CropSNPs

A low-cost genotyping solution for breeding programs

François Belzile
Université Laval

Canadian Barley Symposium
Winnipeg, 24-25 Feb. 2020

UNIVERSITÉ LAVAL

Why do we need/want to fingerprint breeding lines?

- Assess the relationships between lines in a breeding program
- Identify which lines would make the best crosses/progeny
- Verify if F1 seed are indeed the result of the intended cross
- Perform selection of the best lines
  - For specific traits
  - For genetic makeup
- Identify advanced breeding lines that could become the next generation of parents

Genetic fingerprinting technologies

"SNP chip"

~ 50,000 SNPs
$50-100/line

Genotyping-by-sequencing ("GBS")

~ 50,000 SNPs
$20-30/line

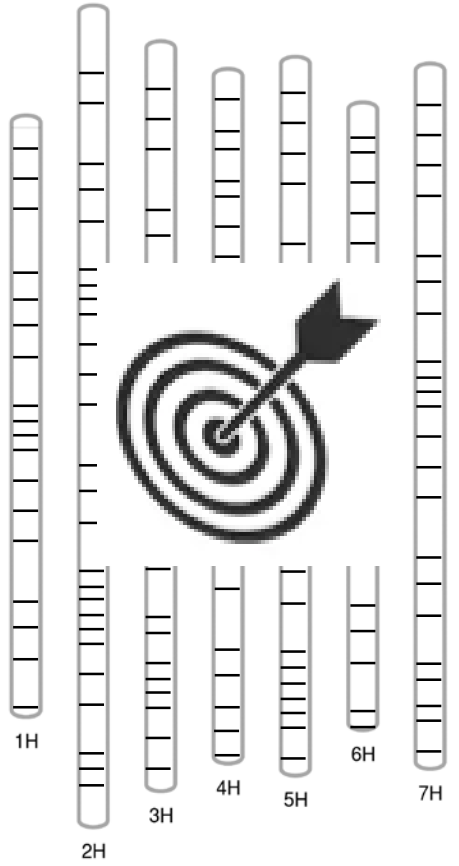
The challenge: fingerprinting a lot of lines

- Basic research often requires genotyping hundreds of lines
- Breeding can require the genotyping of tens of thousands of lines

The opportunity: lower marker density is ok

- For basic research/marker discovery, a dense marker coverage is required
- For most breeding work, a much lower density (100s to 1000s)

Basic research



Breeding

CropSNPs project
(joint effort for barley and soybean)

- Step 1: Assess existing genetic diversity (@high density)
  - Assemble collections representative of Canadian breeding programs
    - 500 barley lines/500 soybean lines
  - Characterize these with lots of markers (~50,000 SNPs)
- Step 2: Identify subset of well-distributed markers
  - Build genetic maps to provide information on the position of markers
  - Select desired subset of markers for low-density genotyping
- Step 3: Design low-cost genotyping solution
  - Develop assays that minimize per-sample cost
  - Validate and optimize performance

Step 1: Assess genetic diversity
Assembled genetic material

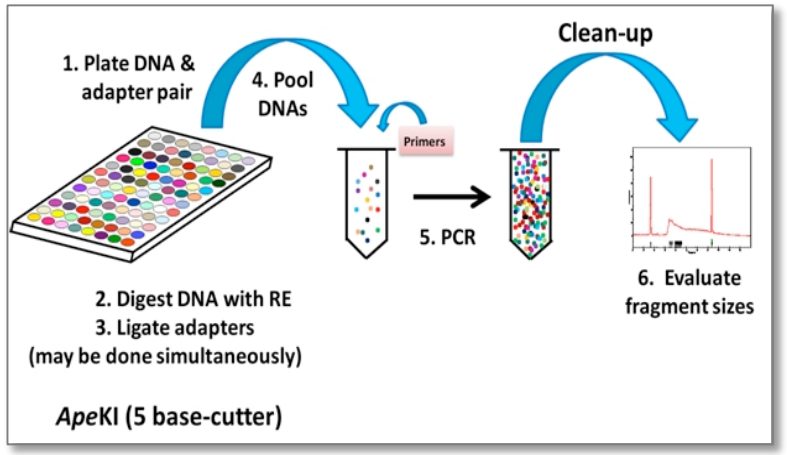
- Objective: 500 barley lines
  - 294 from Western Canada (SK and MB) – 2-row
  - 324 from Eastern Canada (ON and QC) – 6-row
- Provided by breeders as being representative of their breeding program

Step 1: Assess genetic diversity

The process

Leaf tissue

DNA



GBS libraries

High-throughput Sequencing

Bio-informatic analysis

Step 1: Assess diversity

Results for high-density marker coverage

Markers

Lines

	33305: 623555738	33306: 623708635	33307: 623708711	33308: 623714756	33309: 623714761	33310: 623714789	33311: 623727873	33312: 623885178	33313: 623885188	33314: 623885199	33315: 623885258	33316: 623901198	33317: 623901240	33318: 623901291	33319: 624277302	33320: 624282731	33321: 624282740	33322: 624306515	33323: 624306561	33324: 624307439	33325: 624346484
M122	A	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
Sedna	A	T	G	C	T	G	T	T	C	T	T	G	A	A	C	C	C	T	C	C	G
SQO003L	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
SQO00901	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
SQO00902	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
UL0745	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
UL1082	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
UL155	A	T	G	C	T	G	T	K	C	T	W	R	M	A	C	C	C	T	C	C	G
UL164	A	T	G	C	T	G	T	T	C	T	T	G	A	A	C	C	C	T	C	C	G
UL166	A	T	G	C	T	G	T	T	C	T	T	G	A	A	C	C	C	T	C	C	G
UL1681	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
Tupper	R	A	S	A	T	G	T	G	C	Y	A	G	A	G	C	C	T	C	C	C	G
ACStephen	A	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
25CFQ41337	G	A	G	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
01FSM2991	A	A	C	A	T	G	T	G	C	T	A	G	A	G	C	C	T	C	C	C	G
01FSM3901	G	A	C	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
21FHD164	G	A	S	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
21FHD172	G	A	C	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
27CFQ39B126	G	A	C	A	T	G	T	G	C	T	A	A	C	A	C	G	T	C	C	C	A
27CFQ54D78	A	T	C	A	C	A	G	G	C	T	A	G	A	A	C	C	T	C	T	C	G

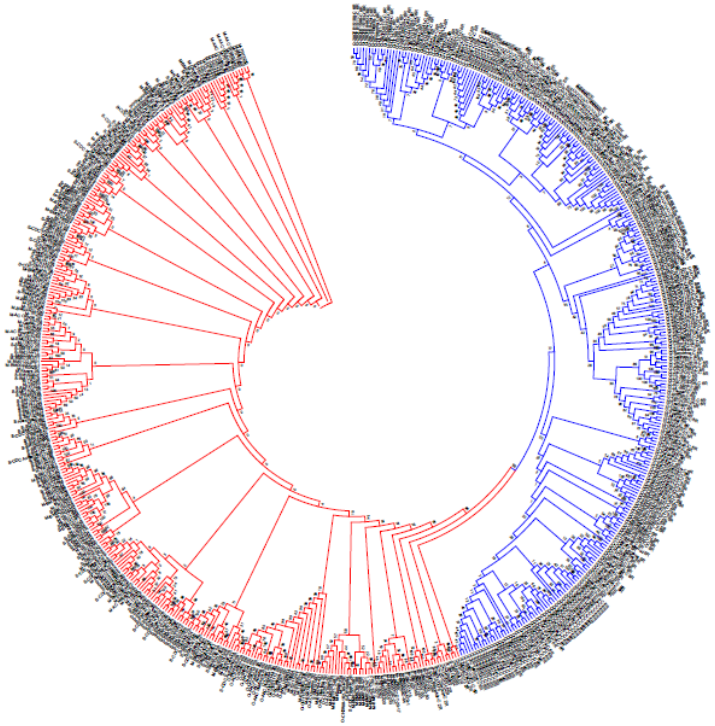
- Number of markers

Program	# of SNP markers
Saskatoon	54,822
Brandon	51,441
Ottawa	64,204
Quebec	66,659
TOTAL	79,889

Step 1: Assess diversity

Results for high-density marker coverage

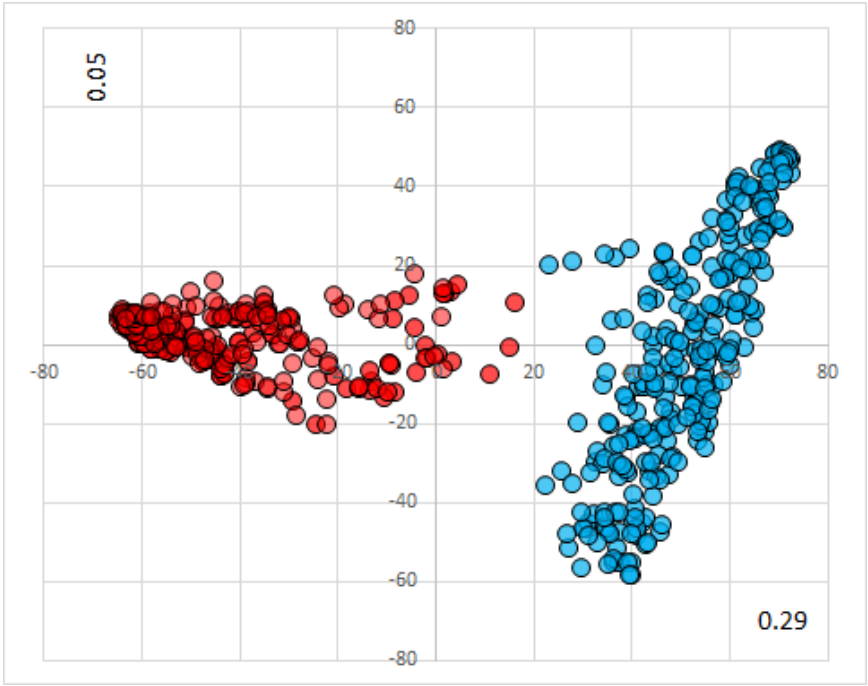
- How are these lines related to one another?



Phylogenetic tree

2-row

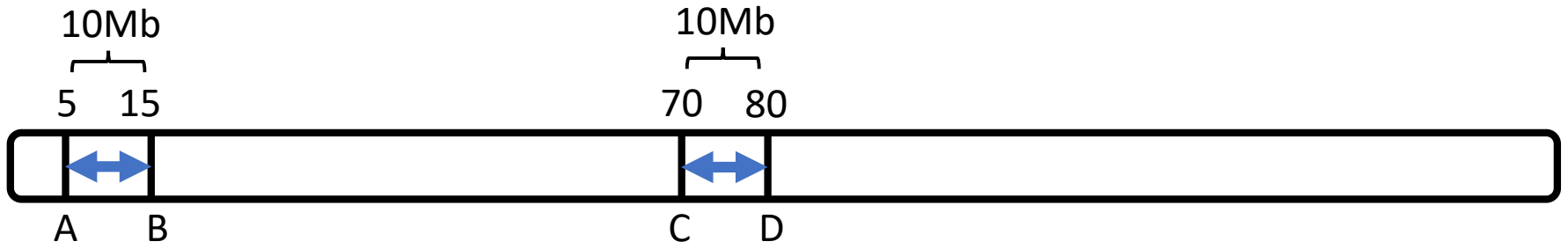
6-row



Principal component analysis

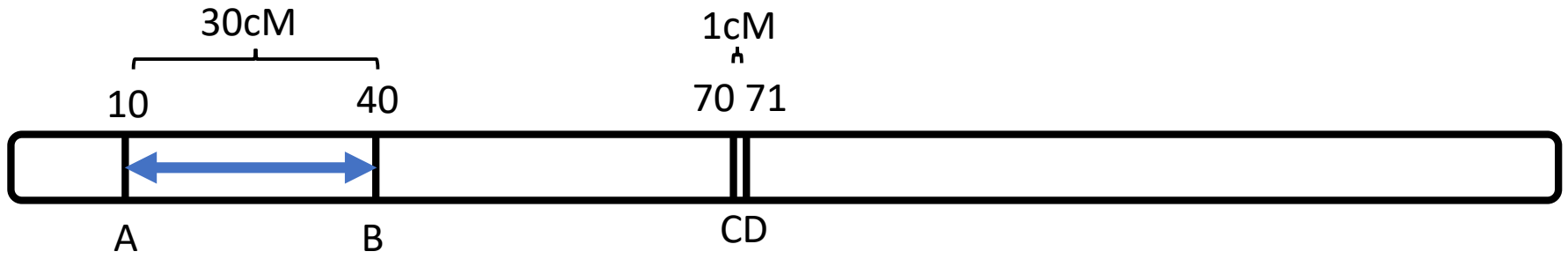
Step 2: Subset of well-distributed markers

Physical position



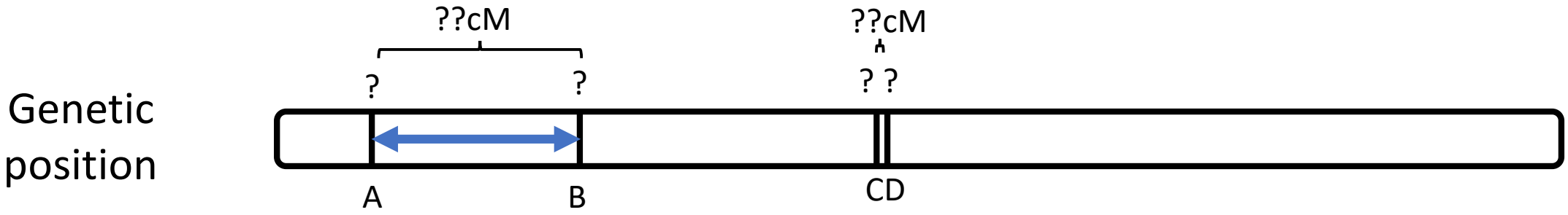
How often will two markers this far apart (physical distance) be inherited together in the progeny?

Genetic position



The genetic map reflects the shuffling of genes that occurs in a breeding program

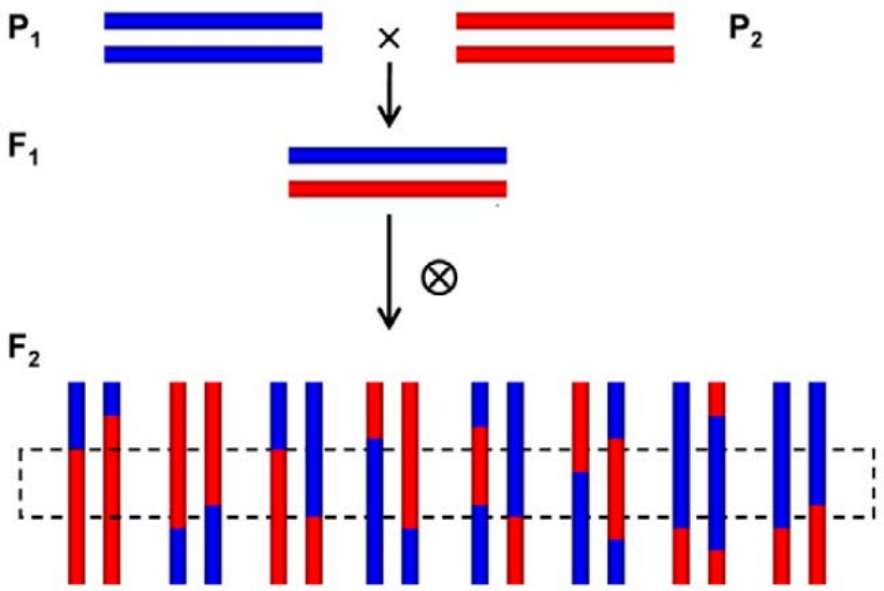
Step 2: Subset of well-distributed markers



To determine the genetic distance between pairs of markers,
we need to build genetic maps

Step 2: Subset of well-distributed markers

Genetic mapping



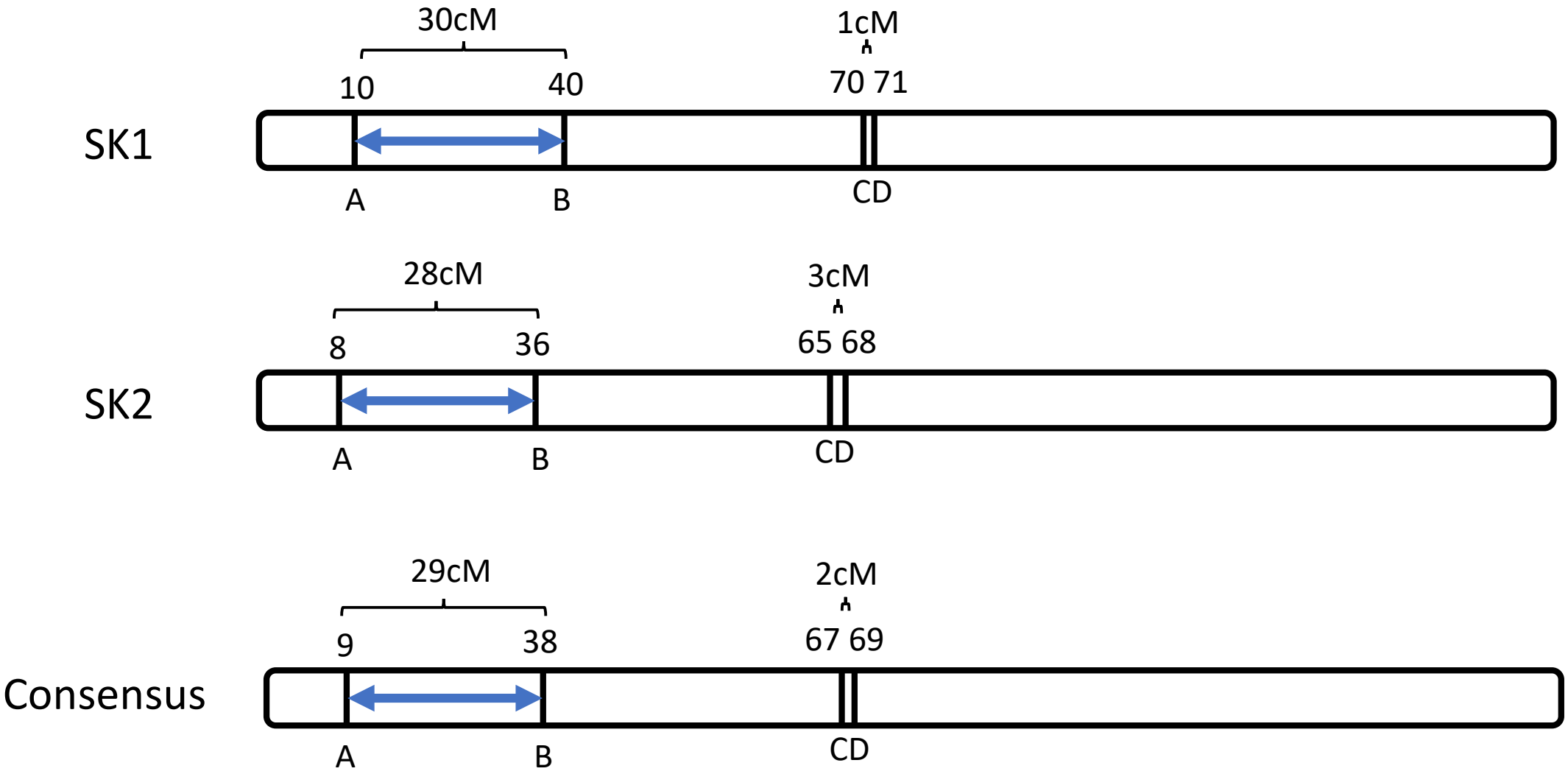
- Make a cross between two lines differing at numerous positions in the genome
- Assemble collections of progeny lines (~200-300/cross)
- Perform genotyping on the individual progeny lines with a few thousand markers
- Calculate the genetic distance between marker pairs

Genetic mapping – In progress

- Saskatoon:
  - 2 populations of 200 lines
  - Genotyping completed
- Brandon
  - 2 populations of 200 lines
  - Genotyping completed
- Ottawa
  - To be determined
- Quebec
  - 2 populations totaling >400 lines
  - Genotyping soon

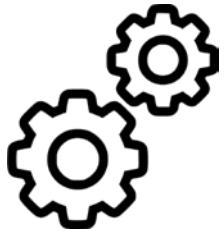
Ready to build individual maps for each cross

Genetic mapping – Next Steps



CropSNPs project
(joint effort for barley and soybean)

- Step 1: Assess existing genetic diversity (@high density)
  - ✔ ☐ Assemble collections representative of Canadian breeding programs
    - 500 barley lines/500 soybean lines
  - Characterize these with lots of markers (~50,000 SNPs)
- Step 2: Identify subset of well-distributed markers
  - Build genetic maps to provide information on the position of markers
  - Select desired subset of markers for low-density genotyping
- Step 3: Design low-cost genotyping solution
  - Develop assays that minimize per-sample cost
  - Validate and optimize performance



Collaborators

Aaron Beattie
Univ. of Saskatchewan

Ana Badea
AAFC-Brandon

Raja Khanal
AAFC-Ottawa

Brian Boyle
Université Laval

UNIVERSITÉ LAVAL

My team

Vincent-Thomas St-Amour
Amina Abed
Martine Jean

Thanks to our funders!