ABSTRACT : 161

GENOTYPING BY SEQUENCING AS A RAPID MEANS TO CHARACTERIZE MATURITY GENES IN SOYBEAN:GENOTYPING OF KNOWN ALLELES AND DISCOVERY OF NOVEL ALLELES AT THE *E3* LOCUS

TARDIVEL A, SONAH H, BELZILE F AND O'DONOUGHUE L S[1]

Dépt. de Phytologie, Faculté des Sciences de l'agriculture et de l'alimentation and Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC G1V 0A6, Canada

[1]CÉROM, Centre de recherche sur les grains inc., 740, chemin Trudeau, Saint-Mathieu de Beloeil, QC J3G 0E2 Canada

*E mails: francois.belzile@fsaa.ulaval.ca, louise.odonoughue@cerom.qc.ca

In Eastern Canada, earliness is a trait of considerable importance given the short growing season. Eight maturity or photoperiod insensitivity loci (E1 to E8) have been identified and mapped and, in four cases (E1 to E4), the genes have been cloned. The aim of this work was to develop a set of single nucleotide polymorphism (SNP) markers/haplotypes allowing breeders to rapidly assess the alleles present in their germplasm at these loci. Thanks to the tremendous throughput of modern DNA sequencing technology, it has become possible to develop genotyping by sequencing (GBS) approaches capable of rapidly identifying and genotyping thousands of SNPs across the entire genome. We have developed and optimized GBS protocols for soybean and have used it to characterize a collection of 144 lines relevant for the study of *E* genes, including many near isogenic lines (NILs), as well as a broader set of 305 lines representative of the germplasm in use in public breeding programs in Eastern Canada. In our collection of Canadian germplasm, a total of 4 distinct haplotypes in and around the *E3* locus were observed: one corresponded to the wild-type Harosoy-*E3* allele and another to the mutant Harosoy-*e3* allele that differs by a 13.3 kb deletion in the 3' end of the *GmPhyA3*. A third haplotype was observed among 17 lines that shared an early-flowering phenotype. Sequencing of the *GmPhyA3* gene revealed that these lines shared a novel allele of the gene characterized by a premature stop codon in the third exon. A fourth haplotype is currently undergoing molecular characterization. These first results carried out at the *E3* locus suggest that this approach is a powerful method for rapid allelic characterization on a large set of lines. Extension and application of this methodology to other maturity genes/loci will be extremely useful for breeding purposes.

Genotyping by sequencing as a rapid means to characterize maturity genes in soybean: genotyping of known alleles and discovery of novel alleles at the E3 locus

Louise S. O'Donoughue

CEROM
Centre de recherche sur les grains inc.

World Soybean Research Conference

Durban, South Africa

21st February 2013

Soybean is a short day plant

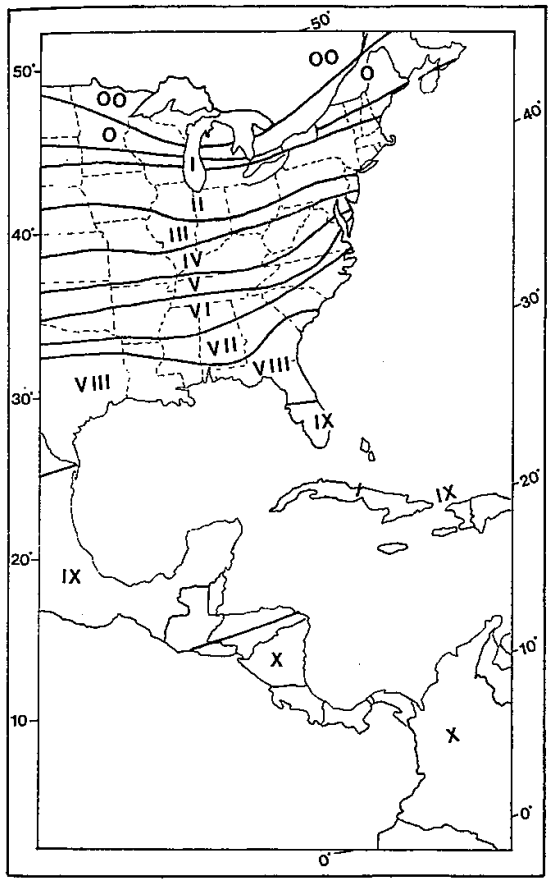
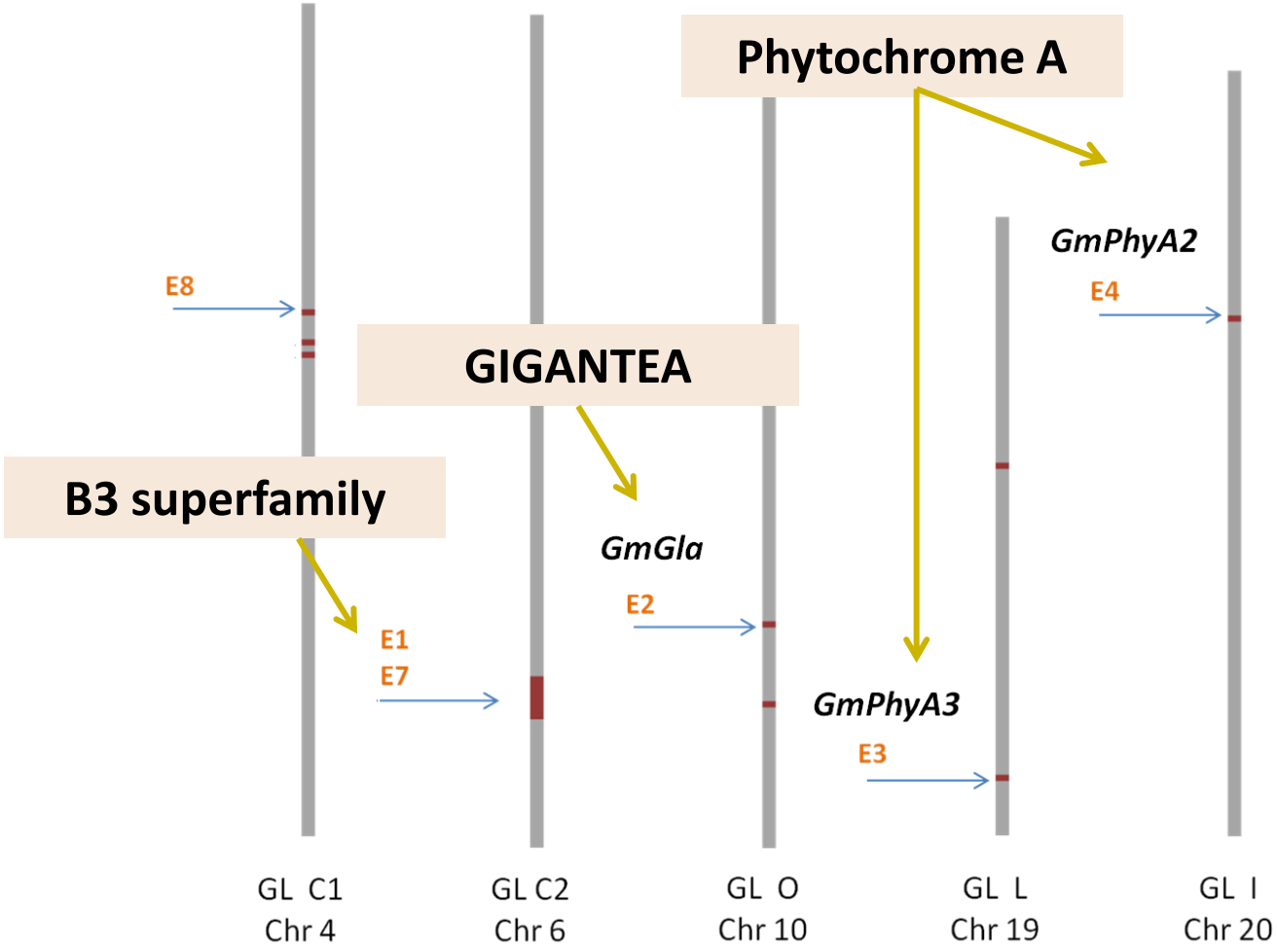


Figure 1.5. Geographical zones of the American continent where soybean maturity groups 00 through X are best grown. From Whigham and Minor (1978). Copyright by Academic Press, Inc.

Soybean is a photoperiod sensitive species. It is known as a short day plant as it requires short days for flower initiation. Under Eastern Canadian latitudes, soybean cultivars have one or more photoperiod insensitivity genes.

9 maturity loci and about 36 QTLs affecting maturity are known – four maturity genes have been cloned

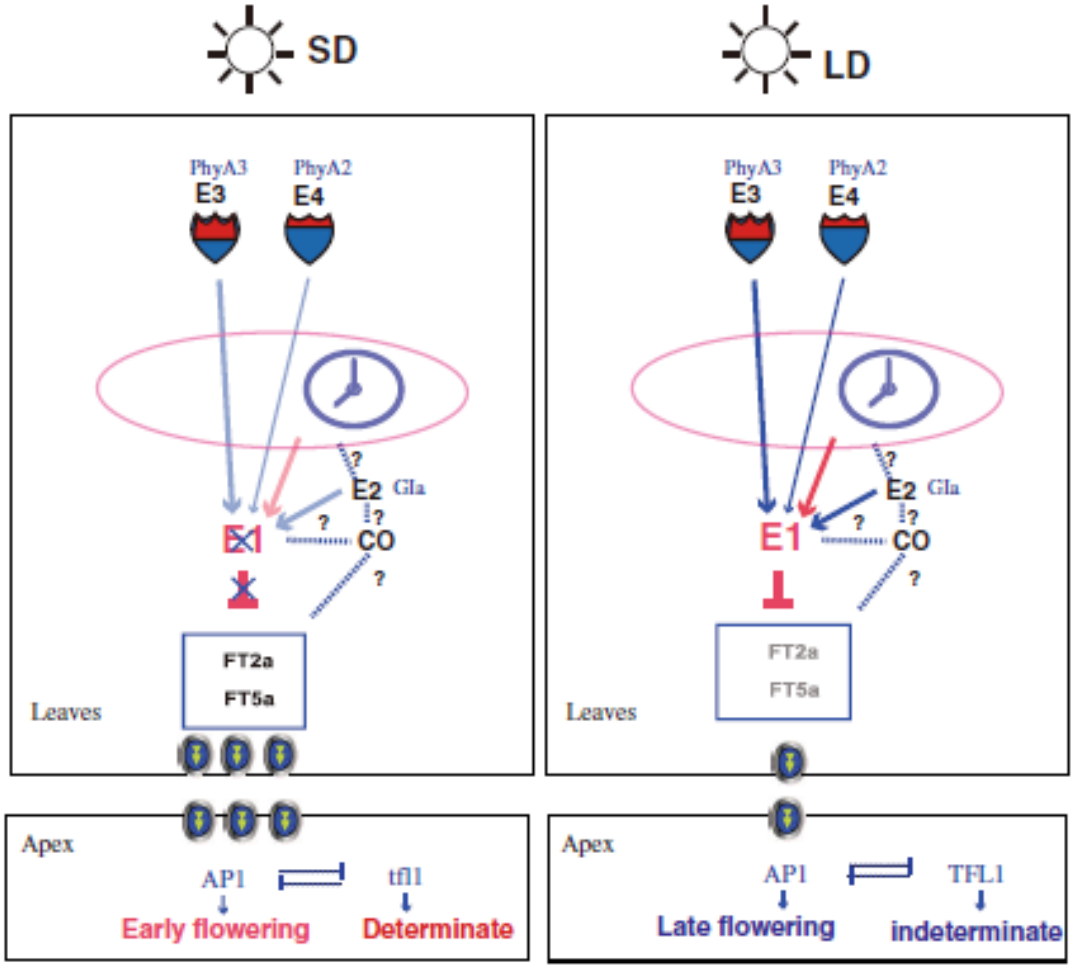


AAFC, Harvey Voldeng, Malcolm Morrison, Elroy Cober 1990 to 2012; Kong et al., 2010; Watanabe et al.,2009 ,2011; Xia et al 2012

Proposed signalling pathway

E1 not expressed

FT expressed



E1 expressed

FT not or little expressed

Xia et al, 2012

Photoperiod insensitivity leads to early maturity

Photoperiod insensitivity usually results from the inactivation of one or more of the E genes and is usually a recessive trait.

The exact genetic basis of early maturity in any given line is largely unknown

Canadian Gene pool

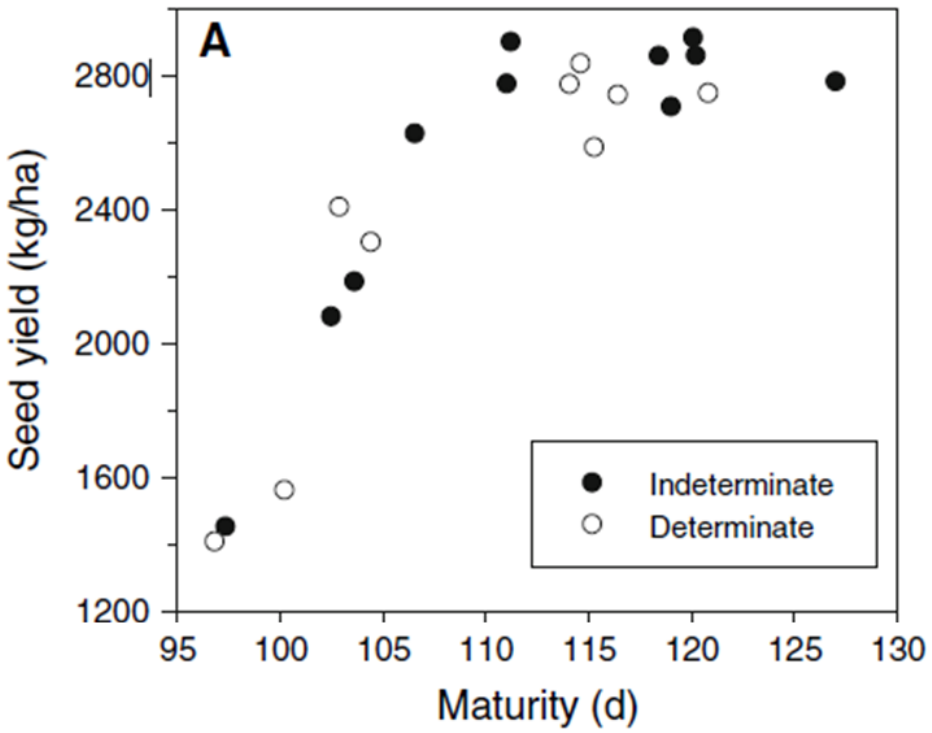
e1? e3?

e2?

e4?

Other sources of early lines

Yield is correlated with maturity



Harosoy NILS

Cober et al, 2010

Breeding objectives

The overall objective of the breeding programme is to develop high-yielding early maturity soybean varieties in order to expand the distribution of soybean production in Canada by;

1. **Gaining a better understanding of the genetic basis for early maturity in Canadian breeding material**

2. **Identifying, if possible, the best combination of maturity genes and alleles to maximize yield**

3. **Developing and using markers for selection of early maturity genes**

First step

Characterizing early parental material and the Eastern Canadian Gene Pool for their status at one of the known maturity locus, the E3 gene, using a Genotyping by Sequencing approach

Plant material

Early parental collection

53
Near Isogenic Lines

Contrasting for maturity genes (E) in five genetic background
Developed by Dr Voldeng and others at AAFC
Genetic basis of maturity relatively well known

91
Early potential Parental lines

Maturity 00 and 000
Varied geographic origins
Genetic basis of maturity unknown

Eastern Canadian collection

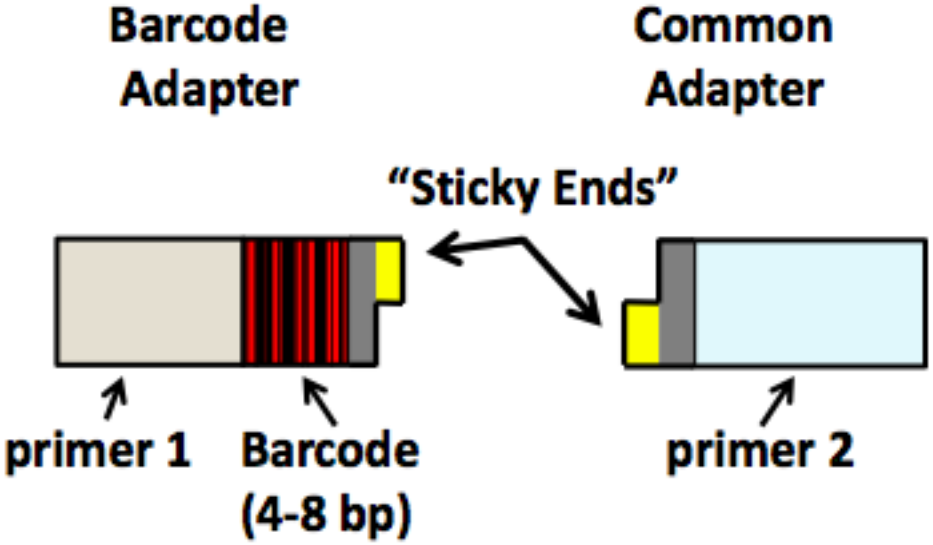
305 cultivars and lines

Representative of the diversity present in the Eastern Canadian gene pool
Maturity 000 to II
Genetic basis of maturity unknown

Marker tool – Genotyping by sequencing GBS

simultaneous identification and detection of a large number of markers

- Restriction digestion of genomic DNA with a methylation sensitive enzyme (*Ape*K1) targets gene rich regions

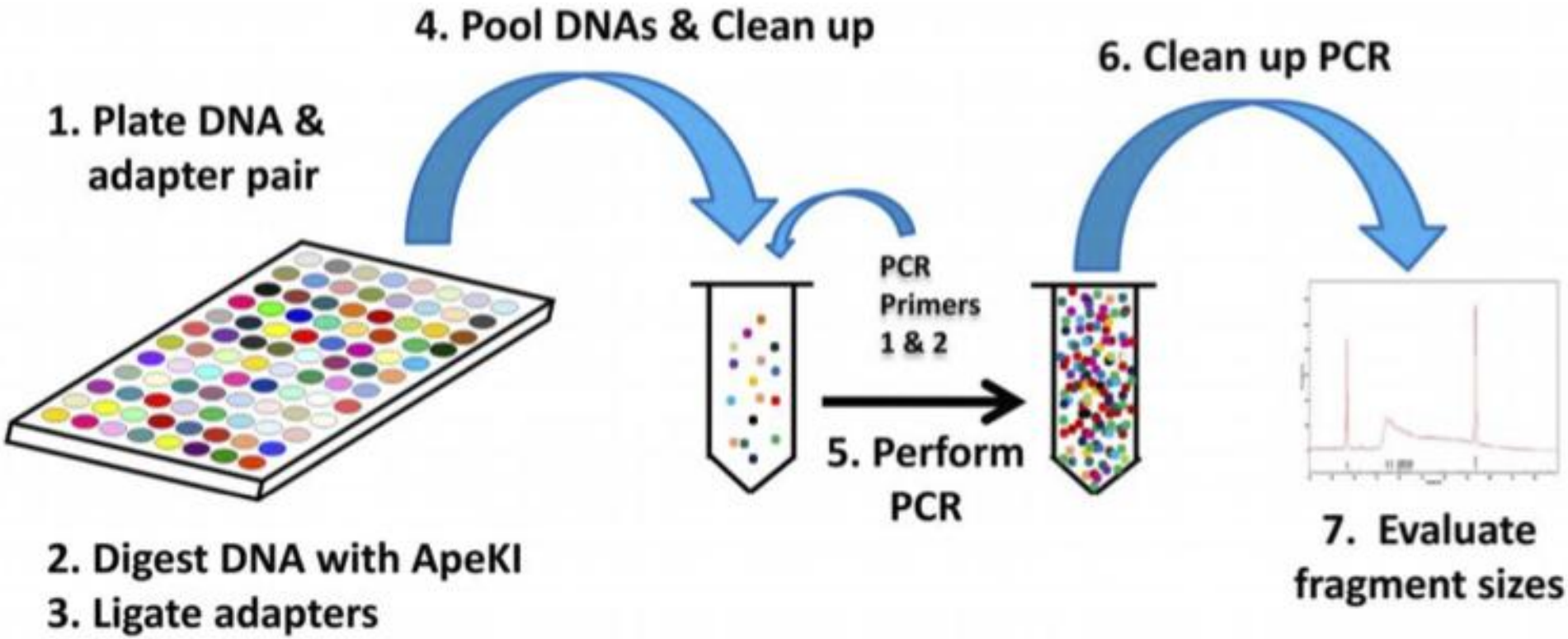


- Ligation of 2 flanking adaptors to the fragments allows PCR amplification. The barcoded adapter permits pooling of many individuals and traceability.

- Sequencing on Illumina machine generates reads of 100 bp flanking *Ape*K1 restriction sites

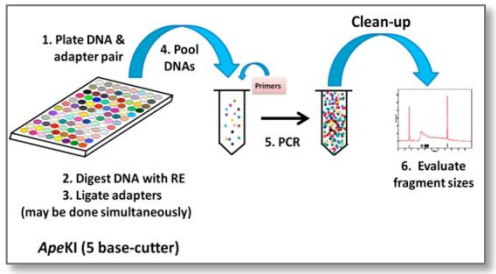
Elshire et al, 2011; Sonah et al,2013

GBS step by step



Elshire et al, 2011; Sonah et al 2013

GBS step by step



Illumina Sequencing by synthesis

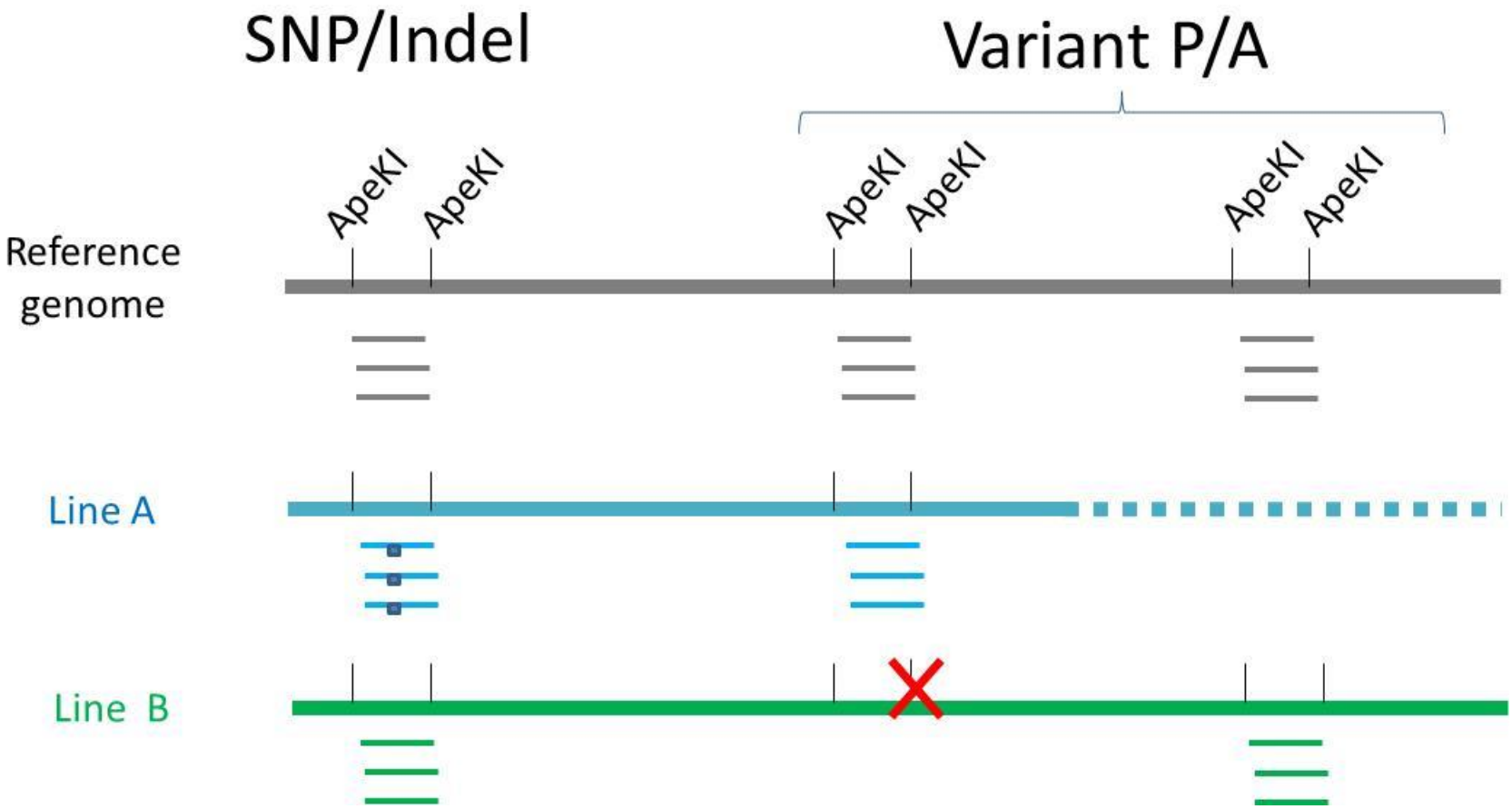
Bioinformatic analysis

IGST-GBS Pipeline

Sequence editing
Alignment to reference genome
SNP Calling (detection, filtration, imputation)

Elshire et al, 2011; Sonah et al 2013

Types of markers and polymorphisms detected by GBS

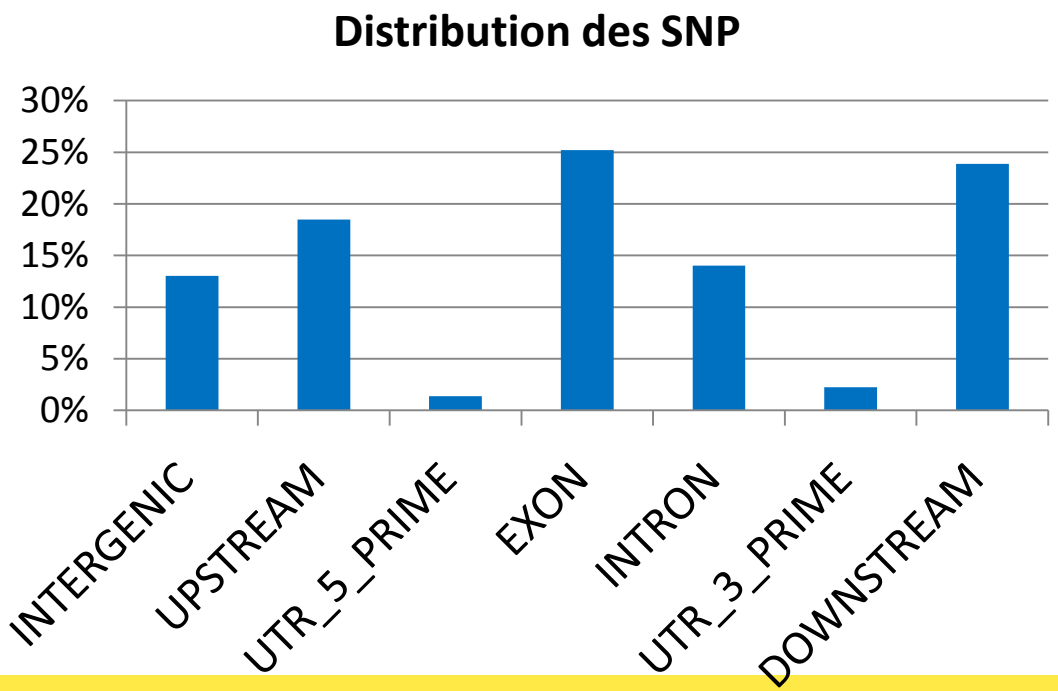


Markers obtained by GBS on our collections

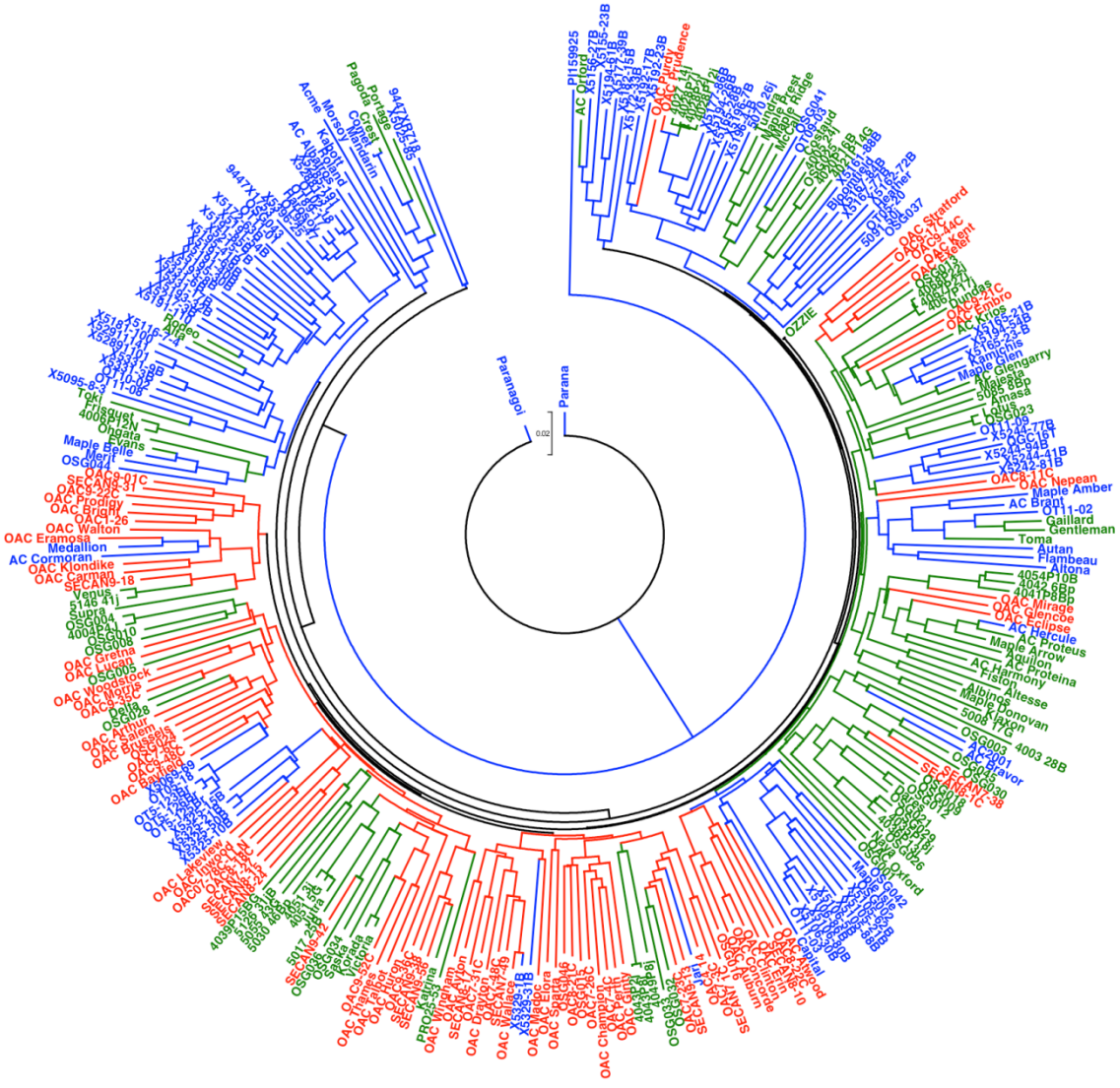
28 101 polymorphisms obtained on the 449 lines of our two collections (Early parental and the Eastern Canadian)

- **26 920 SNP**
- **1261 indel**

→ **Filtration MAF0,02 = 10157 SNP**



Family tree of Eastern Canadian collection based on GBS data

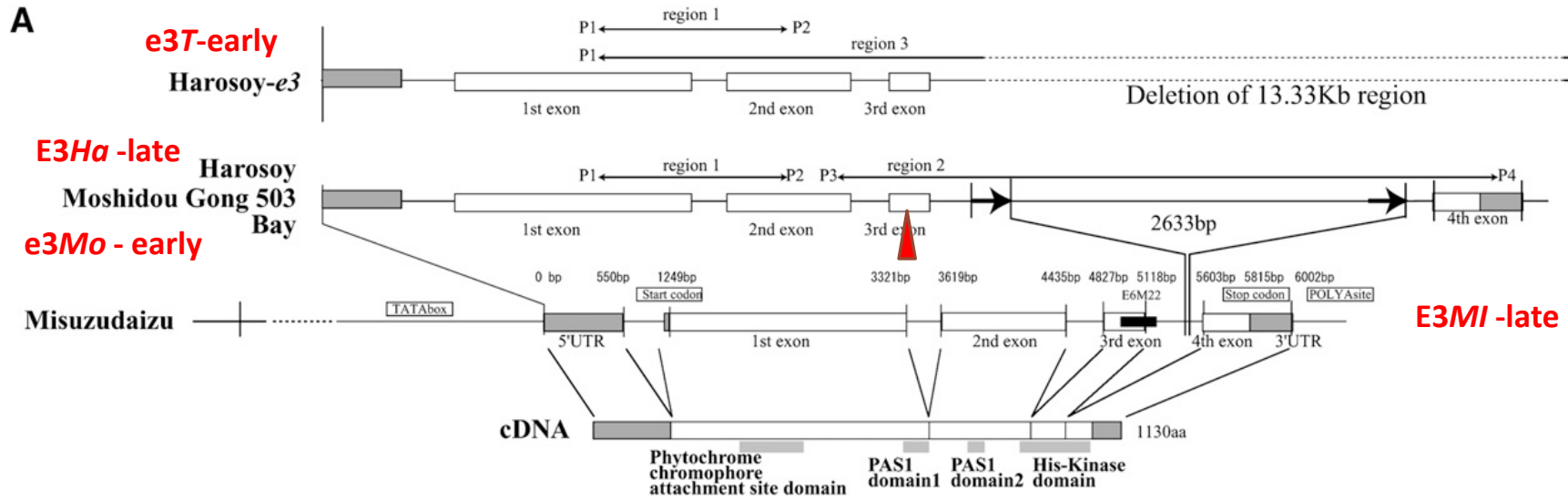


Red = Guelph

Blue = Ottawa

Green = Beloeil

The E3/e3 gene (GmPhyA3)

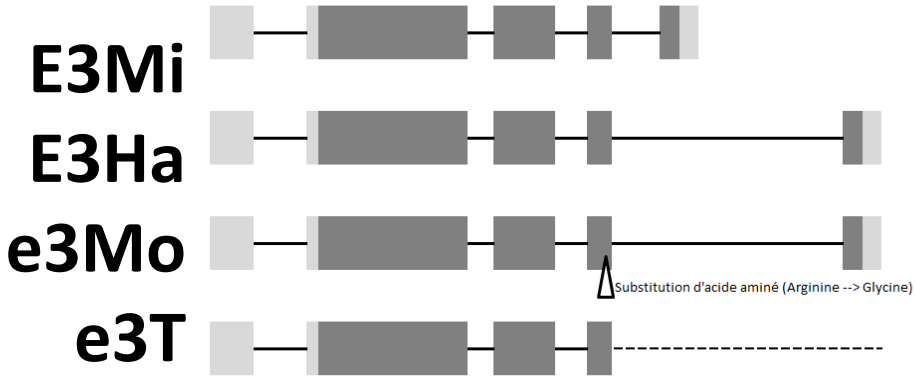


The *e3T* early phenotype is caused by the deletion of a 13-kb region including the fourth exon (Watanabe et al., 2009)

Characterizing E3 in our breeding material

1-Identify haplotypes in the E3 region

2- relate haplotypes to known E3/e3 alleles

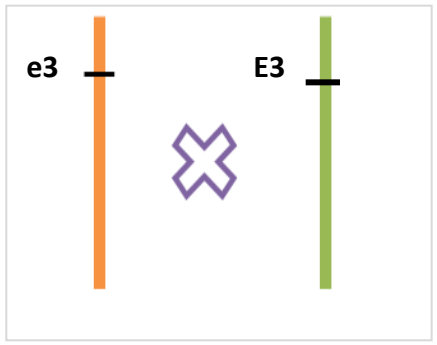


3-Characterise distinct and different haplotypes by Sanger sequencing of E3

Watanabe et al, 2009

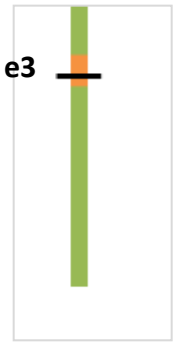
Characterizing E3 in our breeding material

Near-isogenic lines; a very precious tool



PI 196529 Maple Arrow

Repeated backcrosses to Maple Arrow while selecting for early maturity

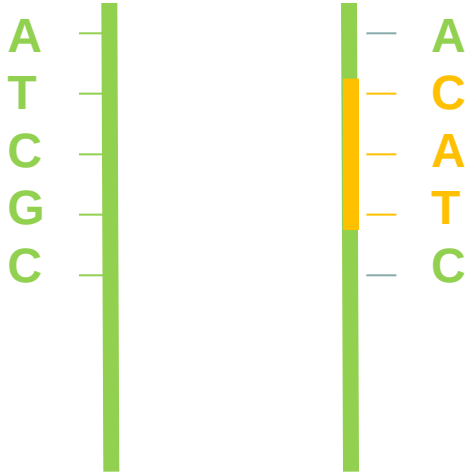


Early maturity e3 allele in a Maple Arrow background

Characterizing E3 in our breeding material

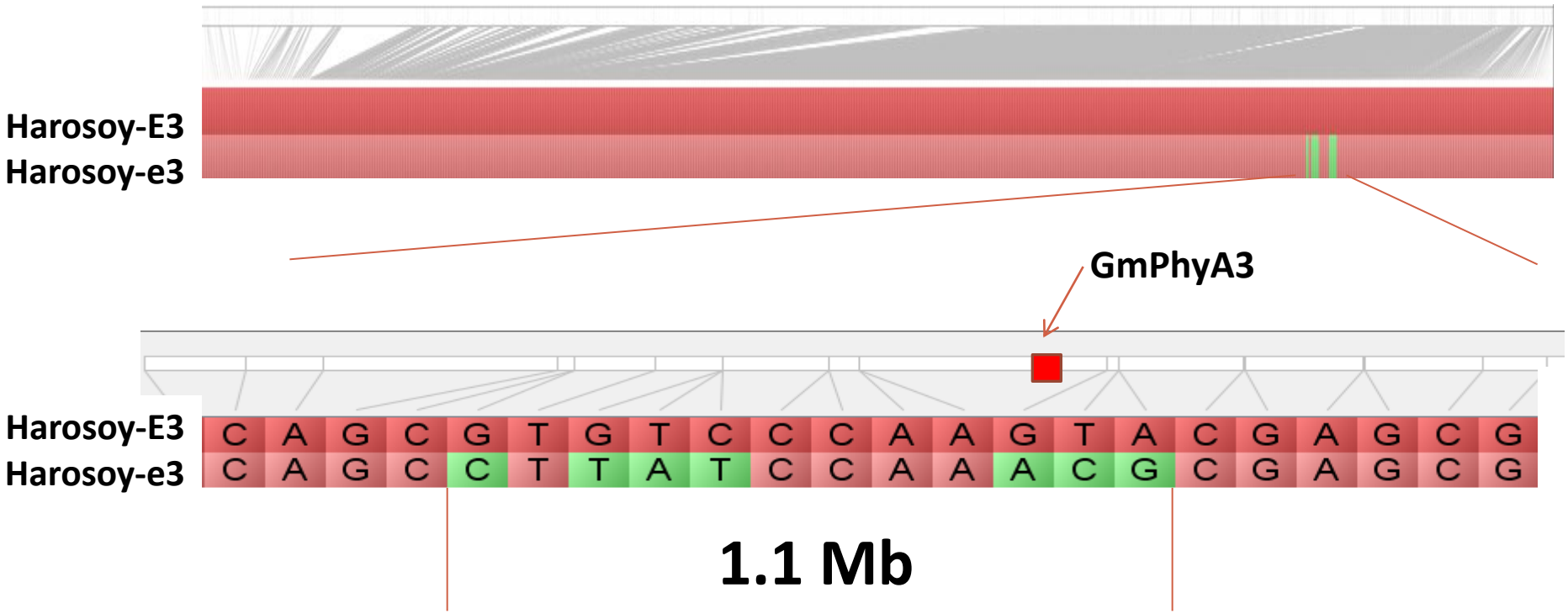
The Genotyping by Sequencing technology was used to generate thousands of markers for each early line including the near isogenic line collection.

The near isogenic lines and the sequence of the E3 gene allowed us to locate the region of interest and polymorphic markers in these regions.

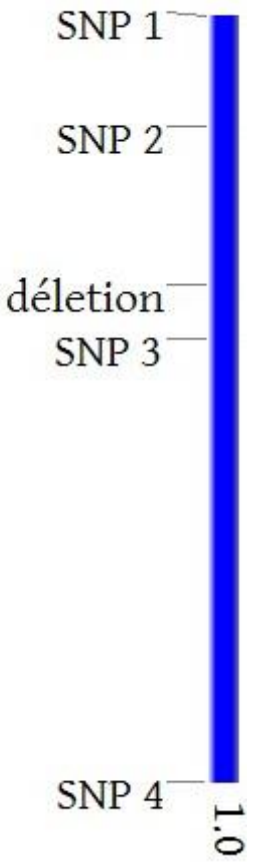


Characterizing E3 in our breeding material

7 of the 507 SNPs on chromosome 19 are polymorphic in the E3 region



Characterizing E3 in our breeding material

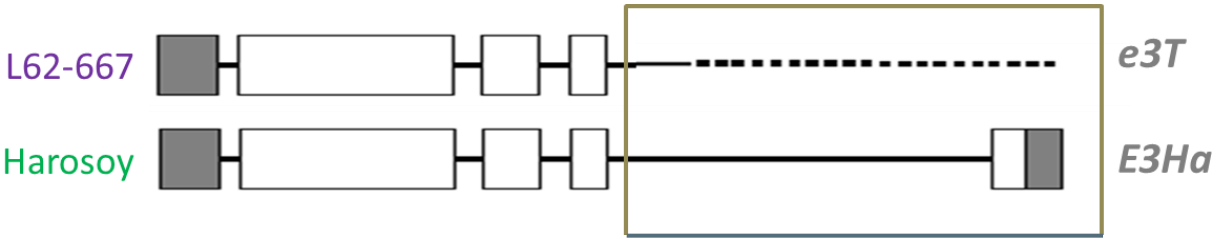


GmPhyA3

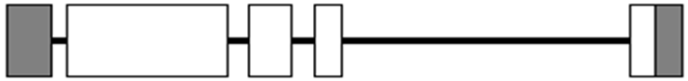
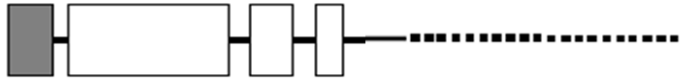
Covers 800 kb
1,6 % of chromosome 19

4 diagnostic SNP markers
+
1 P/A marker characterizing the e3T deletion

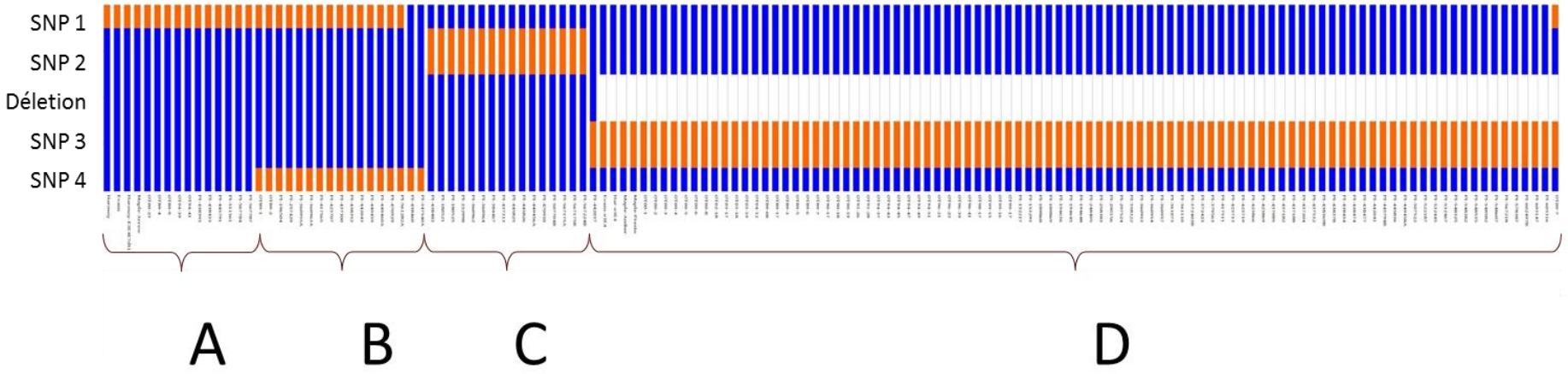
were selected for haplotype construction



Characterizing E3 in our breeding material; The early parental and NIL collection

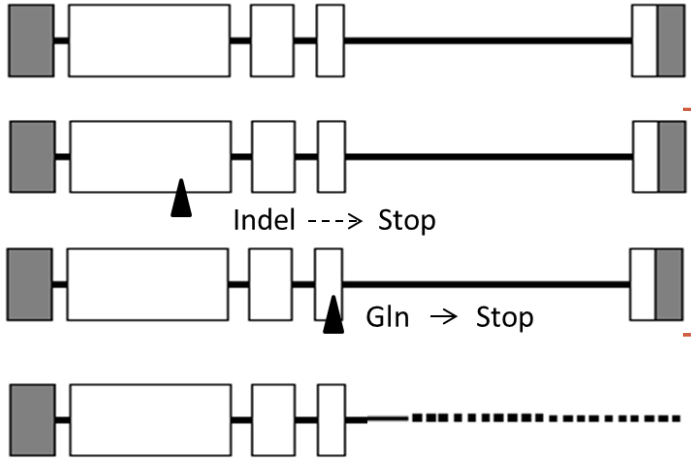
A	**late**	*E3Ha*	
B	**early**	**?**	
C	**early**	**?**	
D	**early**	*e3T*	

Sanger sequencing

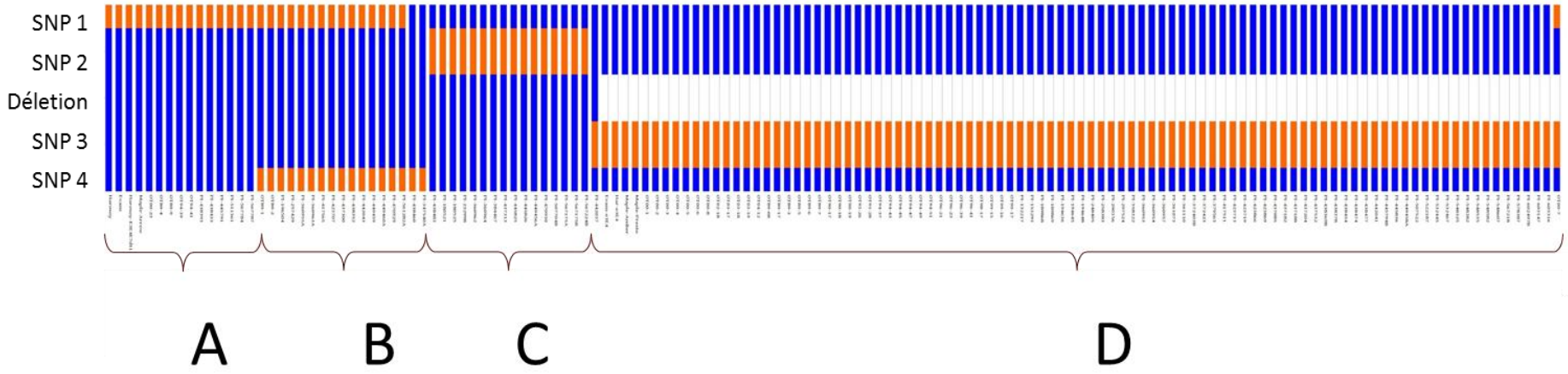


Characterizing E3 in our breeding material; The early parental and NIL collection

A	**late**	*E3Ha*
B	**early**	*e3P2*
C	**early**	*e3P1*
D	**early**	*e3T*

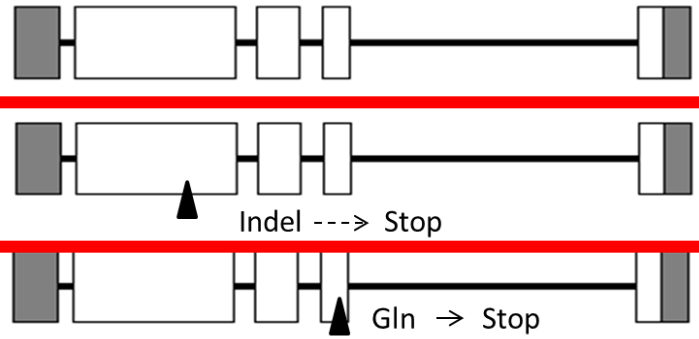


Premature stop codon

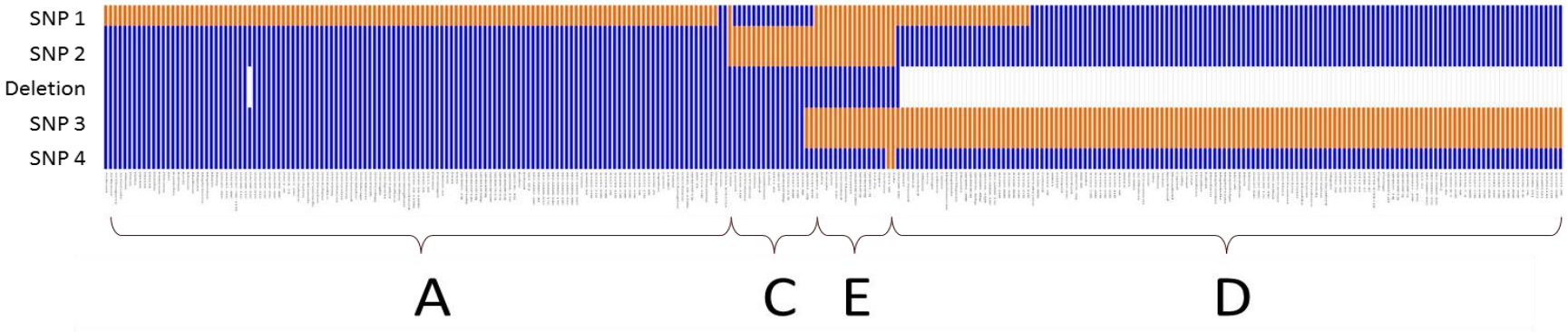


Characterizing E3 in our breeding material; The Eastern Canadian collection

A	**late**	*E3Ha*
B	**early**	*e3P2*
C	**early**	*e3P1*
D	**early**	*e3T*
E	**unknown**	

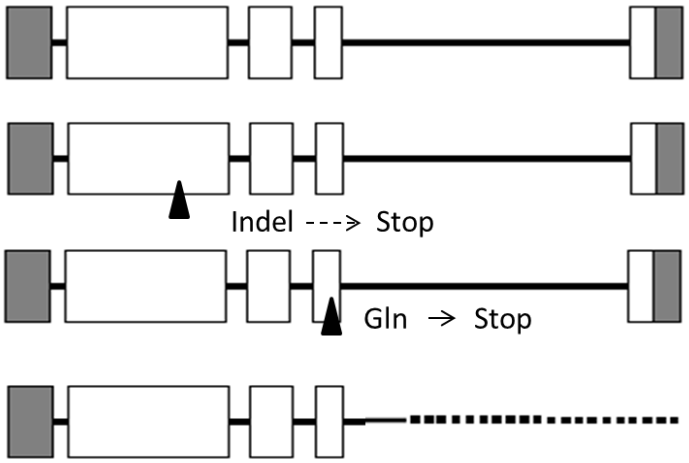


Missing Haplotype B

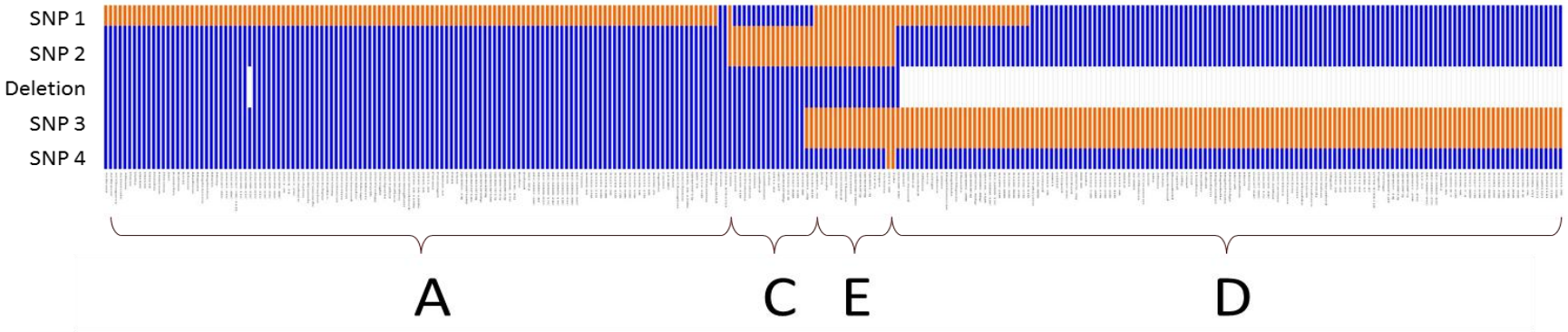


Characterizing E3 in our breeding material; The Eastern Canadian collection

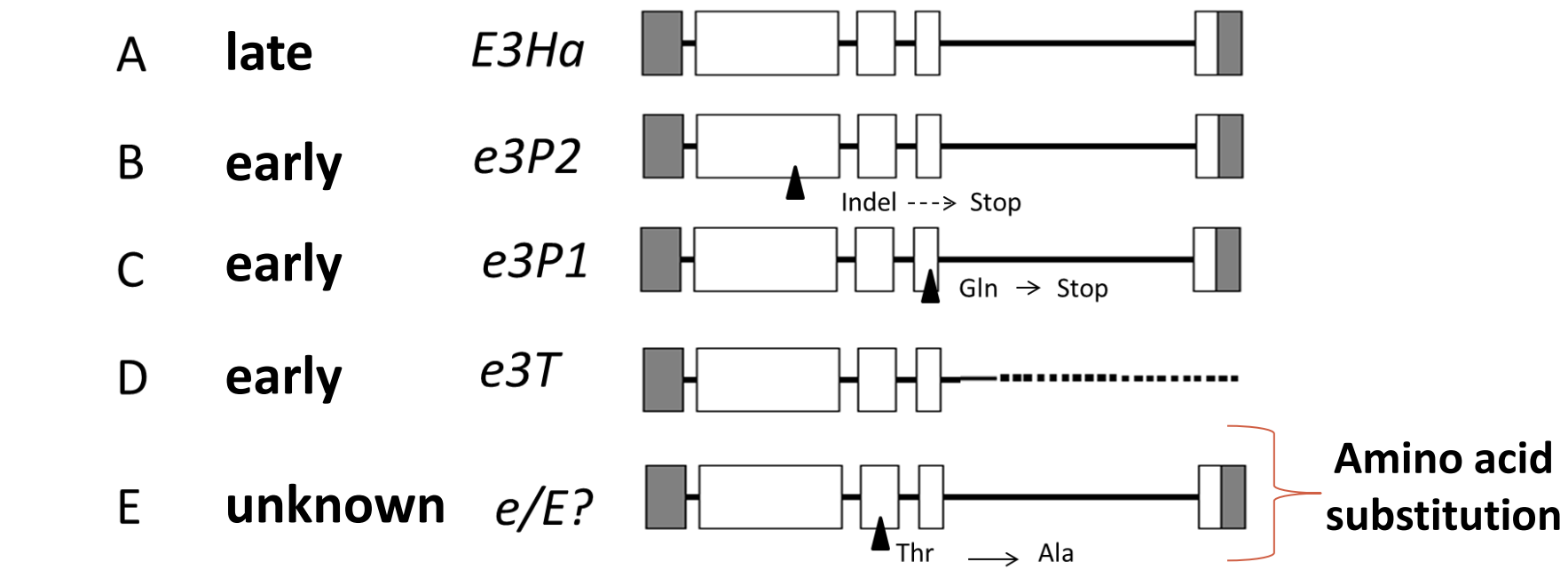
A	**late**	*E3Ha*
B	**early**	*e3P2*
C	**early**	*e3P1*
D	**early**	*e3T*
E	**unknown**	***?***

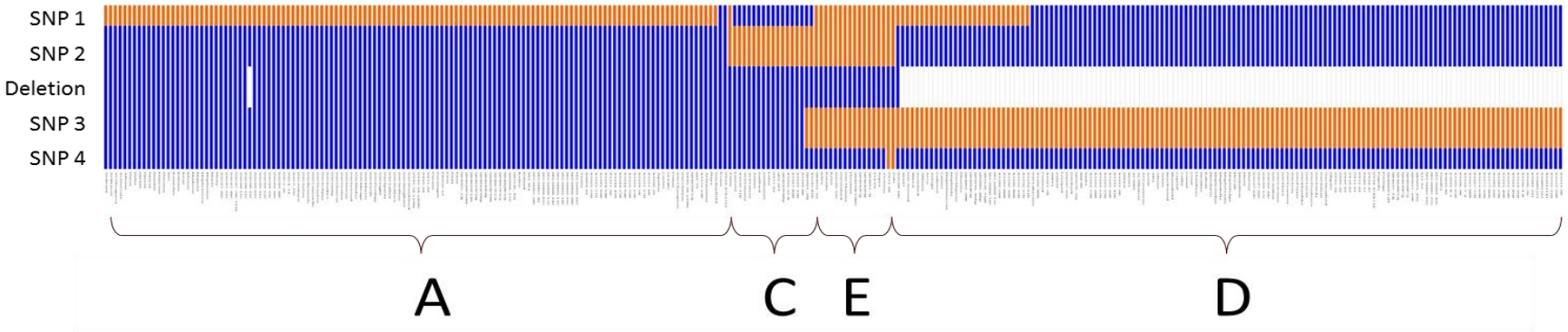


Sanger sequencing



Characterizing E3 in our breeding material; The Eastern Canadian collection





Conclusion

- **This work has allowed the identification of genetic profiles which represent functional and non-functional alleles of the *E3* locus. Thus, the GBS approach is a very powerful tool for the rapid characterization of large collections of soybean lines in terms of their status at this maturity locus.**

What next?

- Characterize the early parental collection and Canadian Northeastern gene pool for the other known maturity genes E1, E4 and E2 is underway
- Characterize the early parental and Eastern Canadian collections for maturity, agronomic characteristics and quality
- Study the relationship between the various maturity gene combinations and agronomic characteristics including yield and identify the best combinations for maturity and yield
- Incorporate the markers as selection tools into the ongoing early maturity breeding program

Acknowledgements

- Université Laval
  - Aurélie Tardivel
  - François Belzile
  - Humira Sonah
- AAFC
  - Elroy Cober
- University of Guelph
  - Istvan Rajcan
- CÉROM
  - Maryse Desrochers

This work is funded jointly by The Canadian Field Crop Research Alliance and Agriculture and Agri-Food Canada under the Developing Innovative Agri-Products Program (DIAP)

CÉROM
Centre de recherche sur les grains inc.

UNIVERSITÉ LAVAL

MERCI

CÉROM
Centre de recherche sur les grains inc.

Bâtiment CÉROM

740, chemin Trudeau
Saint-Mathieu-de-Beloeil (Québec) J3G 0E2

Téléphone : 450 464-2715
Télécopieur : 450 464-8767
Site internet : www.cerom.qc.ca
Courriel : cerom@cerom.qc.ca

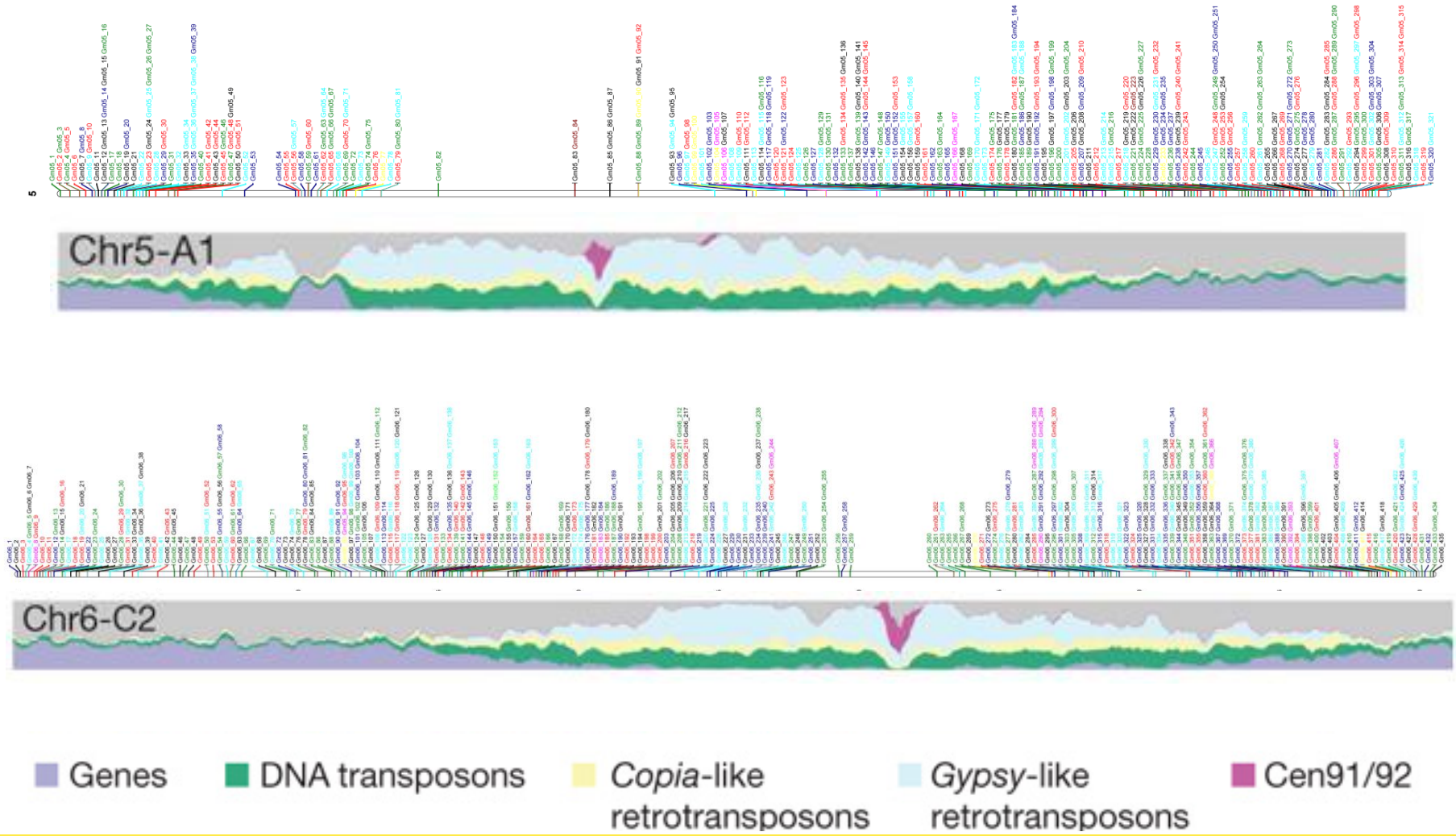
Crédit photo : Mario Maurice - graphisme - pilardesign.com

Objectives

Objectives;

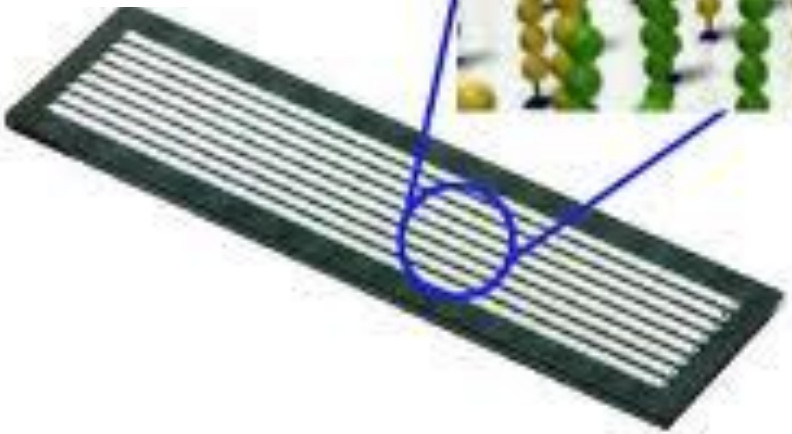
- **Characterize parental germplasm for their status at known maturity loci and evaluate their effect on yield**
- **Identify new loci involved in the control of maturity and their possible effects on yield**
- **Identify the best combinations of maturity genes and alleles for the development of high yielding early maturity soybeans**
- **Develop makers for selection in the breeding program to target the desired gene combinations and accelerate the breeding process**

Chromosomal distribution of GBS SNPs



Sonah et al, 2013

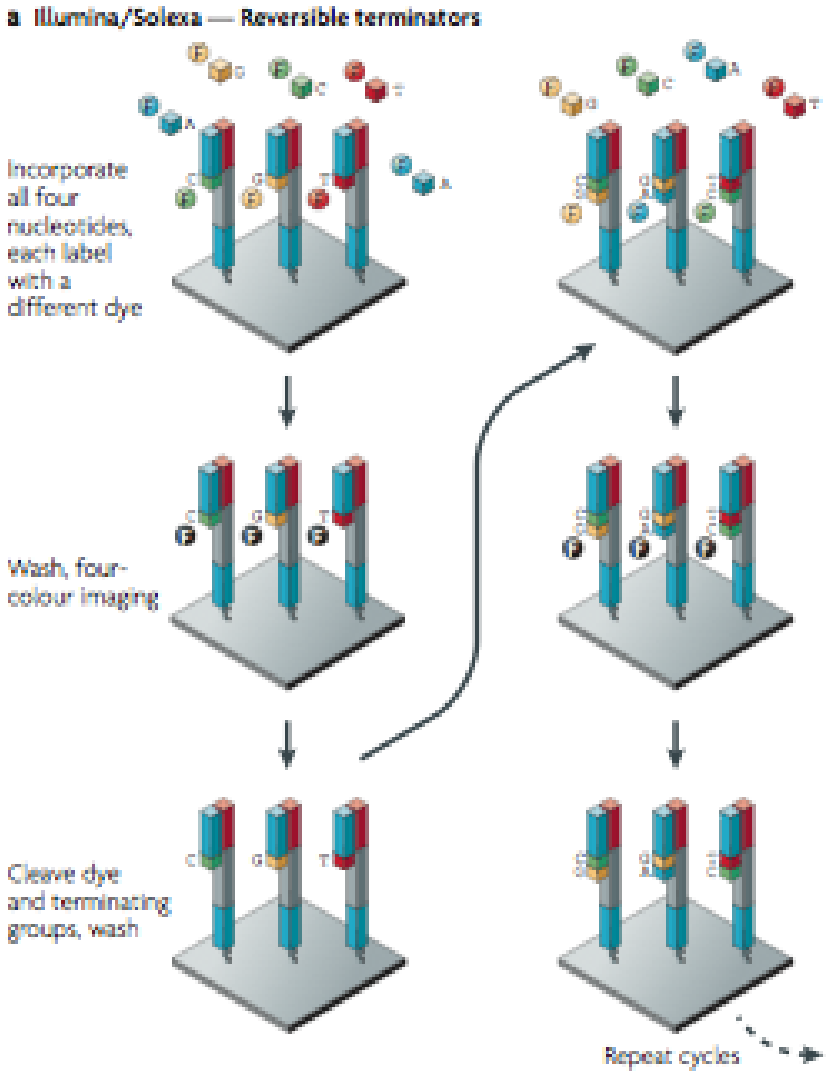
Illumina flow cell

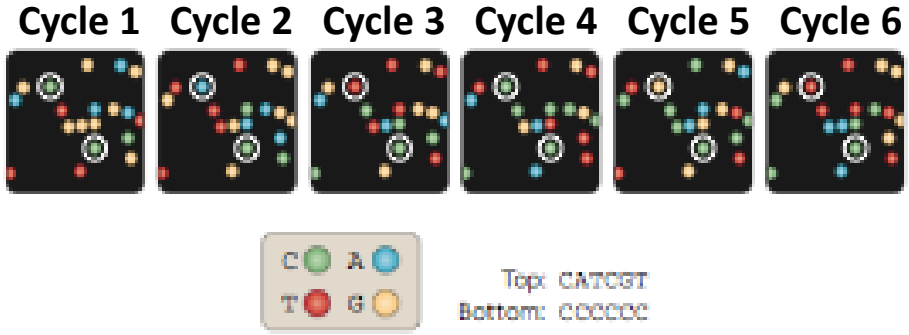


Floor of the flow cell is covered with capture oligos that can lead to amplification of captured fragments

- **1 flow cell has 8 lanes**
- **Each lane can produce up to ~200 million reads**

Illumina sequencing by synthesis





- At each cycle, one and only one labeled base is added
- Imaging at each cycle determines which base had been added to each cluster
- After 100 cycles, the sequence of exactly 100 bases has been determined for each cluster and this produces a "read"
- For each base in the read, a quality score is assigned to indicate how good the base call was

(Metzker, 2010)

How much does it cost?

- DNA extraction ~5$
- GBS library preparation (96-plex) ~7-10$
- Sequencing (96-plex) ~12-15$

For less than 30$ per sample, we can obtain many thousand SNPs per individual

How long does it take?

- DNA extraction and quantification/QC — 1 wk.
- GBS library preparation (UL-IBIS) — 1 wk.
- Sequencing — ~4 wk.
- Analytical pipeline — ~2 wk.

In about 2 months, you can obtain a lot of data with very little hands-on time