Theor Appl Genet (2014) 127:211–229
DOI 10.1007/s00122-013-2211-9

ORIGINAL PAPER

Characterization of the genetic changes in a multi-generational pedigree of an elite Canadian soybean cultivar

Christopher M. Grainger · Istvan Rajcan

Received: 25 February 2013 / Accepted: 3 October 2013 / Published online: 20 October 2013

Abstract Genotyping through the pedigrees of elite soybean [*Glycine max* (L.) Merr.] cultivars developed by a breeding program represents an opportunity to explore and characterize various molecular and genetic changes that are a direct result of long-term selection by soybean breeders. For soybeans bred for Ontario Canada, one such elite cultivar was OAC Bayfield, which had exceptional commercial success as well as being a parent of a number of successful cultivars developed by multiple independent breeding programs. A total of 42 genotypes from six different breeding programs, comprising the multi-generational pedigree of OAC Bayfield were genotyped with molecular markers and chromosomal inheritance was tracked throughout the pedigree. Cluster analysis showed high congruence with the known pedigree and identified three distinct ancestral groups. The ancestral genotypes contained the majority of the rare alleles, with the cultivar CNS having the greatest number of unique alleles. The graphical genotype profile for the 20 chromosomes revealed conserved allelic composition which has been assembled in certain chromosomes in the form of specific linkage blocks, which were either a result of recombination involving ancestral linkage blocks or linkage blocks introduced from the cultivar Fiskeby-V. The identification of highly structured, conserved genomic regions are important for future breeding efforts as they are indicators of preferentially selected regions, or conversely, may be a contributing factor to low genetic gains due to mass fixation across a breeding program's germplasm.

Abbreviations

MG	Maturity group
cM	Centimorgan
PIC	Polymorphism information content
SSR	Simple sequence repeat
LG	Linkage group

Introduction

Continued improvement of soybean cultivars relies on genetic variation between parents to produce a population of individuals in which there is enough phenotypic variation to improve traits of interest. As phenotypic variation is positively associated with genetic diversity (Moose and Mumm 2008), it is important for a soybean breeding program to maintain genetic diversity for continued genetic gain of a particular trait. The narrow genetic base of North American soybean breeding programs is well documented, with 35 ancestral genotypes accounting for 95 % of the genes found in modern cultivars (Gizlice et al. 1994). This number is further reduced when comparing southern versus northern environments. For the northern genetic base, 10 ancestors account for 80 % of the genes, with only five cultivars; Lincoln (24 %), Mandarin (17 %), Richland (11 %), AK Harrow (7 %) and Mukden (5 %) accounting for ~65 % of all genes in modern northern germplasm (Gizlice et al. 1994). Combined with intense selection over multiple generations in applied breeding programs, this process has led to concerns over the ability to make continued genetic gains for cultivar improvement as well as increased vulnerability to disease and insect epidemics (Hyten et al. 2006).

Communicated by A. H. Paterson.

C. M. Grainger · I. Rajcan (✉)
Department of Plant Agriculture, Crop Science Building, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada
e-mail: irajcan@uoguelph.ca

Widening the genetic base of a breeding program can be accomplished through plant introductions or other exotic material (Tanksley and McCouch 1997). However, breeders must balance this need against maintaining the commercial performance of their germplasm (Moose and Mumm 2008), as incorporating exotic material can have negative consequences. One such consequence is breaking apart favourable gene linkages (assembled through recombination) that breeders have selected for in a particular growing environment over time. This is a major reason for the general strategy of "elite by elite" crossing that many soybean breeders employ, ensuring that linkage blocks associated with performance are transmitted to subsequent cultivars.

Preferential chromosome transmission from a specific parent has been previously reported by Lorenzen et al. (1996). The authors found that various soybean cultivars from different breeding programs contained large linkage groups (essentially whole chromosomes) that were inherited intact from a particular parent. The authors suggested that this conserved transmission is due to selection, which is maintaining chromosome regions containing favourable combinations of alleles, or selecting for suppressed recombination in certain chromosomes to avoid a break-up of linkages in these favourable regions. Characterizing the process of linkage block formation/allele fixation within a breeding program can provide valuable knowledge with respect to the specific genetic changes that have occurred in the development of elite cultivars. One strategy for characterizing the genetic changes in the development of elite cultivars is to genotype the members of the pedigrees that comprise elite cultivars. As pedigrees represent a record of breeder manipulations (Shoemaker et al. 1992), genotyping not only the ancestors of commercially successful cultivars, but also elite cultivars developed from these landmark cultivars can be of great value in identifying genomic regions of importance. As selection (both artificial and natural) purges unfavorable alleles while maintaining favourable allele combinations in the form of linkage blocks, the net result is further reduction in the genetic diversity within elite germplasm.

From a molecular breeding perspective, it is of interest to track the inheritance of alleles in pedigrees to investigate the effects of long-term breeder selection within a breeding program.cultivars Sjakste et al. (2003) demonstrated the utility of using graphical genotypes (GGT) in investigating various aspects of the breeding process in Latvian barley cultivars, especially in the identification and transmission of conserved linkage blocks. Genotyping of pedigrees was also used as a strategy to track the origin of susceptibility and resistance to the fungal pathogen *Verticillium dahliae* in North American potato cultivars (Simko et al. 2004). While there have been studies done using molecular markers to track chromosome transmission in soybean pedigrees (Lorenzen et al. 1995, 1996), no such study has been done in elite germplasm adapted to the specific growing environments of Ontario, Canada.

OAC Bayfield (Tanner et al. 1998) represents a landmark cultivar for soybean growers in Ontario. Developed by the University of Guelph soybean breeding program from a cross performed in 1985, it was commercially released in 1994. At its peak in 1998, it was grown on over 400,000 acres which represented ~20 % of the total soybean acreage in Ontario for that year. From 1994 to 2004, the estimated economic value of OAC Bayfield to the Ontario economy is in excess of $750 million [Agricultural Research Institute of Ontario (ARIO) 1998]. This value is considerably higher given its role as a parent/grand parent in a number of commercially successful cultivars such as OAC Wallace, OAC Champion OAC Kent and OAC Drayton. The use of OAC Bayfield as a parent is not limited to cultivars developed from the University of Guelph. Through germplasm exchange, it has been used in independent breeding programs for cultivar development purposes. Given the overall impact OAC Bayfield has had on commercial soybean breeding in Ontario, the objective of this study was to conduct a characterization of both genetic diversity and chromosome microsatellite allele composition changes through multiple generations of the breeding process, involving OAC Bayfield's pedigree.

Materials and methods

Plant material

The use of OAC Bayfield for crossing in multiple independent soybean breeding programs in Canada allowed for not only a vertical, but also horizontal investigation into allele inheritance through the pedigree (Fig. 1). Seed for the genotypes comprising the pedigree of OAC Bayfield was obtained from various organizations/breeding programs depending on the source of the material (Table 1). For ancestral genotypes (i.e., pre-OAC Bayfield), accessions were obtained from either the Plant Gene Resources of Canada (PGRC) in Saskatoon, SK, Canada or the United States Department of Agriculture (USDA) soybean germplasm collection in Urbana-Champaign, IL, USA. Seeds for the remaining genotypes were either collected directly from the University of Guelph soybean breeding program or received from various collaborating soybean breeding programs, which included; Pioneer Hi-Bred, Ridgetown College, La Coop fédérée, Semences Prograin and Agriculture and Agrifood Canada (AAFC). Two sibling cultivars of OAC Bayfield, OAC Salem and OAC Brussels, were also included for comparison purposes with regards to genotype profiling.

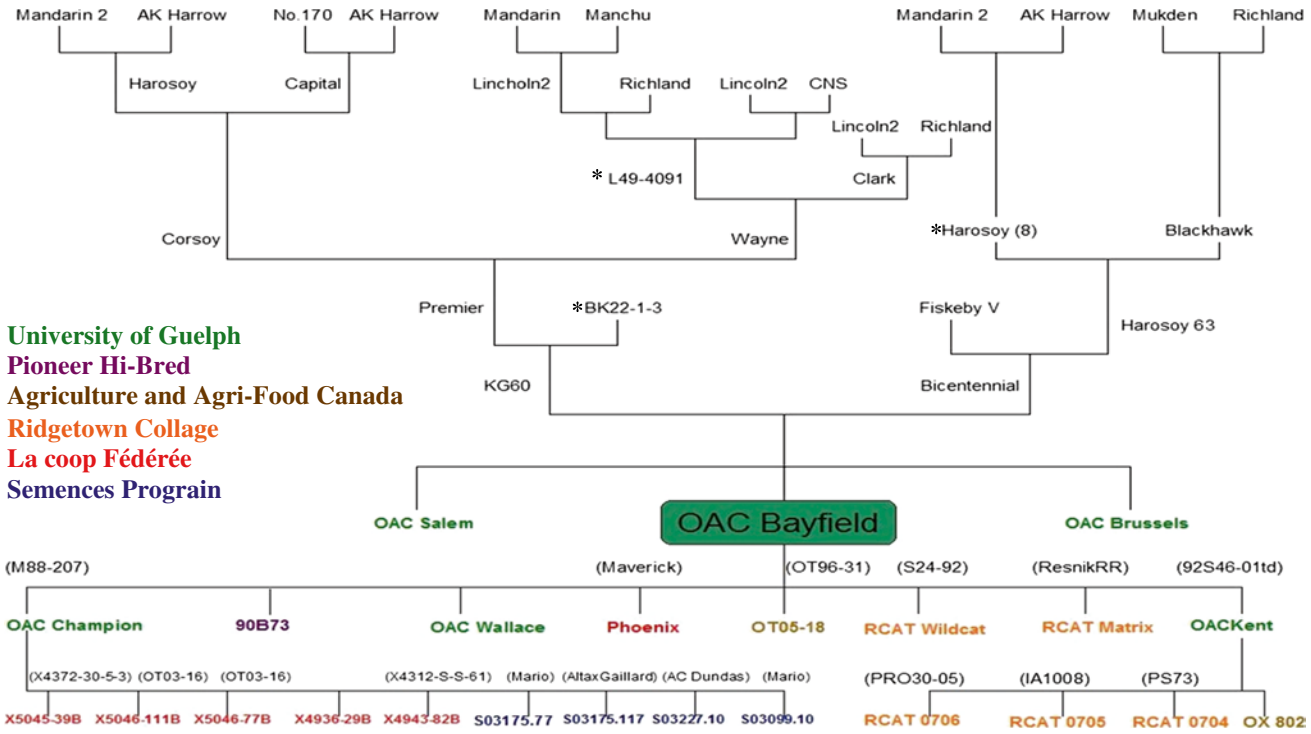


Fig. 1 Pedigree diagram of OAC Bayfield depicting the ancestors of OAC Bayfield as well as cultivars/breeding lines that have been derived from it. For the OAC Bayfield derived genotypes, the *colored names* correspond to the breeding program that developed it. The cultivars/breeding lines in *brackets* are the alternate parent that used in the cross for a given genotype used in the study. The cultivars with an *asterisk* indicate *lines* that were unavailable for study

DNA extraction and genotyping

DNA was extracted from ~15–20 freeze-dried leaf punches or from seed using the Sigma Genelute™ Plant genomic DNA extraction kit (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's protocol. Template DNA concentration and purity was measured with a ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE) and was standardized to 5 ng/μl for use in PCR reactions. Simple sequence repeat (SSR) markers were selected based on the composite linkage map (Song et al. 2004) in SoyBase (http://www.SoyBase.org), with genome-wide genotyping being performed at a density of approximately one SSR marker per 10 cM for each chromosome. PCR amplification was performed based on the Schuelke method (Schuelke 2000), with the final PCR cocktail consisting of the following: 3 μl of 20 % Trehalose, 4.06 μl of molecular grade H_2O, 1.5 μl of 10X PCR buffer, 1.5 μl of 25 mM $MgCl_2$, 1.0 μl of 3 mM dNTP mix, 0.12 μl of 4 μM M13 tailed forward primer, 0.48 μl of reverse primer, 0.48 μl of 4 μM "universal" M13 primer labeled with either 6FAM, VIC, NED or PET fluorescent dyes (Applied Biosystems, Foster City, CA), 0.4 μl of 2.5 U/μl Sigma Jumpstart™ taq polymerase and 3 μl of template DNA for a total reaction volume of 15 μl. Amplification reactions were performed using Stratagene thermocyclers (Agilent Technologies Inc, Santa Clara, CA) with the following cycling profile; an initial denaturation at 94 °C for 5 min followed by a two-step cycling profile, with 30 cycles of 94 °C for 30 s, 56 °C for 45 s and 72 °C for 45 s followed by 8 cycles of 94 °C for 30 s, 53 °C for 45 s and 72 °C for 45 s with a final extension at 72 °C for 10 min. Completed PCR products were pool-plexed to combine up to four SSR markers at a time for fragment analysis conducted at the University of Guelph Advanced Analysis Centers Genomics Facility using the ABI 3730 DNA Analyzer (Applied Biosystems). The size standard 500-LIZ was used as the internal size standard and PCR fragment sizes were determined using the software program GeneMarker® (Softgenetics, State Collage, PA), using the local southern sizing algorithm.

Data Analysis

Molecular genotype profiles for all lines were visualized as GGT using the software program GGT 2.0 (Berloo 2008). Designation of alleles was done in a chronological fashion with the foundation cultivars being genotyped first and subsequent lines being genotyped accordingly. The cultivar Mandarin-2 (also referred to as Mandarin (Ottawa) in the literature) was arbitrarily chosen as the reference genotype

Table 1 Description of germplasm comprising the pedigree of OAC Bayfield

Cultivar	Source/accession number	Maturity group
Mandarin[a]		
Mandarin 2	PI 297532	0
AK Harrow	PI 548298	III
No.171[a]		
Manchu	PI 548365	III
Mukden	PI 548391	II
Richland	PI 548406	II
Harosoy	PI 548573	II
Capital	CN 33259	II
Lincoln 2	PI 548362	III
CNS	PI 548445	VII
Harosoy (8)[a]	`	
Blackhawk	PI 548516	I
Clark	PI 548533	IV
L49-4091[a]		
Corsoy	CN 42220	II
Wayne	PI 548628	III
Harosoy 63	CN 33248	II
Fiskeby V	CN 33256	000
BK22-1-3[a]		
Premier	CN 35596	II
KG60	PI 556785	0
Bicentennial	CN 42824	0
OAC Salem	University of Guelph	00
OAC Bayfield	University of Guelph	0
OAC Brussels	University of Guelph	0
OAC Champion	University of Guelph	0
OAC Kent	University of Guelph	II
OAC Wallace	University of Guelph	0
RCAT Wildcat	Ridgetown College	I
RCAT MatRix	Ridgetown College	I
RCAT 0706	Ridgetown College	II
RCAT 0705	Ridgetown College	II
RCAT 0704	Ridgetown College	II
S03175.77	Semences Prograin	0
S03175.117	Semences Prograin	0
S03227.10	Semences Prograin	0
S03099.10	Semences Prograin	0
OT05-18	Agriculture Canada	00
OX-802	Agriculture Canada	II
90B73	Pioneer Hi-Bred	0
Phoenix	La Coop fédérée	00
X5045-1-S1-S1-39-B	La Coop fédérée	00
X5046-1-S1-S1-111-B	La Coop fédérée	00
X5046-1-S1-S1-77-B	La Coop fédérée	00
X4936-1-S1-S1-1-29-B	La Coop fédérée	00
X4943-1-S1-S1-1-82-B	La Coop fédérée	00

[a] Cultivar/line unavailable for study

when creating the microsatellite allelic series for each SSR marker. Thus, the genotype profiles of the other lines used in the study are relative to Mandarin-2. For classifying a chromosomal region as a linkage block, the criteria of Lorenzen et al. (1996) were used, where three consecutive informative loci derived from a particular ancestor being inherited as a unit for a minimum of two generations constitute a linkage block. Thus, linkage blocks were considered to be regions of the chromosome in which consecutive markers that were able to discriminate among alleles at particular loci were inherited as a unit, regardless of the linkage (cM) distance between markers. For single generation inheritance, four consecutive informative loci are required for a region to be classified as a linkage block. This software was also used to create a genetic distance matrix based on the simple matching coefficient between cultivars (Mohammadi and Prasanna 2003). This matrix was used to construct a neighbor-joining phylogenetic tree to assess the congruence of the genotype data to the known pedigree record as well as quantify the genetic distance between members of the pedigree. Summary statistics related to genetic diversity such as allele number per marker, allele frequency and polymorphism information content (PIC) were calculated using PowerMarker v3.25 (Lui and Muse 2005). To test for non-random Mendelian inheritance between a pair of parents and their progeny, a Chi square analysis was done using marker alleles that were polymorphic between a given set of parental genotypes.

Results

Genetic diversity

Genotyping with the 161 markers produced a total of 644 alleles, with the number of alleles per marker varying from 1 to 9, resulting in an overall average of 4 alleles/marker (Tables 1, 2). The average number of alleles per locus on a chromosome basis (min 3.1, max 4.9) was consistent with the overall average. Thus, there were no chromosomes with extreme diversity loss or excess on an average basis. However, out of the 161 markers, 14 had a major allele frequency >0.95, which corresponds to ~9 % of the sampled genome in which there was little or no allelic diversity found.

Polymorphism information content (PIC) values ranged from a low of zero to 0.847 (Sat 417), with larger values being of greater use when genotyping additional/unknown germplasm (e.g., for identification purposes), as these markers are more likely to be polymorphic among various soybean cultivars. Cluster analysis grouped the cultivars into six groups (Fig. 2), which in general, showed high congruence with the known pedigree record. The dendrogram revealed three distinct ancestral groups (i.e., pre-OAC Bayfield

Table 2 Summary statistics of microsatellite diversity in the pedigree of OAC Bayfield

Chromosome	Marker	Major allele frequency	Allele number	PIC[b]
1	Satt184	0.45	4.0	0.585
	Sat353	0.32	5.0	0.639
	Satt531	0.88	2.0	0.187
	Satt221	0.89	3.0	0.181
	Satt548	0.76	2.0	0.292
	Satt370[a]	0.97	2.0	0.047
	Satt436	0.41	4.0	0.654
	Satt129	0.61	3.0	0.443
Average			3.1	
2	Sat227	0.41	7.0	0.717
	Satt157	0.28	9.0	0.780
	Satt558	0.78	3.0	0.300
	Satt296	0.90	2.0	0.157
	Satt266	0.56	2.0	0.371
	Satt428	0.48	4.0	0.591
	Satt703	0.65	3.0	0.447
	Satt274	0.82	4.0	0.285
Average			4.2	
3	Sat379[a]	0.97	2.0	0.047
	Satt152	0.50	7.0	0.629
	Satt530	0.56	6.0	0.595
	Satt125	0.58	2.0	0.366
	Satt387	0.79	3.0	0.308
	Satt339	0.60	4.0	0.503
	Satt234[a]	1.00	1.0	0.000
	Satt22[a]	1.00	1.0	0.000
Average			3.2	
4	Satt565	0.68	4.0	0.458
	Soygpatr	0.80	3.0	0.297
	Satt396	0.92	2.0	0.129
	Satt578	0.73	2.0	0.311
	Satt136	0.45	5.0	0.541
	Satt670	0.47	5.0	0.567
	Satt524	0.88	2.0	0.187
	Satt164	0.53	5.0	0.534
Average			3.5	
5	Satt684	0.39	4.0	0.632
	Satt382	0.47	7.0	0.628
	Satt050	0.73	3.0	0.359
	Satt385	0.44	5.0	0.632
	Satt174[a]	0.97	2.0	0.045
Average			4.8	
6	Satt681	0.67	6.0	0.464
	Satt227	0.94	2.0	0.097
	Satt281	0.59	6.0	0.569
	Sat336	0.41	6.0	0.678
	Satt322	0.60	2.0	0.363
	Satt643	0.40	7.0	0.697
	Satt277	0.40	5.0	0.621

Table 2 continued

Chromosome	Marker	Major allele frequency	Allele number	PIC[b]
	Satt319	0.47	4.0	0.564
	Satt202	0.47	4.0	0.631
	Satt371	0.37	5.0	0.699
	Satt357	0.69	4.0	0.439
Average			4.6	
7	Satt636	0.66	4.0	0.461
	Satt201	0.90	3.0	0.166
	Satt150	0.92	2.0	0.129
	Satt567	0.57	3.0	0.491
	Satt463	0.46	6.0	0.637
	Satt323	0.61	3.0	0.437
	Satt494[a]	1.00	1.0	0.000
	Satt306	0.90	3.0	0.166
	Satt551	0.70	3.0	0.381
	Satt210	0.66	4.0	0.449
Average			3.5	
8	Satt390	0.58	3.0	0.399
	Sat406	0.51	7.0	0.615
	Satt589	0.46	4.0	0.604
	Satt315[a]	1.00	1.0	0.000
	Satt424	0.86	4.0	0.237
	Satt437	0.41	5.0	0.651
	Satt470[a]	1.00	1.0	0.000
	Satt409	0.48	6.0	0.536
	Satt538[a]	0.95	2.0	0.086
	Satt378[a]	0.97	2.0	0.054
Average			3.5	
9	Satt539	0.51	2.0	0.374
	Satt242	0.66	4.0	0.478
	Satt196	0.47	5.0	0.567
	Satt588	0.51	4.0	0.455
Average			4.0	
10	Satt358	0.76	4.0	0.353
	Satt500	0.64	5.0	0.512
	Sat318	0.85	4.0	0.247
	Satt259	0.65	5.0	0.501
	Satt173	0.36	7.0	0.722
	Sat341	0.63	5.0	0.477
	Satt592	0.73	3.0	0.335
	Sat109	0.66	6.0	0.486
Average			4.8	
11	Satt426	0.43	4.0	0.572
	Satt251	0.55	3.0	0.406
	Satt197	0.43	5.0	0.596
	Satt519	0.75	3.0	0.360
	Satt597	0.45	3.0	0.523
	Satt332	0.83	4.0	0.269
	Satt665	0.50	5.0	0.526
	Satt359	0.44	3.0	0.521

Table 2 continued

Chromosome	Marker	Major allele frequency	Allele number	PIC[b]
Average			3.7	
12	Satt353	0.87	4.0	0.217
	Satt568	0.69	3.0	0.428
	Satt009	0.73	2.0	0.312
	Satt442	0.30	6.0	0.750
	Satt469	0.87	3.0	0.213
	Satt222	0.92	2.0	0.126
	Satt302	0.45	3.0	0.511
	Satt181	0.63	4.0	0.414
Average			3.3	
13	Satt030	0.33	4.0	0.662
	Satt269	0.55	3.0	0.528
	Satt423	0.88	4.0	0.208
	Satt160	0.42	5.0	0.572
	Satt516	0.35	5.0	0.682
	Satt663	0.48	5.0	0.609
	Satt335	0.73	2.0	0.311
	Sat375	0.38	8.0	0.723
	Satt490	0.47	3.0	0.474
	Satt554	0.57	6.0	0.589
	AW756935	0.92	2.0	0.134
	Satt395[a]	0.97	2.0	0.049
Average			4.4	
14	Satt577	0.63	5.0	0.538
	Satt467	0.95	2.0	0.086
	Satt304	0.87	2.0	0.194
	Satt474	0.51	4.0	0.488
	Satt534	0.52	6.0	0.604
	Satt560	0.70	7.0	0.465
Average			4.3	
15	Satt213[a]	1.00	1.0	0.000
	Satt411	0.73	3.0	0.335
	Satt720	0.45	5.0	0.666
	Satt212	0.63	2.0	0.356
	Sat107	0.25	8.0	0.810
	Satt369	0.54	4.0	0.540
	Satt553	0.37	9.0	0.760
	Satt230	0.61	4.0	0.502
Average			4.5	
16	Satt249	0.73	4.0	0.392
	Satt693	0.33	4.0	0.636
	Satt529	0.57	3.0	0.402
	Satt215	0.46	4.0	0.569
	Satt244	0.57	6.0	0.493
	Satt431	0.71	3.0	0.349
	Satt712	0.70	2.0	0.329
Average			3.7	
17	Satt328[a]	1.00	1.0	0.000
	Satt458	0.42	8.0	0.668
	Satt154	0.30	5.0	0.668

Table 2 continued

Chromosome	Marker	Major allele frequency	Allele number	PIC[b]
	Satt208	0.60	2.0	0.364
	Satt389	0.41	6.0	0.650
	Satt311	0.51	4.0	0.447
	Satt464	0.87	2.0	0.198
	Satt186	0.40	5.0	0.640
	Satt386	0.70	3.0	0.358
Average			4.0	
18	Satt235	0.78	3.0	0.330
	Satt324	0.57	4.0	0.514
	Satt394	0.74	3.0	0.333
	Satt199	0.66	2.0	0.345
	AF162283	0.25	5.0	0.718
	Satt191	0.47	6.0	0.657
Average			4.1	
19	Satt238[a]	1.00	1.0	0.000
	Satt523	0.67	4.0	0.446
	Satt284	0.43	4.0	0.597
	Satt481	0.64	3.0	0.460
	Satt166	0.65	4.0	0.461
	Satt561	0.71	2.0	0.323
	Satt373	0.39	8.0	0.669
Average			3.7	
20	Satt571	0.60	5.0	0.497
	Satt367	0.56	5.0	0.572
	Satt700	0.80	3.0	0.302
	Sat104	0.57	2.0	0.369
	Sat419	0.48	6.0	0.633
	Satt440	0.55	4.0	0.446
Average			4.1	
	Grand Mean	0.62	3.9	0.435

[a] Major allele Frequency >0.95

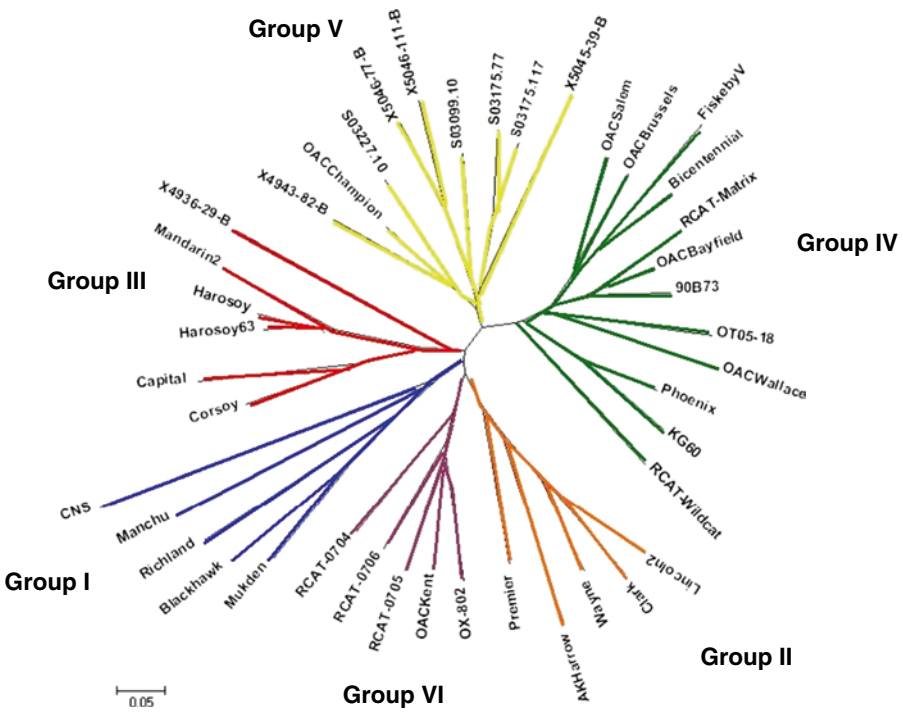
[b] Polymorphism information content

genotypes) that are separate from each other as well as the current cultivars. Ancestral Group I consisted of Mukden, Richland, Manchu, Blackhawk and CNS. Ancestral Group II consisted of AK Harrow, Lincoln2, Clark and Wayne. The final Ancestral Group (III) consisted of Mandarin-2, Harosoy, Harosoy 63, Capital, Corsoy and one of the lines from La Coop fédérée, X4936-29-B. This was the major exception to the overall congruence of the dendrogram to the pedigree record and reasons for this line clustering to an ancestral group are either an erroneous sample or a large contribution of ancestral alleles in this particular line. The OAC Bayfield group (Group IV) consisted of its parents (KG 60 and Bicentennial), its siblings (OAC Salem and OAC Brussels), the progeny derived from OAC Bayfield and Fiskeby-V. The remaining two groups (Group V and VI) are composed of the progeny that was derived from a particular parental cultivar (i.e., either OAC Champion or OAC Kent). Pairwise genetic distance between cultivars ranged from 0.09 between Harosoy and Harosoy 63 to 0.73 between CNS and Clark. CNS, in fact, was the most genetically distant cultivar with an average genetic distance value to all other members of the pedigree of 0.67 (Provided as supplementary data).

Genotyping through the pedigree also reveal a number of rare/unique alleles that were either found in the ancestral genotypes or lines from collaborating breeding programs (Fig. 3). Of the ancestral genotypes, CNS contained the highest number of rare/unique alleles (31) followed by Manchu (13) and Fiskeby-V (10). Line X0545-39B from La Coop fédérée had considerably more unique alleles than any other line derived from OAC Bayfield with 14 unique alleles.

Parental inheritance

Of the parent-offspring combinations that could be tested with a Chi square analysis, Premier, Bicentennial and OAC Brussels all showed significant deviations from the expected ratio of equal parental contribution (Table 3).



Group I : Ancestral group 1
Group II : Ancestral group 2
Group III : Ancestral group 3
Group IV : OAC Bayfield
Group V : OAC Champion
Group VI : OAC Kent

Fig. 2 Neighbour-joining dendrogram of genetic relatedness based on SSR genotyping of cultivars comprising the pedigree of OAC Bayfield

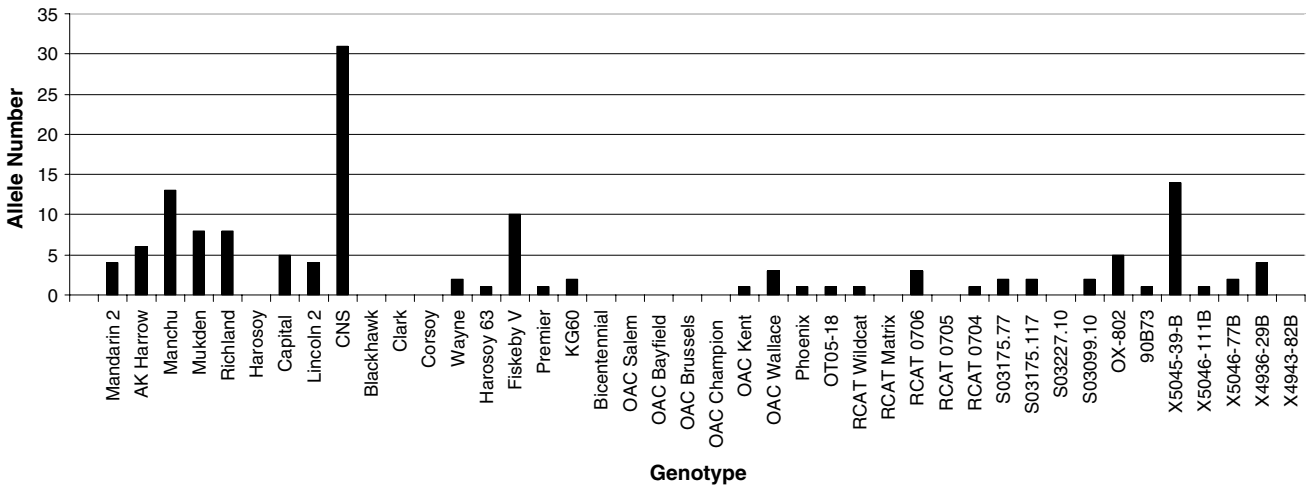


Fig. 3 Histogram of unique alleles within the members of OAC Bayfield's pedigree. Genotypes are ordered chronologically from *left* to *right*

Of particular interest was the significant over-representation of alleles from Fiskeby-V in Bicentennial. This, in turn, is seen in OAC Brussels with the same over-representation from Bicentennial. Of the 57 alleles inherited from Fiskeby-V in Bicentennial, 49 were transmitted from Bicentennial to OAC Brussels. While these alleles came from various regions of the genome, certain chromosomes had large regions in which a group of alleles were transmitted as a unit or a chromosomal region. Approximately 47 % (23 out of 49) of the Fiskeby-V alleles found in OAC

Table 3 Chi square test of deviations from expected genomic contribution of parent cultivars to progeny in the pedigree of OAC Bayfield using polymorphic SSR markers between the parents

Cultivar	Number of loci	Parent 1 (Number of loci)	Parent 2 (Number of loci)	χ^2	*P* Value
Harosoy**	87	Mandarin-2 (67)	AK Harrow (20)	0.19	NS
Blackhawk	78	Mukden (45)	Richland (33)	1.85	NS
Clark**	80	Lincoln 2 (64)	Richland (16)	1.07	NS
Premier	57	Corsoy (14)	Wayne (43)	14.75	<0.01*
Bicentennial	86	Harosoy 63 (29)	Fiskeby-V (57)	9.12	<0.01*
OAC Salem	63	KG 60 (28)	Bicentennial (35)	0.78	NS
OAC Brussels	63	KG 60 (15)	Bicentennial (48)	17.29	<0.01*
OAC Bayfield	66	KG 60 (30)	Bicentennial (36)	0.71	NS

Expected ratio is 50:50 except for those indicated with double asterisks

NS Non significant

* Significant at $\alpha = 0.01$

** One backcross (expected ratio is 75:25)

Brussels were inherited as large chromosomal regions from Chromosomes 6, 8, 11 and 18. The other two siblings (OAC Bayfield and OAC Salem) had a more balanced set of alleles from their parent cultivars KG 60 and Bicentennial.

Microsatellite allele transmission to OAC Bayfield

A graphical genotype profile of each of the 20 chromosomes characterizing the specific transmission of microsatellite alleles/allele combinations through the pedigree of OAC Bayfield is provided in Fig. 4. A striking feature of the transmission pattern through the pedigree is the high level of allelic structure that is built up in particular chromosomes which is conservatively transmitted, with a lack of such structure in others. The allelic structure is primarily in the form of linkage blocks that fall into two general categories: ancestral linkage blocks that are combined through recombination with subsequent conserved transmission, and novel linkage groups introduced from Fiskeby-V.

Chromosomes 1 (LG D1a) and 7 (LG M) provide examples of large linkage blocks derived through recombination of ancestral linkage blocks (Fig. 5). In the case of chromosome 1, there is a linkage block (40.9–61.0 cM) that is shared among a number of the base generation ancestors (AK Harrow, Manchu, Mukden and Richland). This region is conservatively transmitted along the pathway of Mukden/Richland-Blackhawk-Harosoy 63, where a recombination event with Fiskeby-V results in a highly structured chromosome that is transmitted from Bicentennial and is shared among OAC Bayfield, OAC Salem and OAC Brussels. Interestingly, the chromosome composition is nearly identical to that of AK Harrow, a foundation ancestor.

As with chromosome 1, chromosome 7 has a key recombination event between the upper chromosomal region of Harosoy 63 (5.0–50.1 cM) and the lower chromosomal region of Fiskeby-V (60.0–112.1 cM). The subsequent allele composition of the chromosome is transmitted without change through Bicentennial and its progeny, OAC Bayfield, OAC Salem and OAC Brussels.

Chromosomes 8 (LG A2) and 16 (LG J) are examples where novel linkage groups that trace back to Fiskeby-V are conserved among OAC Bayfield, OAC Salem and OAC Brussels (Fig. 6). The uppermost region (9.1–34.0 cM) in chromosome 8 and an 11 cM region (33.9–44.1 cM) in chromosome 16 are both inherited as distinct linkage blocks that are shared among OAC Bayfield and its siblings.

Microsatellite allele transmission from OAC Bayfield

In tracking the chromosomal inheritance from OAC Bayfield for another two generations, two situations emerged: first, the chromosome allelic structure built up in OAC Bayfield remains highly conserved in particular cultivars for particular chromosomes. Second, the transmission of conserved allelic structure found in OAC Bayfield was minimal in the development of other cultivars. Among the eight cultivars with OAC Bayfield as an immediate parent, Chromosomes 1 and 16 exhibited the greatest amount of retained chromosome structure that had a high frequency of conserved transmission (Fig. 7). For both these chromosomes, six of the eight progeny (OAC Wallace, Phoenix, RCAT Wildcat, RCAT Matrix, OT05-18 and 90B73) share a highly conserved chromosome composition, and in the case of chromosome 1, all six progeny cultivars have an identical microsatellite allele composition to that of OAC Bayfield. For both OAC Champion and OAC Kent, the allelic transmission profiles were markedly different than the other six cultivars derived from OAC Bayfield. With OAC Champion, Chromosomes 3, 4, 5, 16 and 19 all have the same allelic

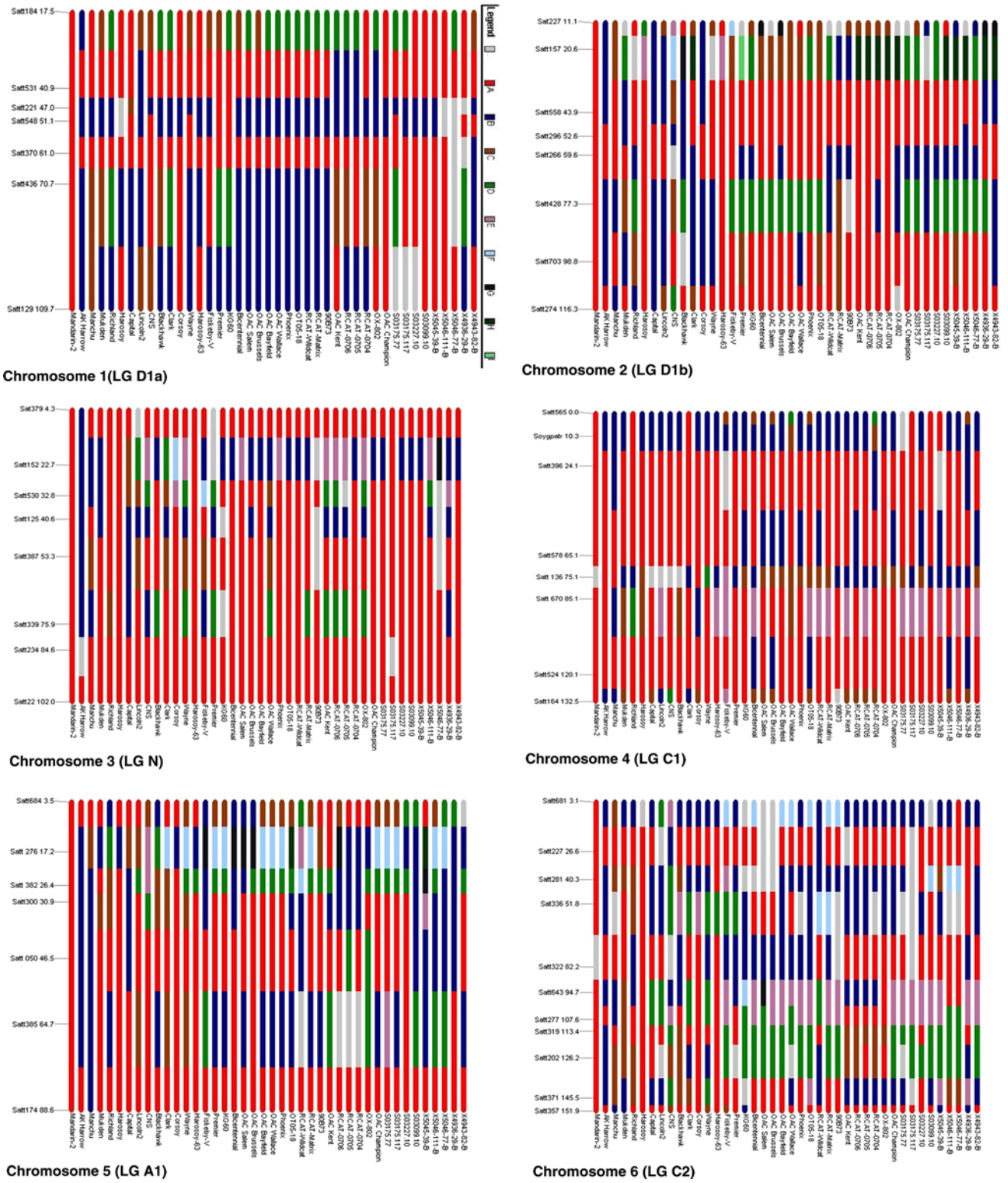
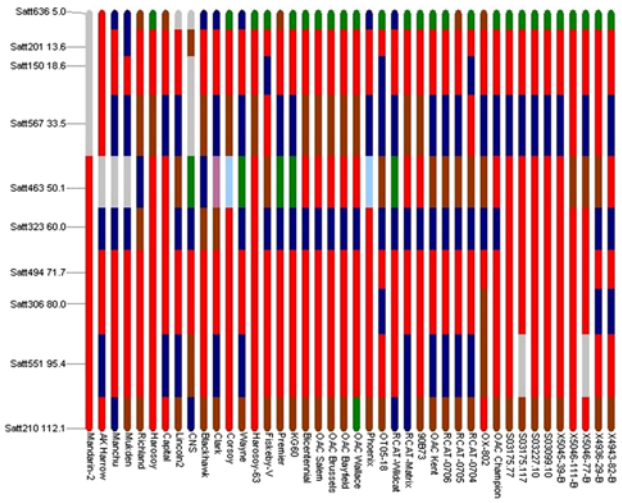
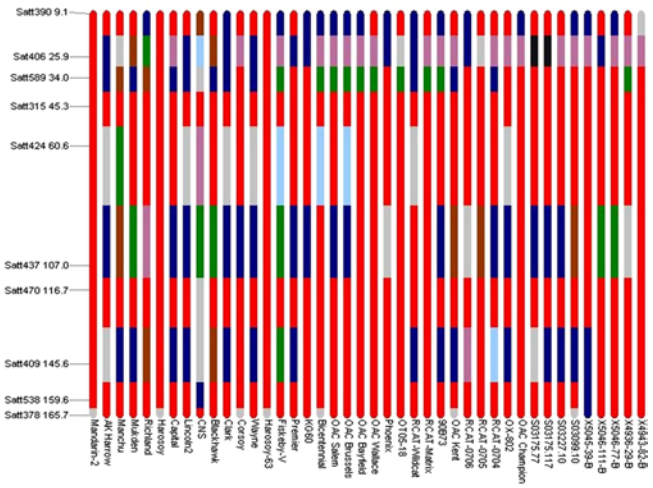


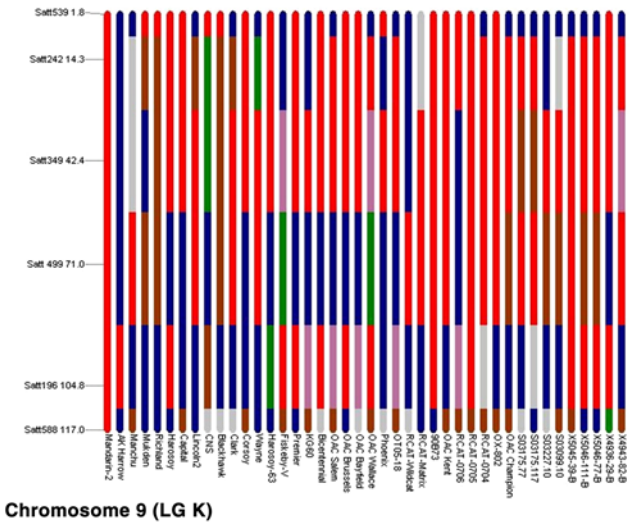
Fig. 4 Graphical genotyping of chromosomal transmission through the pedigree of OAC Bayfield. For each chromosome, the microsatellite allele series at each SSR marker position (cM) is given with each unique allele being assigned a particular color. The pedigree members are placed in chronological order up to OAC Bayfield, with post-OAC Bayfield cultivars/lines placed according to the pedigree. Genotype profiles are relative to Mandarin-2, with individuals with the same allele designation sharing the *same color*, with *grey* indicating missing values



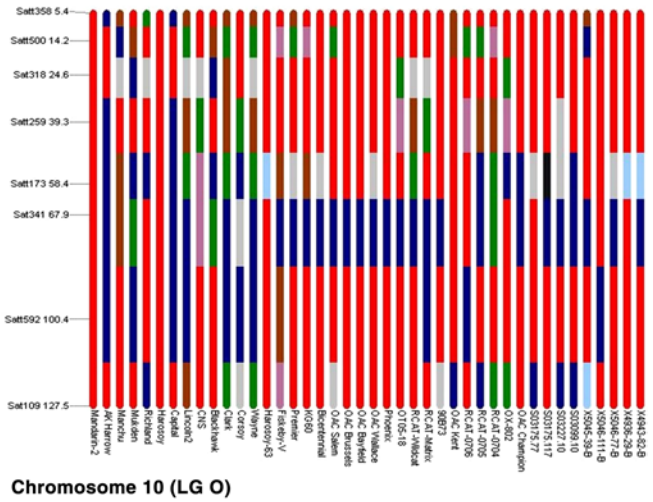
Chromosome 7 (LG M)



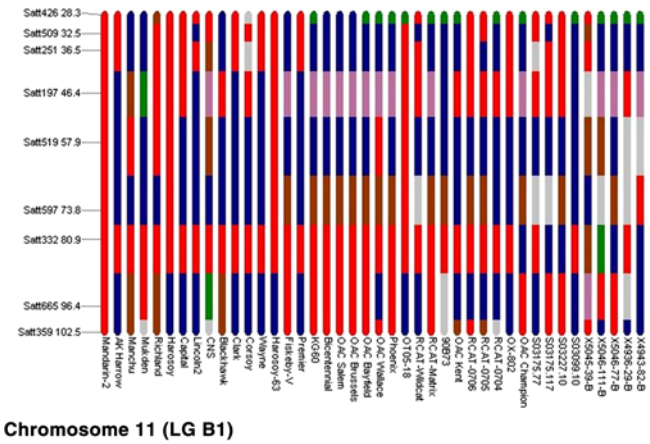
Chromosome 8 (LG A2)



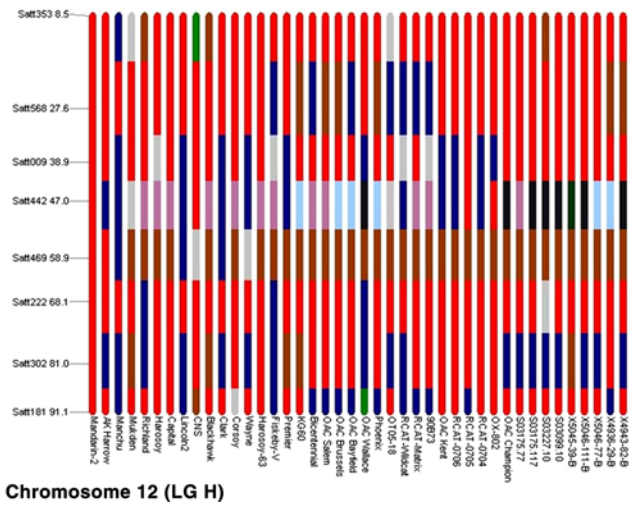
Chromosome 9 (LG K)



Chromosome 10 (LG O)



Chromosome 11 (LG B1)



Chromosome 12 (LG H)

Fig. 4 continued

Fig. 4 continued

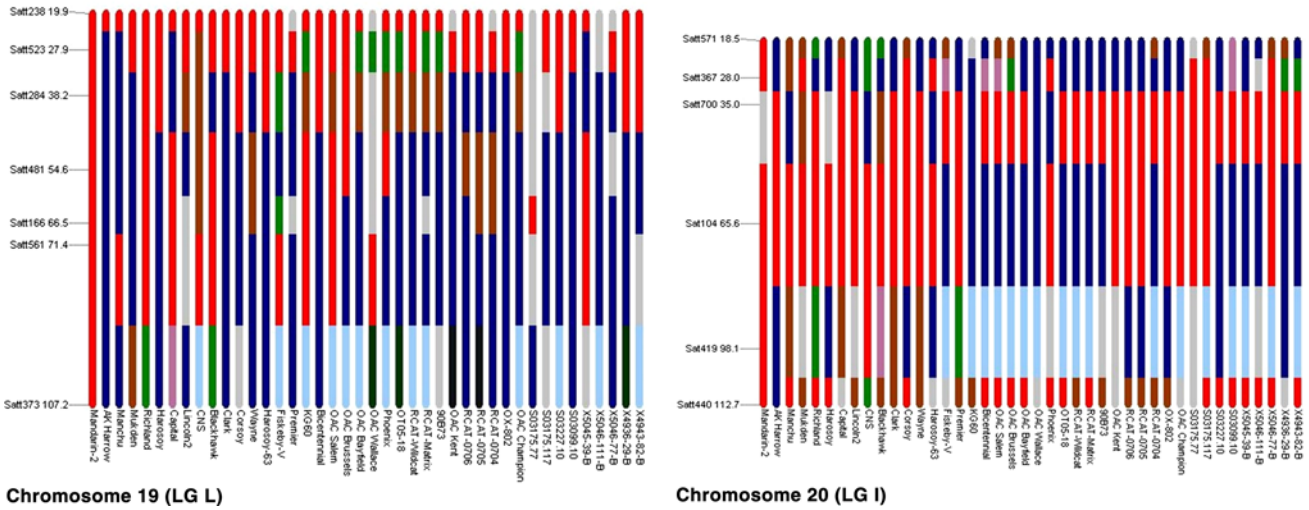


Fig. 4 continued

composition for the entire chromosome as OAC Bayfield (Fig. 8). With OAC Kent, there were no such chromosomes and in fact there were alternate marker alleles at a number of the loci (72/161) when compared to OAC Bayfield, indicating less of an impact of OAC Bayfield with respect to allelic transmission than the other seven cultivars.

In addition, it was possible to track chromosome transmission for a second generation (i.e., generation for which OAC Bayfield is a grandparent) for both OAC Champion and OAC Kent. Although a number of chromosomes shared identical allelic profiles between OAC Champion and OAC Bayfield, this structure was largely broken down in the next generation, except for Chromosomes 3, where 5/9 progeny share the identical allele profile to OAC Champion. In the case of OAC Kent, the differentiated allelic composition (with respect to OAC Bayfield) remains consistent among the lines derived from OAC Kent (RCAT-0704, RCAT-0705, RCAT-0706 and OX-802).

Although a number of chromosomes became differentiated (with regards to allelic profile) from OAC Bayfield over the two generations, there were regions that remained conserved among many of the lines. Chromosomes for which linkage blocks were conserved in a number of lines derived from OAC Bayfield (parent or grand parent) are the upper regions of Chromosomes 4 (13/21), Chromosome 7 (15/21) and Chromosome 11 (12/21) (Fig. 9).

Discussion

Pedigree genetic diversity

As genetic diversity is a necessity for continuous genetic improvement, it was of particular interest to quantify the level of diversity and characterize the nature of the diversity (recombination versus introgression) among the members of the pedigree. OAC Bayfield's pedigree exemplifies the well-known narrow gene pool that most North American soybean breeding programs are based on (Gizlice et al. 1994), with the same foundation ancestors (e.g., Mandarin-2, AK Harrow, Richland, Lincoln 2) occurring on both sides of the pedigree. The cluster analysis revealed as to just how over-represented alleles from certain ancestors are. The allelic diversity becomes reduced with Harosoy 63, as it is in Ancestral group III of the dendrogram, which contains genotypes found on the other side of the pedigree (see Fig. 1). That is, not only are Mandarin-2 and AK Harrow on both sides of the pedigree, but also their alleles are being over-represented through Harosoy 63. This is also supported by the genotypes making up Ancestral group I, which contains the most genetically distant genotypes due to a number of their alleles not being incorporated into future generations. This is especially true for the ancestor CNS. Both the genetic distance values and histogram of rare alleles reflect the small contribution of CNS alleles that were incorporated into future generations of the pedigree. This is not surprising as CNS is a group VII cultivar. There is a legacy effect with CNS as it is found in many North American soybean pedigrees due to it being a source of bacterial pustule resistance (Narvel et al. 2001). This highlights an important issue with regards to widening a genetic base; it is not simply how genetically diverse a germplasm source is, but whether the genetic diversity will be of value for agronomic improvement in a target environment. By genotyping through pedigrees, one can identify specific genomic regions that have been introgressed from particular ancestors when breeding for cultivars for a target environment.

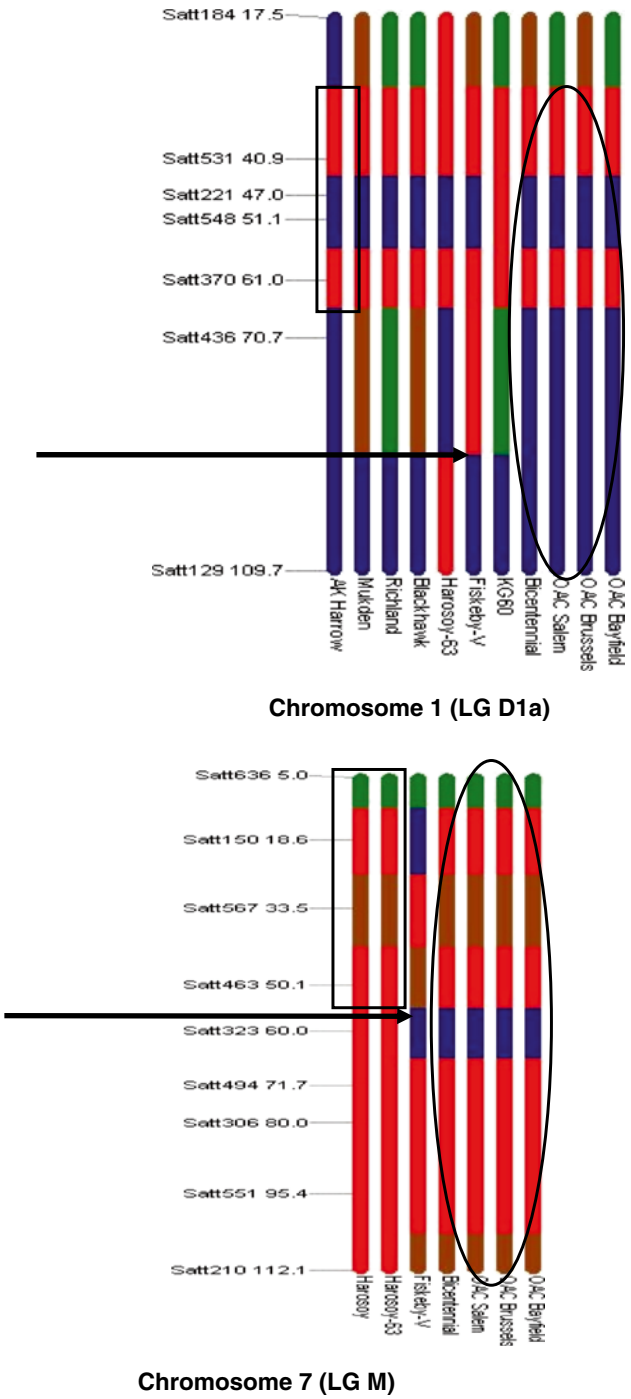


Fig. 5 Transmission of chromosomes 1 and 7 *highlighting* recombination events followed by conserved transmission. *Boxes* surrounding specific chromosomal regions indicate the ancestral linkage blocks, the *arrow* indicates the recombination event and the *oval* indicates new conservatively transmitted region shared by various cultivars

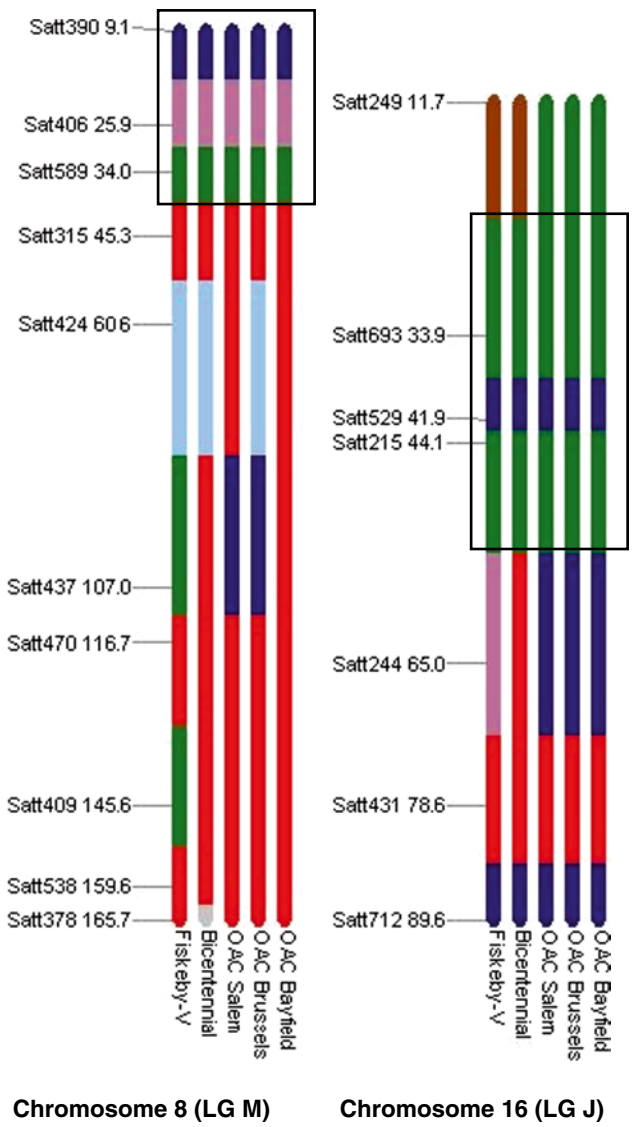


Fig. 6 Linkage blocks which trace back to Fiskeby-V. *Boxes* indicate the specific chromosomal regions are a unique to Fiskeby-V

Impact of Fiskeby-V

Fiskeby-V was a Swedish cultivar developed by Sven A. Holmberg in the 1950s, for the environment of Fiskeby, Sweden, which is many degrees above the most northerly latitudes in which soybean was known to be adapted to at the time (Soyinfo Center 2013). Both the dendrogram and Chi square analysis provide evidence that Fiskeby-V was a key ancestor in the development of OAC Bayfield, while the graphical genotyping indicated that alleles/linkage blocks from Fiskeby-V were a defining genotypic difference between MG 00-1 and MG II genotypes in the pedigree. The dendrogram showed Fiskeby-V clustering with the OAC Bayfield group indicating a greater number of alleles that are shared between Fiskeby-V and all other cultivars in the group. The Chi square analysis supports the result of the cluster analysis as there is a significant over-representation of Fiskeby-V alleles in Bicentennial, which would suggest that through selection, many beneficial

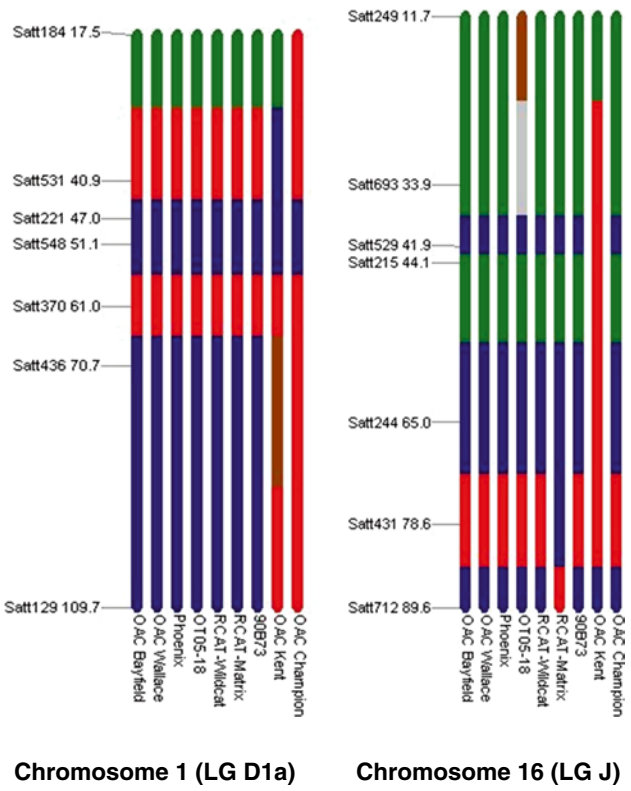


Fig. 7 Highly conserved allelic composition for chromosomes 1 and 16 when tracking microsatellite allele transmission from OAC Bayfield to particular progeny cultivars

alleles have been selectively mined from Fiskeby-V in the development of Bicentennial.

In a search of previously discovered QTL in SoyBase 2011 (http://www.SoyBase.org, 2011), there are both pod maturity (R3 1-2, R7 1-2) and flowering (Fflr 9-3) QTLs on chromosome 16 at ~43 cM of the composite linkage map. All of these QTL lie within the linkage blocks identified in these chromosomes that can be traced back to Fiskeby-V. Furthermore, the Fiskeby-V derived linkage block on chromosome 16 is absent in all of the MG II genotypes (OAC Kent, RCAT-0704, RCAT-0705 and RCAT-0706). These findings also support the study by Fu et al. (2007), who concluded that Canadian cultivars were more genetically related to accessions from Russia, Sweden and Ukraine than to Asian accessions.

Effects of long-term breeding on chromosome composition and transmission

Studies in many crop species have shown that allelic structure is built up through breeding via selection, which results in significant changes to the genome (Sjakste et al. 2003; Wright et al. 2005; Zhu et al. 2007). One change is in the assemblage of linkage blocks containing combinations of favourable alleles, which are then conservatively transmitted to subsequent generations (Lorenzen et al. 1996). An important component of this research was the inclusion of genotypes from multiple independent breeding programs. The genotypes developed represent selections from multiple populations derived from a number of biparental crosses, selected by multiple independent breeders. Therefore, each cultivar or breeding line represents the selection from thousands of progeny and a multitude of possible allele combinations. If breeder's selection was maintaining favourable allele combinations, then it should have been reflected in the genotype data. The results from this study support the latter concept as large linkage blocks on a number of chromosomes were shared among cultivars and breeding lines developed by six different breeding programs, which all have OAC Bayfield as a common parent or grand parent.

The alternative explanation for conserved linkage block inheritance is that they are the result of naturally low recombining regions of particular chromosomes. Ott et al. (2011) used microsatellite markers to investigate the physical distribution of recombination in soybean. They reported higher levels of recombination in the telomeric regions (due to higher gene densities) than in the centromeric and pericentromeric regions and that recombination rates varied greatly among chromosomes. Chromosome 16 had the lowest crossover frequency, whereas Chromosomes 1, 14 and 18 all exhibited low recombination toward the middle of the long chromosome arm (Ott et al. 2011). Conversely, Chromosome 7 had the highest crossover frequency (Ott et al. 2011). These findings are consistent with our findings, as both Chromosomes 1 and 16 had a high frequency of conserved linkage block transmission. Hence, naturally occurring low recombination rate could be the reason why Chromosomes 1 and 16 had linkage blocks that were conservatively inherited across multiple breeding programs. However, we observed substantial allelic structure and conserved linkage block transmission in Chromosome 7, which suggests an impact of breeder's selection, which is maintaining allelic structure given the high crossover frequency as expected based on previous reports (Ott et al. 2011). Further studies into the role that breeder's selection plays in repressing recombination where potentially favourable allele combinations exist are warranted.

The other process we were able to investigate involved the changes in allele frequency through various generations. Of particular interest was determining where allele fixation was occurring, as fixation in the gene rich areas of chromosomes brings up concerns with regards to future genetic gains due to massive erosion of genetic diversity. Performing graphical genotyping of pedigrees offers a means to visualize the process of allele fixation over time in terms of the regions that are being selected within elite

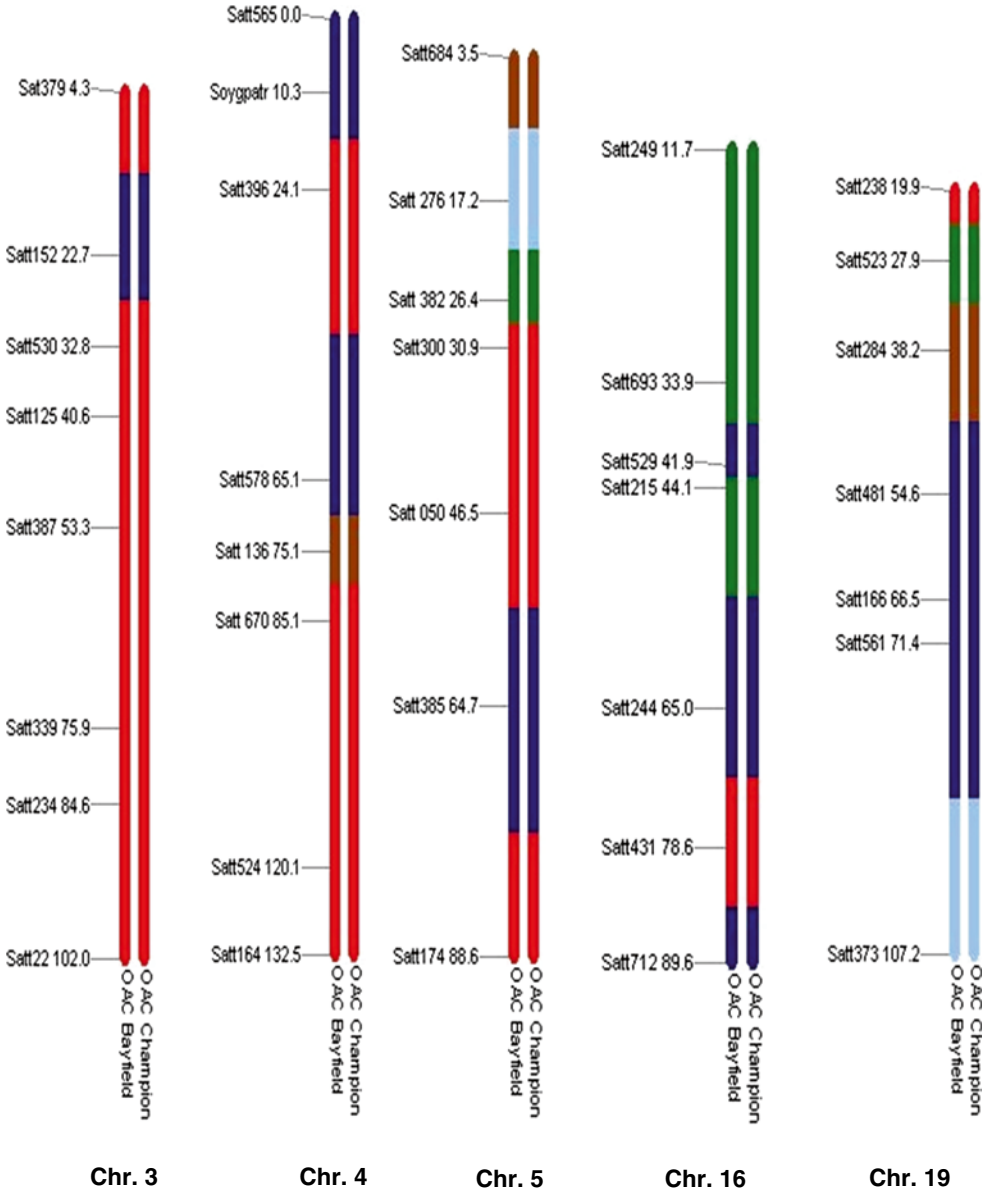


Fig. 8 Chromosomes for which the entire allelic composition is transmitted without change from OAC Bayfield to OAC Champion. SSR maker positions (cM) are given according to the Soybase composite map

germplasm of a breeding program developing cultivars for a target environment.

Conclusion

There is an inherent paradox with regards to favourable linkage blocks and the need to widen a genetic base. If these linkage blocks represent important genomic regions for cultivar development, then a high degree of fixation raises concerns over plateaus in modern soybean improvement (Hyten et al. 2006). However, by breaking apart these linkage blocks through recombination or introgression, a breeder runs the risk of taking steps "backwards" as performance of the cultivars is usually inferior to elite × elite crosses. This is a driving factor in the general breeding strategy of elite x elite crossing for many applied soybean breeding programs as the "cost-benefit" aspect must be addressed when expanding a genetic base for commercial breeding purposes. It is noted that 1 SSR marker per 10 cM is low-density coverage compared to current genotyping technologies (such as genotype by sequencing), and the use of these technologies will be valuable in characterizing linkage blocks in fine detail. However, it has also been shown that elite soybean germplasm exhibits extensive linkage disequilibrium in many areas which results in less markers being required to represent a given genomic region (Lam et al. 2010).

The practical utility of genotyping pedigrees of a breeding program is that it provides a framework in which

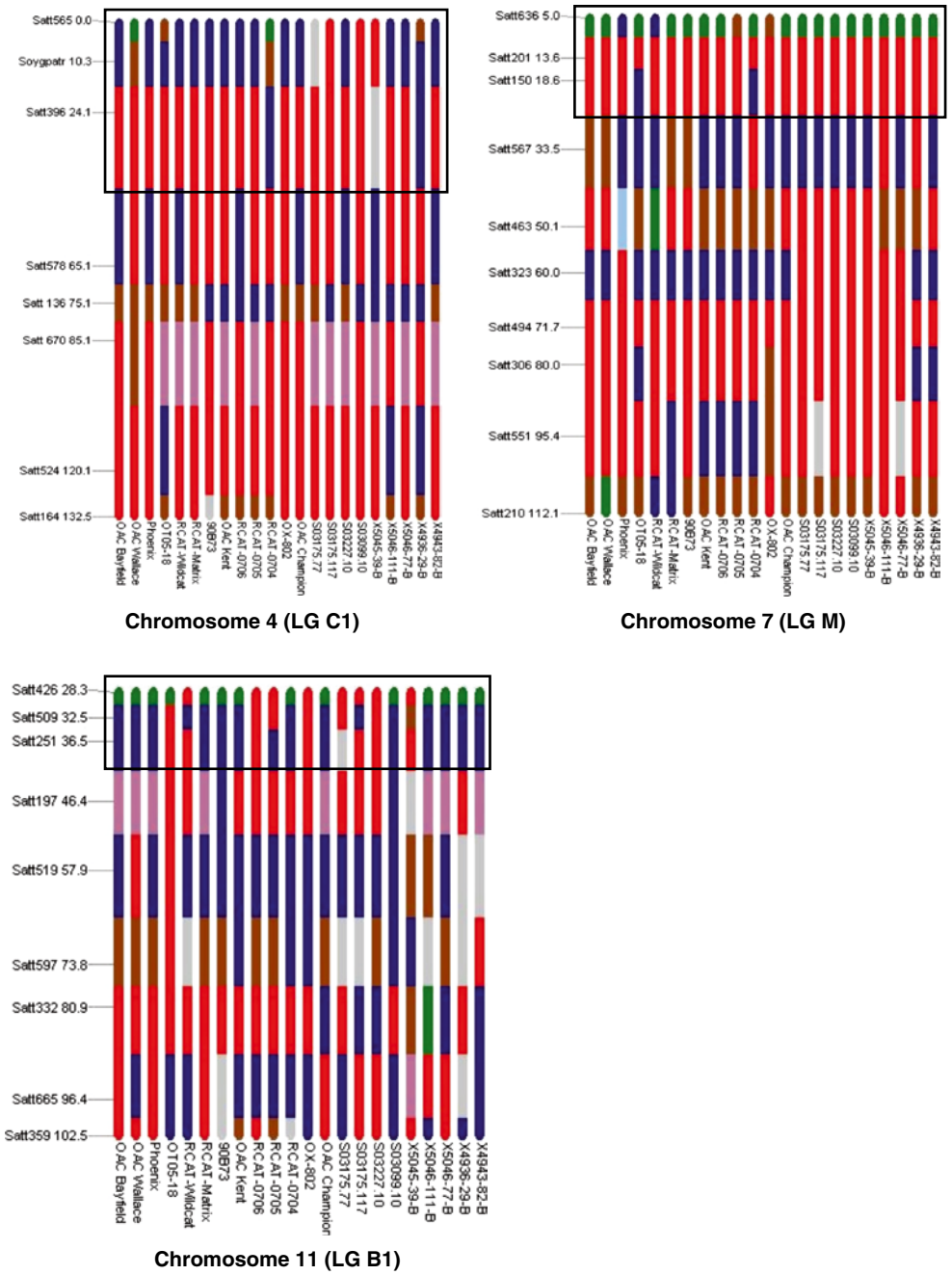


Fig. 9 Chromosomal regions exhibiting conserved transmission across various cultivars/breeding lines derived from OAC Bayfield. *Boxes* indicate the conserved region for each chromosome

genotype profiles allow a breeder to visualize the enrichment for particular alleles/allele combinations. A breeder can then use various markers to move a particular linkage block into different genetic backgrounds to determine phenotypic effects. Conversely, one could investigate the phenotypic effect when highly conserved linkage blocks are broken via recombination. Finally, by characterizing the molecular landscape of individual soybean breeding programs, a breeder can directly compare haplotype profiles of their breeding material to any other breeding material that has been genotyped/sequenced. Given the rapid changes in genomic technologies, combined with the reference genome sequence (Schmutz et al. 2010), the speed and cost of obtaining detailed molecular information in soybean is no longer an issue. The focus shifts to the utilization and incorporation of massive amounts of genomic information in an applied fashion. By gaining insights into the molecular changes through pedigrees of a breeding program,

various strategies can then be developed to incorporate molecular breeding tools to aid in developing breeding objectives aimed at germplasm enhancement in soybean and possibly other crop plants.

Acknowledgments Plant material supplied by the following public and private collaborating breeding programs is gratefully appreciated; Ridgetown College, La Coop fédérée, Semences Prograin, Agriculture and Agrifood Canada (AAFC) and Pioneer Hi-Bred. We thank Angela Hollis and Jeff Gross for conducting the fragment analysis at the University of Guelph Genomics Facility. Funding for this research was generously provided by the Grain Farmers of Ontario (Project #: S2011GE02).

References

Agricultural Research Institute of Ontario (ARIO) (1998) Research accomplishments 1993–1998. 1998 ARIO filed crops review

Berloo RV (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. Heredity 99(2):232–236

Fu YB, Peterson GW, Morrison MJ (2007) Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. Crop Sci 47:1947–1954

Gizlice Z, Carter TE Jr, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947–1988. Crop Sci 34:1143–1151

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. PNAS 103:16666–16671

Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Jun Li, Jian M, Wang J, Shao G, Wang J, Sun S, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059

Lorenzen LL, Boutin S, Young N, Specht JE, Shoemaker RC (1995) Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. Crop Sci 35:1326–1336

Lorenzen LL, Lin SF, Shoemaker RC (1996) Soybean pedigree analysis using map-based molecular markers: recombination during cultivar development. Theor Appl Genet 93:1251–1260

Lui K, Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Sci 43:1235–1248

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Phys 147:969–977

Narvel JM, Jakkula LR, Phillips DV, Wang T, Lee S-H, Boerma HR (2001) Molecular mapping of *Rxp* conditioning reaction to bacterial pustule in soybean. Heredity 92(3):267–270

Ott A, Trautschold B, Sandhu D (2011) Using microsatellites to understand the physical distribution of recombination on soybean chromosomes. PLoS ONE 6(7):e22306

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyoden B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan PB, Specht J, Grimwood J, Rokhsar D, Stacy G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nature Biotech 18:233–234

Shoemaker RC, Guffy RD, Lorenzen LL, Specht JE (1992) Molecular genetic mapping of soybean: map utilization. Crop Sci 32:1091–1098

Simko I, Haynes KG, Jones RW (2004) Mining data from potato pedigrees: tracking the origin of susceptibility and resistance to *Verticillium dahliae* in North American cultivars through molecular marker analysis. Theor Appl Genet 108:225–230

Sjakste TG, Rashal I, Röder MS (2003) Inheritance of microsatellite alleles in pedigrees of Latvian barley varieties and related European ancestors. Theor Appl Genet 106:539–549

Song QJ, Marek LF, Shoemaker RR, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

SoyBase (2011) Soybean breeders toolbox. Retrieved May 2011, from http://soybase.org

Soyinfo Center (2013) Soy from a historical perspective. http://www.soyinfocenter.com. Accessed Feb 2013

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Tanner JW, Luzzi BM, Gostovic P, Montminy W, Hume DJ (1998) OAC Bayfield soybean. Can J Plant Sci 78:625–626

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effect of artificial selection on the maize genome. Science 308:1310–1314

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24(3):875–888