Development and Implementation of Biotechnological Tools for Rapeseed Breeding in Southern Chile

Federico Iñiguez-Luy, PhD.
Group Leader - Oilseed Breeding and Genomics
Genomics and Bioinformatics Unit
CGNA

federico.iniguez@cgna.cl

SLAC, Passo Fundo
20th August, 2014

CGNA
Centro de Genómica Nutricional Agroacuícola
INNOVANDO DESDE LA ARAUCANÍA

El Centro de Genómica Nutricional Agroacuícola, está ubicado en la Región de La Araucanía, es un centro de investigación científica-aplicada orientado al desarrollo de cultivos estratégicos para la alimentación y procesos tecnológicos que requiere la agroindustria.

Los productos de su trabajo en investigación y desarrollo van en directo beneficio de toda la cadena agroalimentaria:

Agricultores que disponen de semillas de calidad generadas por el CGNA, orientadas a producir más proteína, ácidos grasos, aceites esenciales y otras características especiales.

Agroindustria que cuenta con materias primas y productos de origen vegetal con gran calidad nutricional y valor agregado.

Consumidores finales que accederán a una mejora en su alimentación.

El CGNA se compone de profesionales formados en universidades del primer mundo y del medio nacional. Doctorados, postdoctorados, profesionales de pregrado y estudiantes de pre y post grado, cubren una serie de disciplinas que transforman al centro en un polo de atracción para hacer ciencia y desarrollar know-how desde La Araucanía.

Este centro cuenta con más de una decena de colaboradores internacionales en distintos proyectos. Universidades y otros centros de prestigio internacional conforman una red global de trabajo en ciencia y tecnología.

El CGNA posee financiamiento base del Gobierno Regional de La Araucanía y el Programa Regional de CONICYT, fue fundado por la UFRO e INIA.

Por segundo año consecutivo coordina a nivel nacional y desde Temuco el Fascination of Plants Day logrando reunir a diversas Instituciones chilenas en torno a esta celebración.

www.cgna.cl

Our Challenges (1)

•Respond to the increasing demand for **proteins, fatty acids and soluble fiber** for animal feed and human consumption

-Linseed
-Rapeseed
-Yellow Lupins

✓**Specific nutritional profiles in grains**

✓**Food technology of agricultural products**

Our Challenges (2)

- Develop **germplasm, advance lines and varieties** with high adaptability and especific nutritional profiles to a wide target of enviroments, with emphasis on the central south region of Chile.

✓**Local Genetics**

Our Challenges (3)

- Respond to an ever changing and complex **climatic scenario** through the study of the genome of our three key crops: Flaxseed, Rapeseed and Yellow Lupins

✓Drought
✓High solar
✓ radiation
✓Fluctuant temperatures
✓Increased disease incidence

Rapeseed (*B. napus*) Importance in Chile

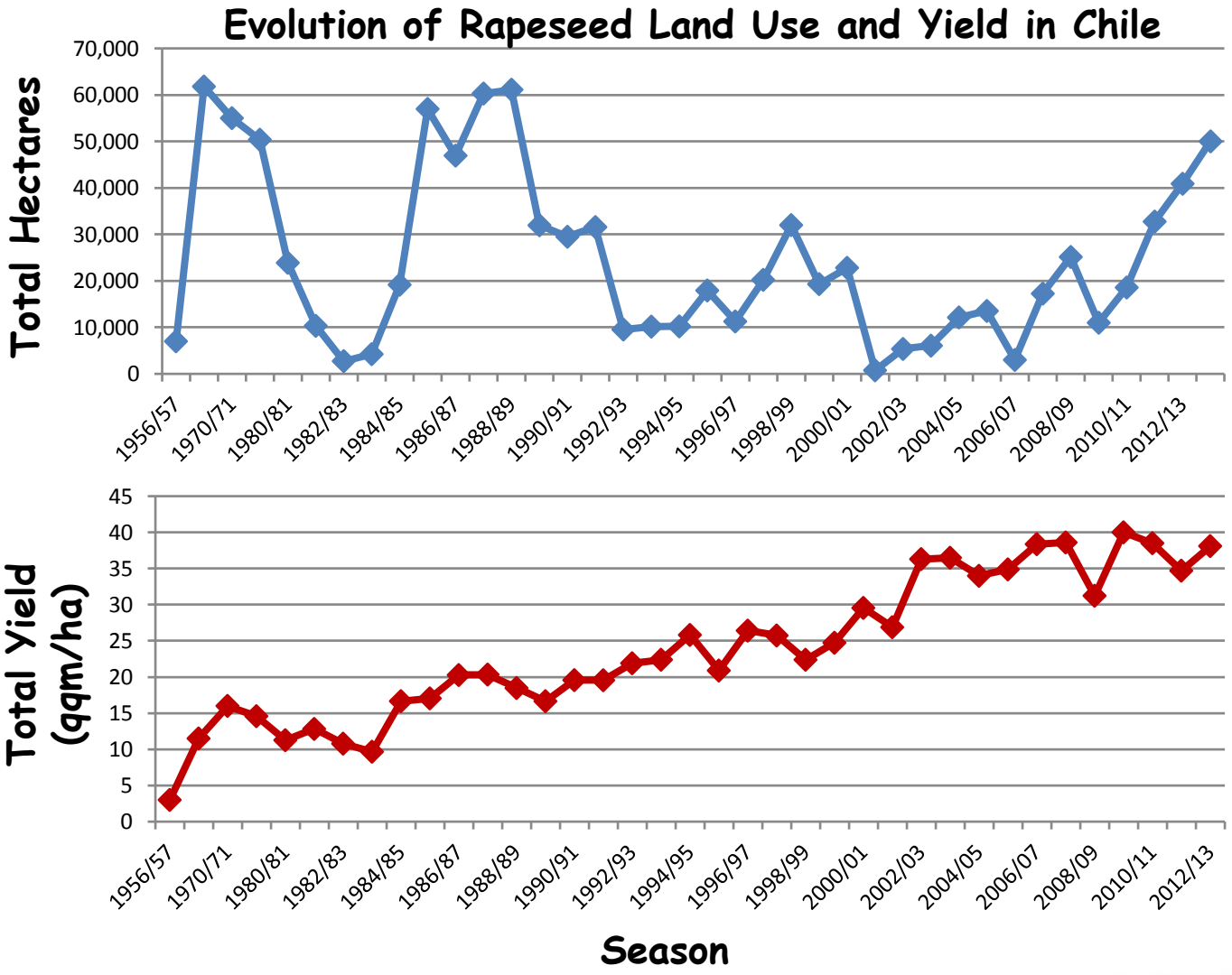
•*Brassica napus,* commonly known as rapeseed, raps or colza is the second oilseed crop with ~ 14% of the total oil production and grows well in most temperate regions of the globe.

•In Chile, rapeseed production and yield per hectare have grown considerably in the last 30 years. This is due to an increased demand for the production of oil, both for the salmon cluster and recently for human consumption.

•The main export destinations for the domestic oil production reside in Latin America with Colombia taking about 85% of national exports. Next is Argentina, with about 9%.

•The main origin of Chilean imports of rapeseed oil is from Canada, with about 90% of the imported oil. The second source is Argentina.

Rapeseed Importance in Chile



Rapeseed Projections in Chile

•International prices have set favorable conditions for rapeseed production.

•A growing demand from the salmon industry and high potentials for oil production oriented to human consumption.

•New development of oil crushing plants in southern Chile.

•The importance that rapeseed has in the rotation scheme with annual crops especially cereals.

•It is estimated that the total area sown with rapeseed will keep on rising and will surpass its historical record of 60,000 hectares.

Rapeseed Research at CGNA

Identification of breeding lines with high adaptability to central south Chile and a wide target of environments

Development of double haploid (DH) and recombinant inbred (RI) lines

Fingerprinting and genotyping of parental (DH and RI) lines

Gene Introgression and evaluation of elite background

Agronomical and nutritional evaluation of selected DH and RI lines

Developed inbred lines, OP varieties and hybrids exhibiting high agronomical potentials and containing added nutritional value

Development and Implementation of Genomic Tools

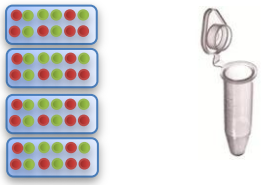
Marker Discovery: Hight-throughput Single Nucleotide Polymorphism (SNP) discovery in areas of the genome of *Brassica napus* that have been historically associated to agronomical and nutritional traits.

SNP Markers
Discovery Using Sequence Capture Tecnology

Selection of genomic regions of interest

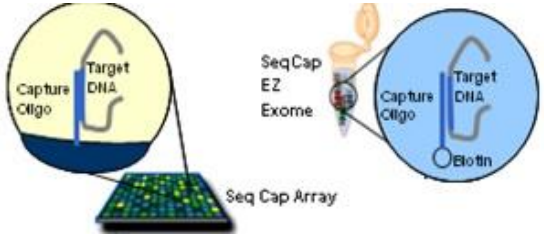
Targeted regions

Sequence Capture Platform Design



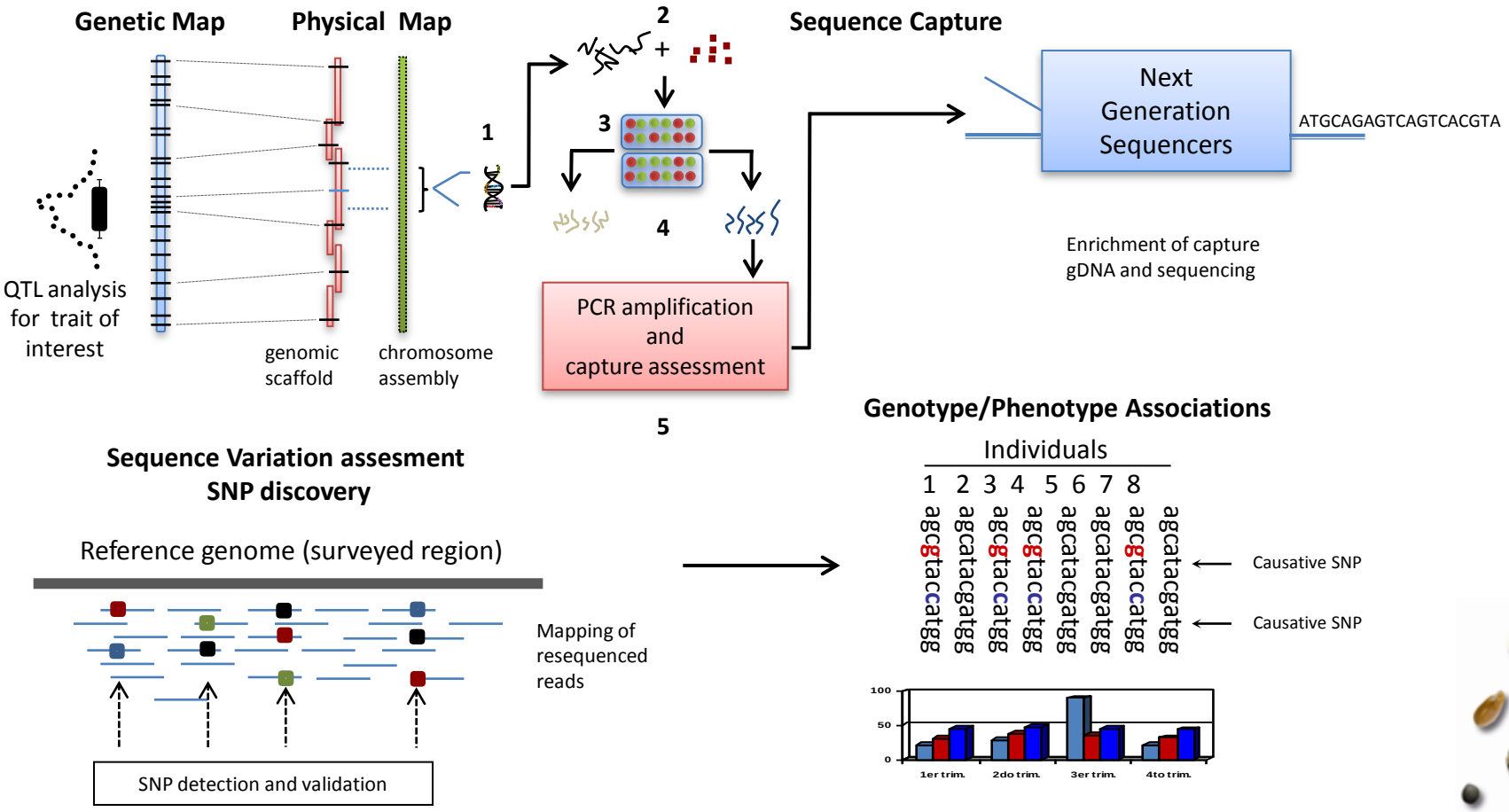
Solid or Liquid

Comparative analysis of captured genomic areas



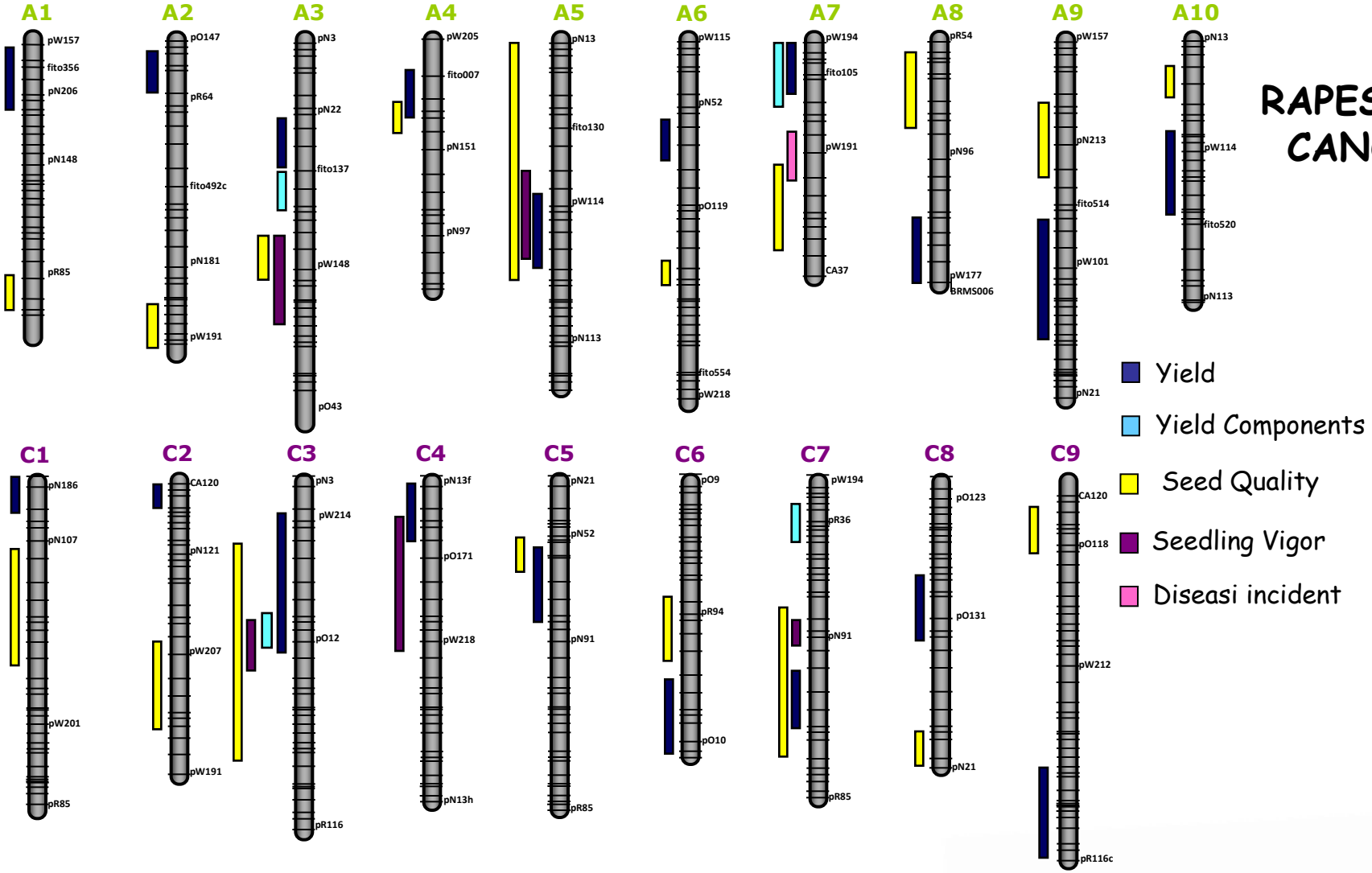
Capture Oligo DNA hybridization

SNPs Discovery using Sequence Capture and Next Generation Sequencing



FONDECYT Reg. 1100732

Areas of the *B. napus* Genome Interrogated



(Snowdon and Iniguez-Luy, 2012; Clarke et al. 2013)

Objectives

I) Interrogate 47 regions of the *B. napus* genome that explain agronomical and nutritional traits (identified via genetic analysis - QTL-).

II) Resequencing of 10 highly contrasting *B. napus* lines.

III) Discover and characterize (eg. genic vs intergenic, coding vs intronic sequences and type of substitution - transition vs transversion) SNP type markers in the regions interrogated.

IV) Validate the developed SNP markers (potential use and effectiveness of discovery).

Results

Sequence reads mapping summary using multiple reference sets.

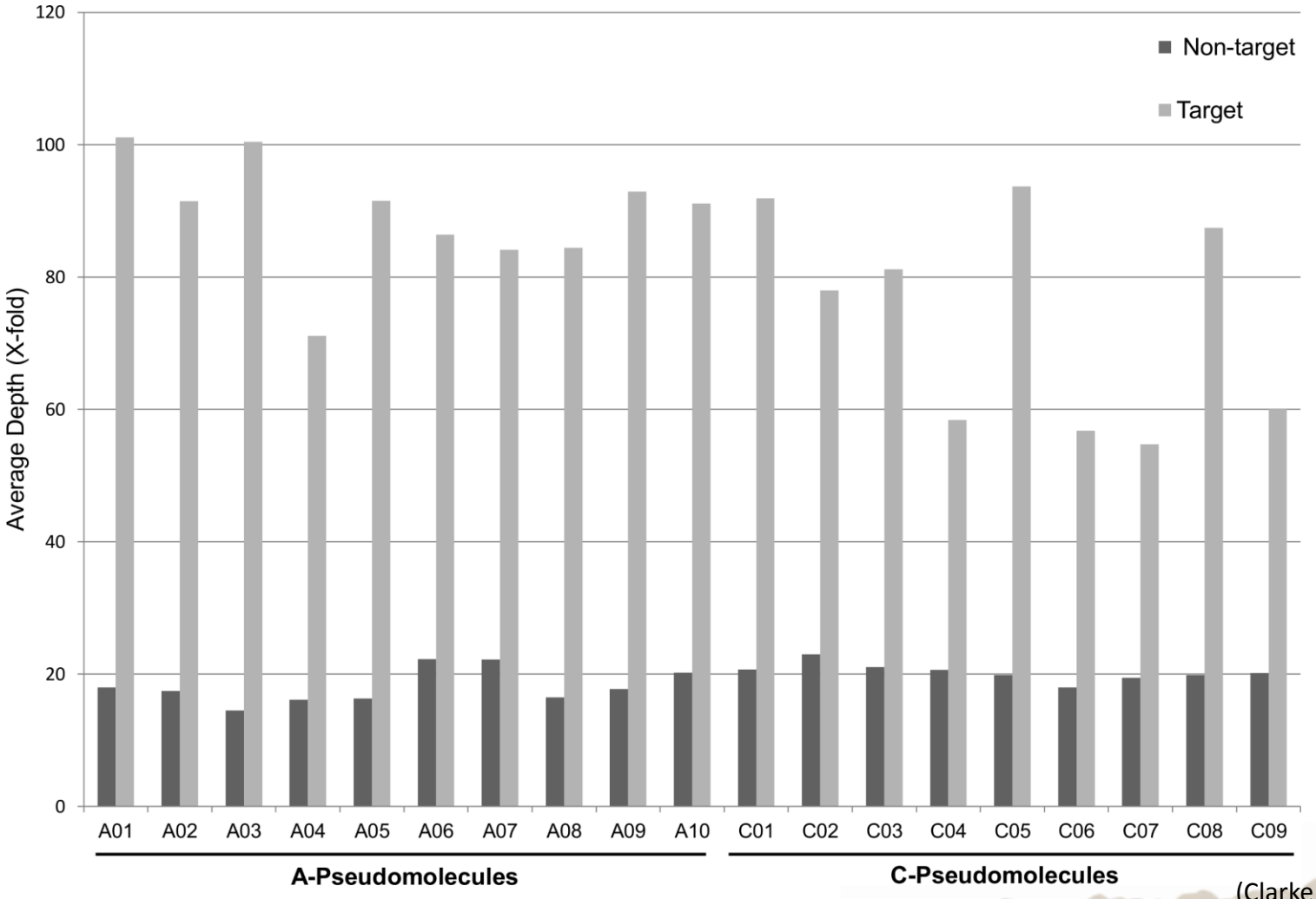
		454 Chemistry Data					
		Capture Design Reference (883 Sequences)		Capture Design with Orthologues (883 + 1074 Sequences)		A & C Genome Pseudomolecules (19 Sequences)	
Lines	SR	RM	RMp	RM	RMpi	RM	RMpi
DH12075	1,289,496	414,569	32	457,448	10.3	775,554	69.5
PSA12	826,680	241,867	29	279,959	15.7	497,595	77.7
Express	827,074	226,406	27	271,137	19.8	475,214	75.3
V8	711,244	190,312	27	230,314	21	411,708	78.8
Tapidor	778,116	230,934	30	266,100	15.2	453,606	70.5
Ningyou	803,553	240,828	30	277,338	15.2	480,899	73.4
Rainbow	742,283	207,465	28	248,757	19.9	432,873	74
YN-429	735,005	219,859	30	247,101	12.4	427,669	73.1
CGNA1	742,361	201,604	27	241,700	19.9	426,791	76.6
CGNA2	717,016	195,811	27	233,568	19.3	412,298	76.5

SR = Sequence Read, **RM** = Reads Mapped, **RMp** = % of Reads Mapped
RMpi = Reads Mapped % Increment.

(Clarke et al., 2013)

Results

Sequencing coverage obtained for the interrogated regions



(Clarke et al., 2013)

Results

Total number of detected SNPs and final number of candidate SNPs

Filtering Criteria	Number of SNPs Excluded	Number of SNPs Remaining
None	0	2,740,205
Multiple Variants hits	33,917	2,706,288
Sequencing Bias*	2,111,439	594,849
Flanking Sequence**	5,482	589,367
Candidate SNPs		**589,367**

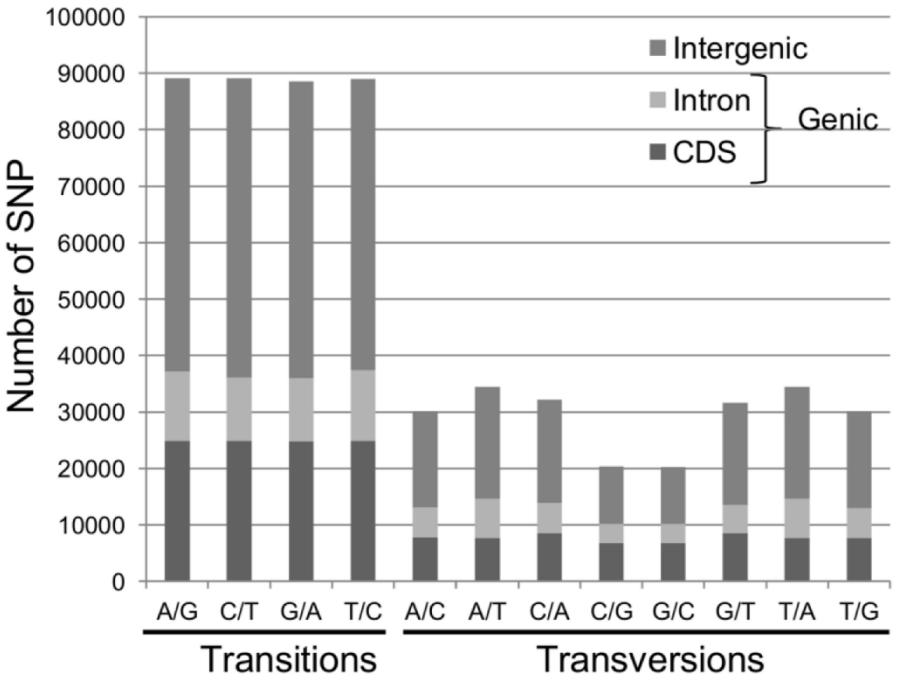
* Removal of SNPs containing only heterozygous and bias SNP calls.

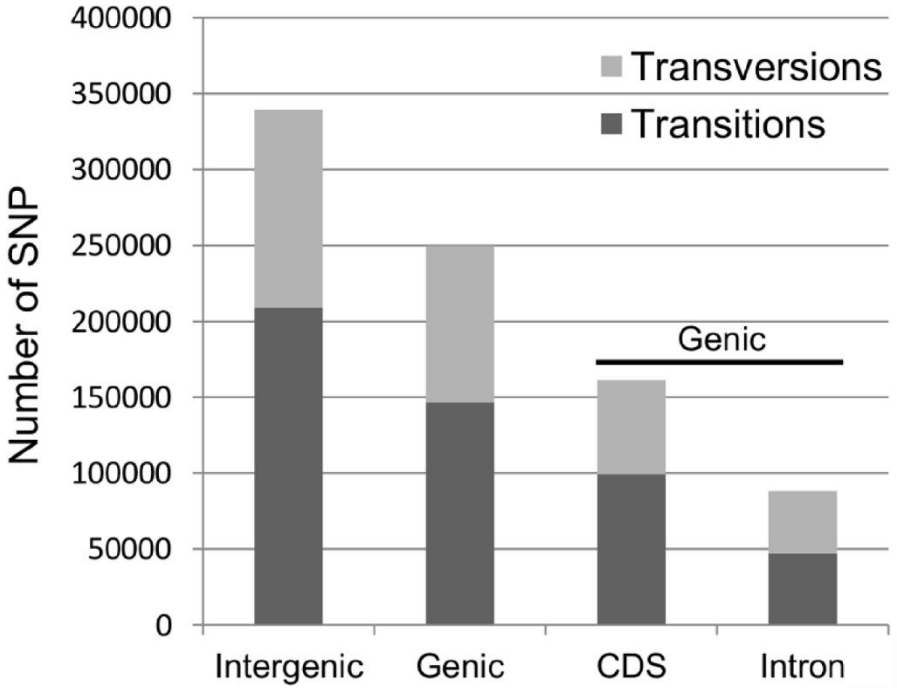
** Removal of SNPs not meeting KASPar or Illumina Infinium flanking sequence requirements.

(Clarke et al., 2013)

Results

SNP charaterization: Type of mutation and SNP location





(Clarke et al., 2013)

SNP Marker Validation

1) KASPar Assay

- *B.napus* Diversity Set
- DH12075xPSA12
- V8xExpress

2) 60K Illumina Infinium array

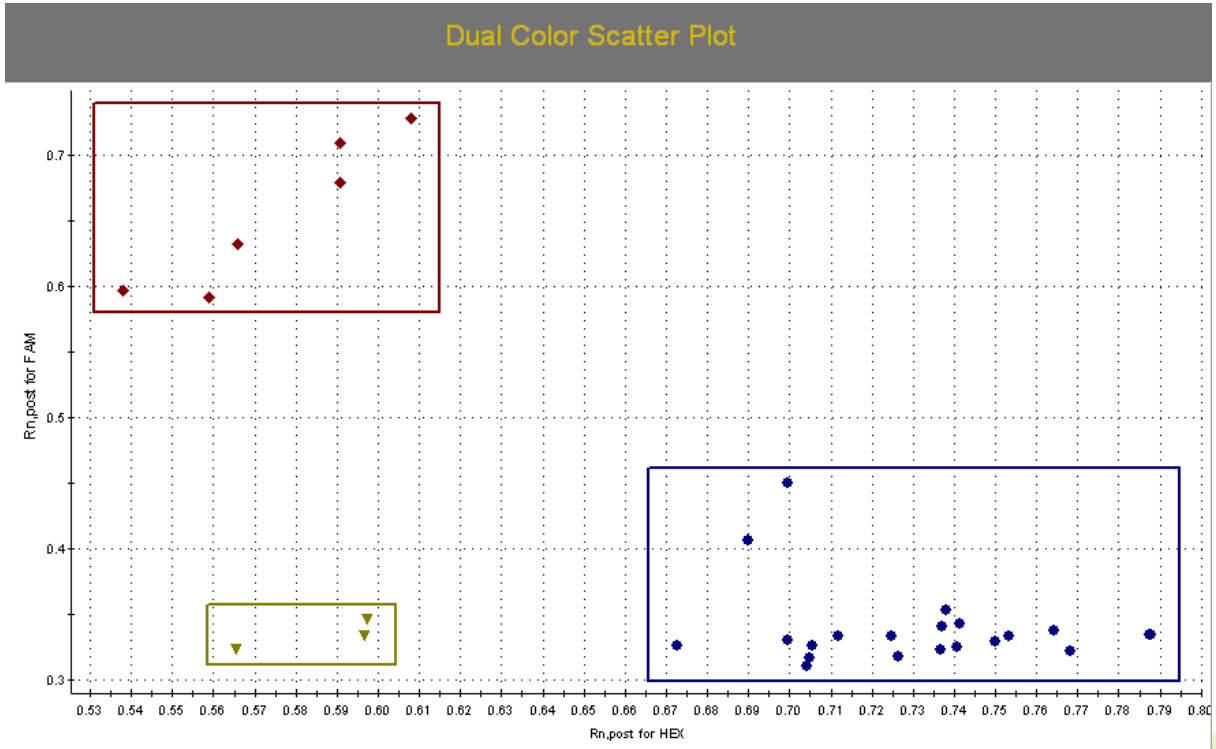
- *B.napus* Diversity Set
- Polymorphic SNP Determination (%)
- Presence of True Allele Determination

3) Bi-parental Mapping

- Genomic SNP Confirmation and validation of the sequence captiure design.

SNP Validation in a *B. napus* Diversity Set

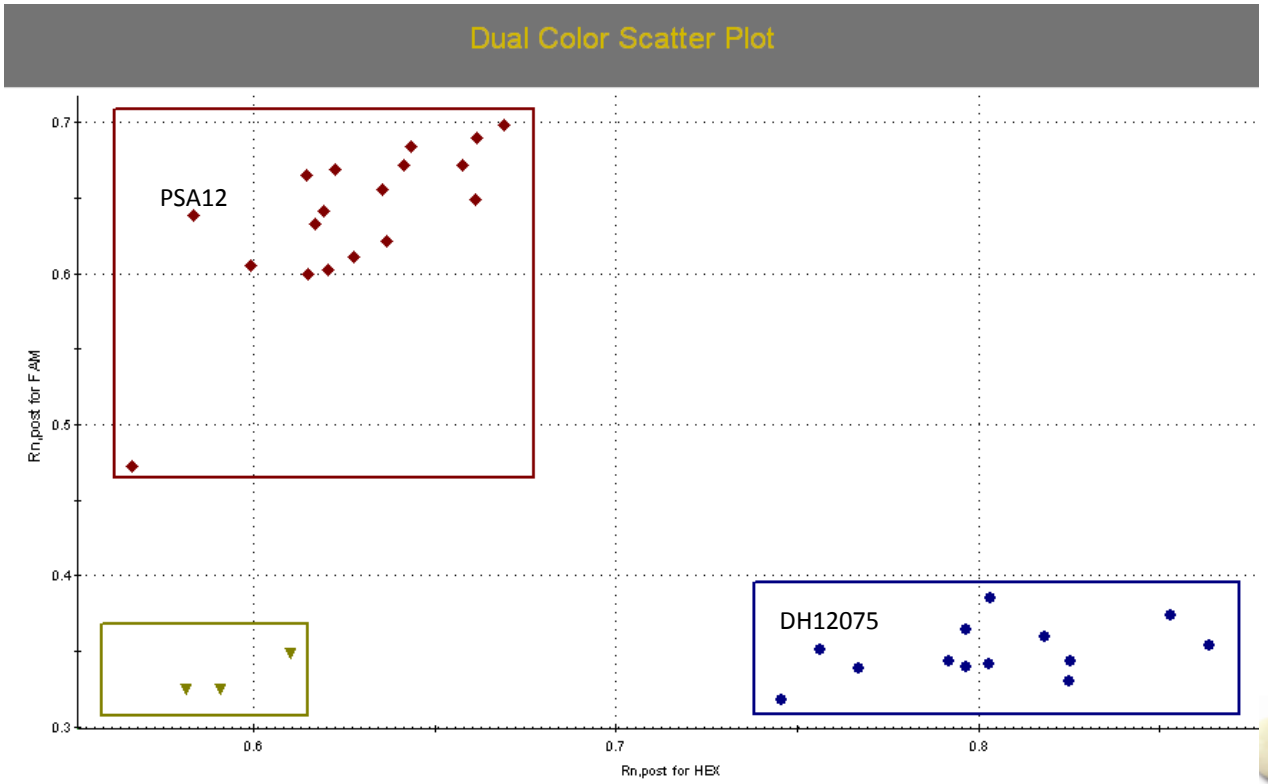
9	10	11	12
Mendel	Express	Topas	PSA12
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
Hamburger	Lirajet	Polo	V8 M6
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
Samourai	Capitol	Yudal	Express M7
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
Quinta	Quantum	Scoop	
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
Ningyou 7	Westar	Paroll	
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
V8	Mohican	YN01	
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
Glacier	Global	Stellar	
REF HEX FAM	REF HEX FAM	REF HEX FAM	
Tapidor	Dunkeld	DH12075	
REF HEX FAM	REF HEX FAM	REF HEX FAM	



- 100 SNPs
- 27 *B. napus* lines

SNP Validation in a Spring type DH population: PSA12xDH12075

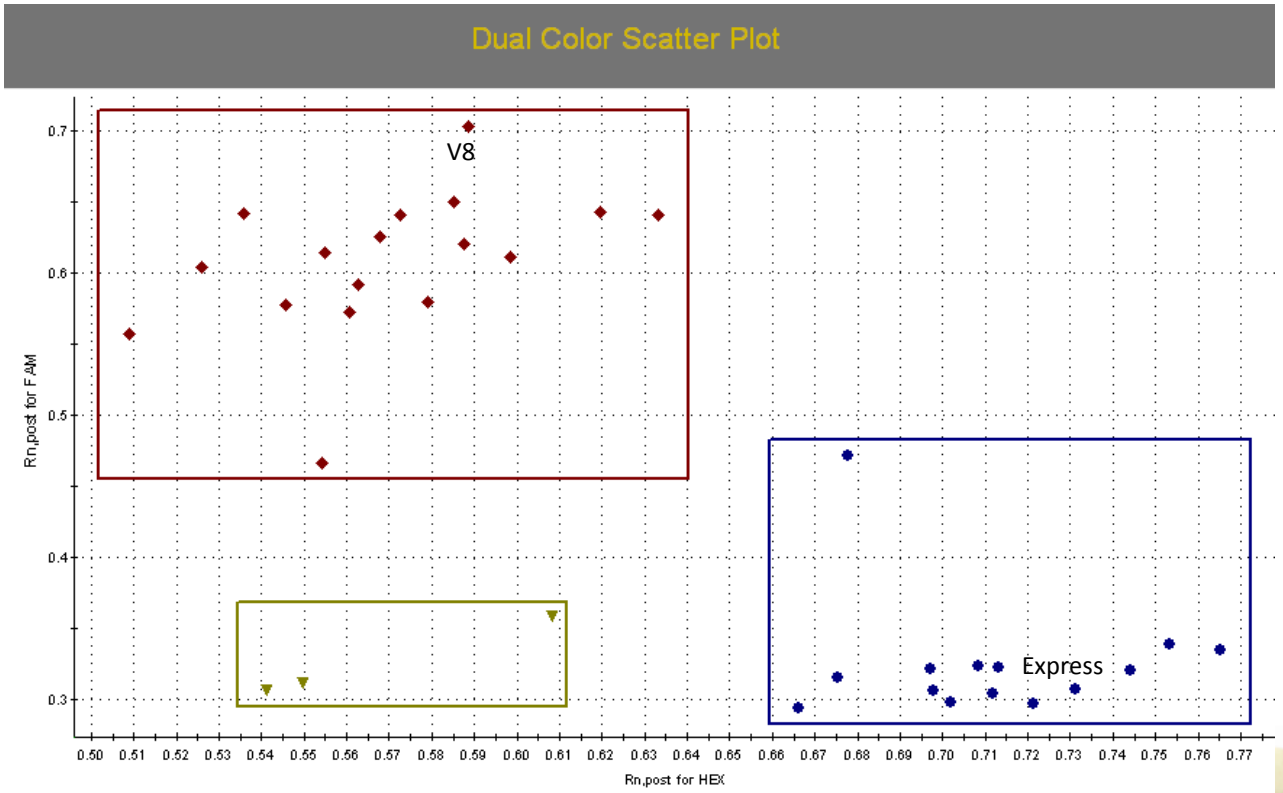
9	10	11	12
SG5	SG100	SG178	SG285
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG19	SG115	SG204	SG300
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG22	SG120	SG209	SG309
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG30	SG130	SG214	SG330
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG35	SG135	SG225	SG333
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG75	SG165	SG249	
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG80	SG168	SG254	DH12075
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
SG85	SG175	SG270	PSA12
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM



- 19 SNPs
- 30 DH 12075 x PSA12 individuals + parents

SNP Validation in a Winter type DH population: V8xExpress

5	6	7	8
3033-1	3009-1	3019-1	3029-1
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3001-1	3010-1	3020-1	3030-1
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3002-1	3011-1	3022-1	3031-1
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3003-1	3012-1	3023-1	3032-1
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3004-1	3014-2	3024-1	3034-1
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3005-1	3015-1	3025-1	
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3006-1	3016-1	3026-1	V8
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM
3007-1	3018-1	3027-1	Express
REF HEX FAM	REF HEX FAM	REF HEX FAM	REF HEX FAM



- 37 SNPs
- 29 V8 x Express individuals + parents

Results

KASPar assay SNP validation summary

	Diversity Set	Parental lines specific SNP*		
		DH12075 x PSA12	V8 x Express	Both populations
Individuals tested	25	30	29	-
Total SNP tested	100	19	37	44
Amplification type				
No amplification	12	0	4	8
Monomorphic	9	1	5	6
Multiple Loci	8	2	0	6
Polymorphic**	71	16	28	24
% PA	88	100	89	82
% PS	**71**	**84**	**78**	**55**

Singleplex KASPar SNP validation assay of a set of 100 discovered SNP markers.

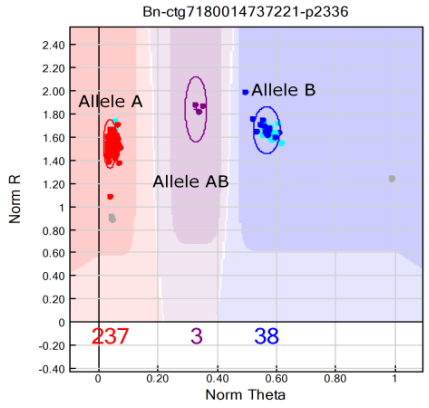
Abbreviations: PA = Positive Amplification; PS = Polymorphic SNPs.
* Represents markers specifically designed from SNPs that showed polymorphism for a specific set of reference mapping parental lines. For instance, polymorphisms between the spring type parents, (DH12075 and PSA12); polymorphisms between the winter type parents, (V8 and Express); and polymorphisms detected for both sets of parental lines at the same SNP locus.
** Polymorphic SNPs also include two dominant (presence/absence of the tested allele) marker types.

(Clarke et al., 2013)

SNP Marker Validation using a 60K Illumina Infinium array

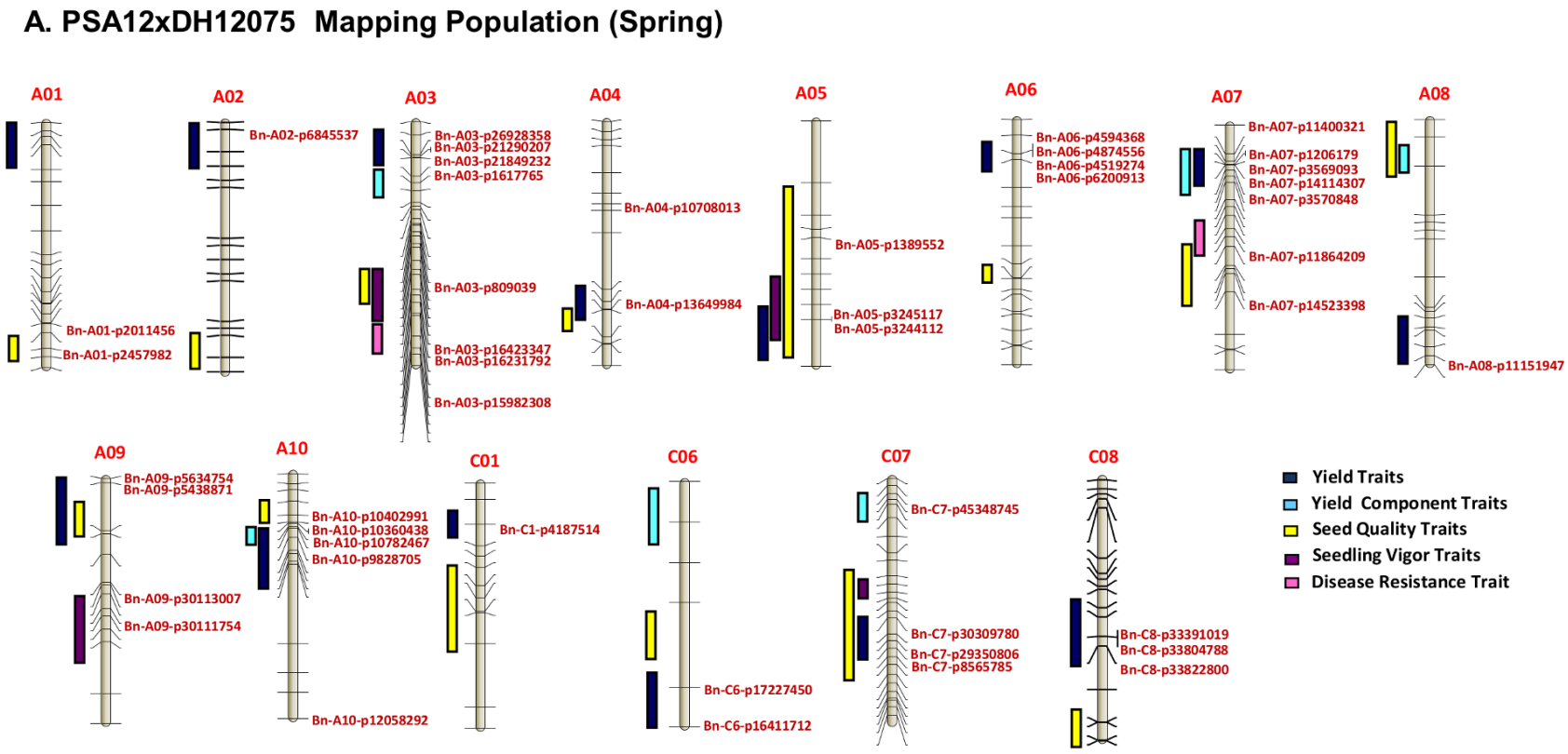
- Brassica 60K iSelect 24x1HD Custom Genotyping Beadchips.
- Infinium HD Assay Ultra Protocol
- 100 *Brassica napus* lines
- 4333 SNPs evaluated

Total Number of SNPs	4333	Percentage
Failed SNPs	363	-
SNPs with positive amplification	3970	**91.6%**
SNPs with multiple loci	1529	**38.5%**
Polymorphic SNPs	2441	**61.5%**



(Clarke et al., 2013)

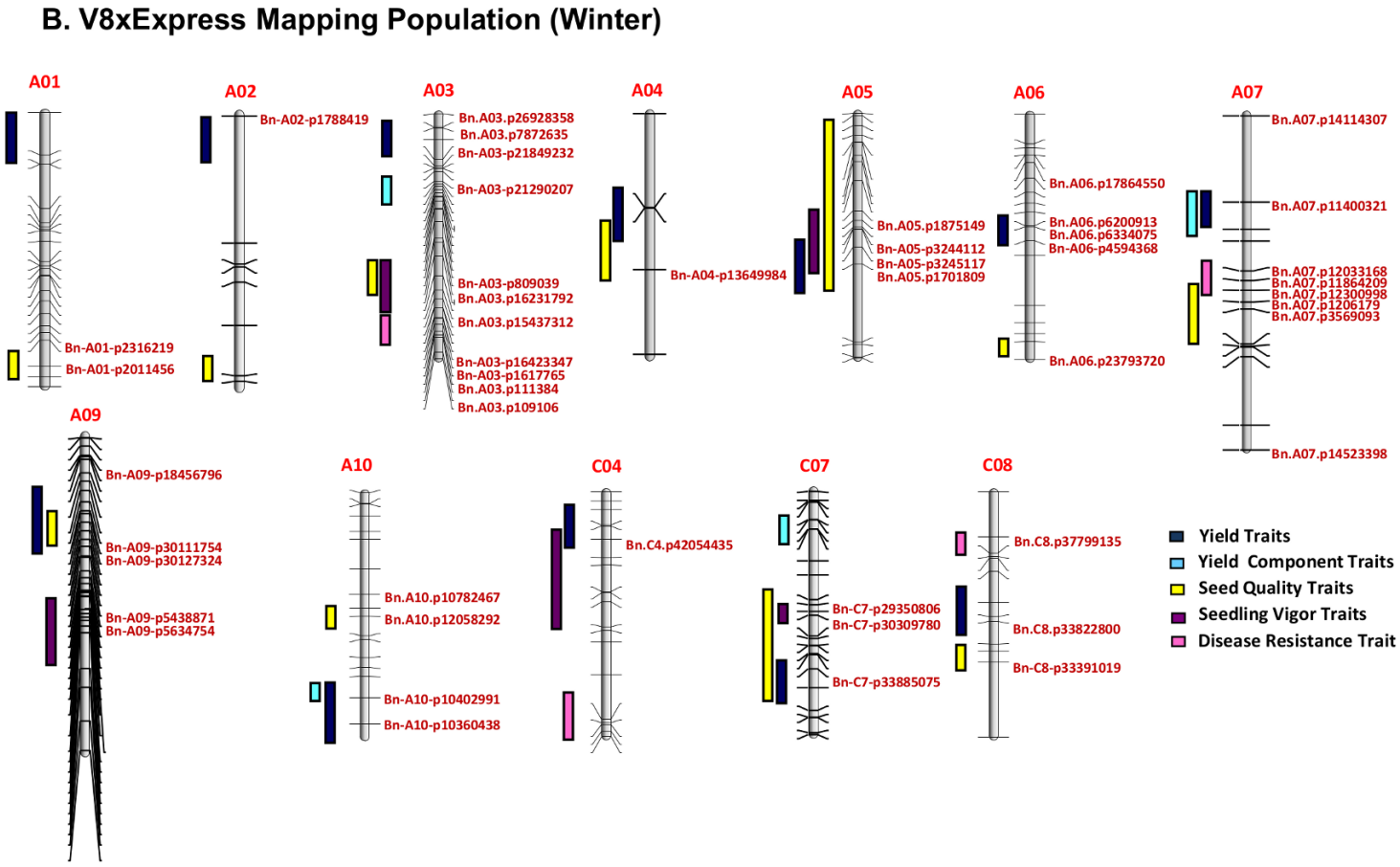
Bi-parental Mapping SNP marker Validation



44 out 48 SNPs markers (91.6%) mapped to the interrogated genomic regions (QTLs) in the DH12075 x PSA12.

(Clarke et al., 2013)

Bi-parental Mapping SNP marker Validation



45 de 48 SNPs markers (93.7%) mapped to the interrogated genomic regions (QTLs) in the Express x V8.

(Clarke et al., 2013)

Conclusions SNP Development

I) It was possible to combine "Sequence Capture " and NGS technologies to interrogate 11% of the *B. napus* genome that explained characters of agronomical and nutritional interest.

II) Using the completed *A* and *C* genome sequences of *B. napus* proved to be crutial to effectively mapped and assembled the haplotype sequence reads for the ten lines studied.

III) More than 500,000 SNPs were discovered in all 47 regions surveyed, including 42% genic and 58% intergenic SNPs.

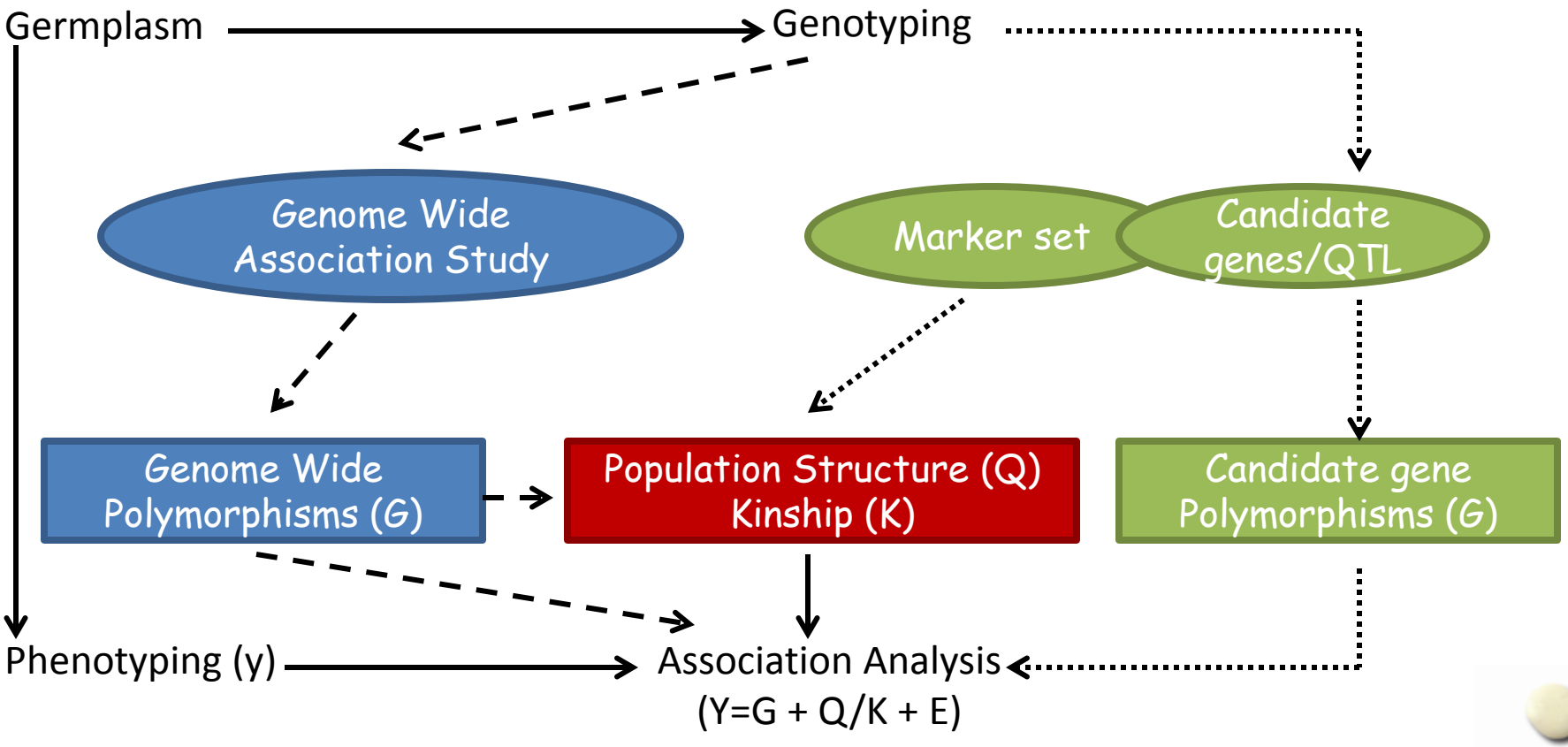
IV) The SNP validation rate was close to 80% demostrating an effective combination of molecular tools and bioinformatics.

V) International Brassica 60K SNP array contribution (Illumina)

Development and Implementation of Genomic Tools

Application: Association mapping of seed quality traits in Brassica napus L. using GWAS and candidate QTL approaches.

Association Mapping Metodologies



Identification of SNP markers associated to seed quality traits using a GWAS and a candidate QTL approach

Modified from Zhu y col., 2008

Objectives

I) Evaluate six seed quality traits in an association panel composed of 89 lines that were phenotyped in two environments during two seasons in a CRBD in Germany and Chile.

II) Select 140 SNP markers (11 QTLs , ~11.5 Mpb) for the cQTL approach and 5506 SNP markers (genome wide) for the GWAS approach.

III) Analyze the population structure (K) and the relative kinship (Q) among the lines that compose the association panel.

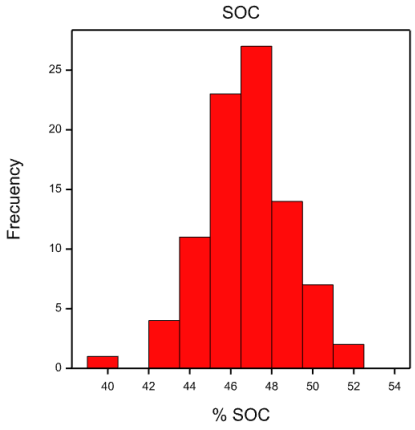
IV) Find significant and positive SNP/phenotype associations using conservative cQTL and GWAS (examination of P-P graphs).

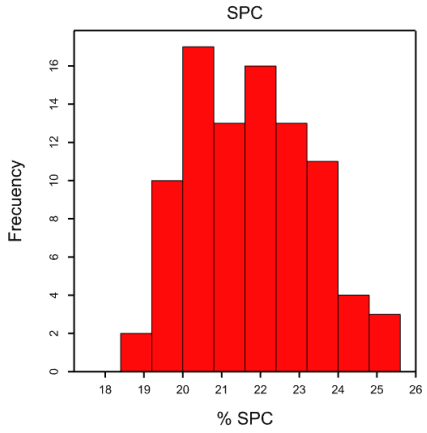
V) Determined the allelic effects of the significantly associated SNP markers (Box plot analysis).

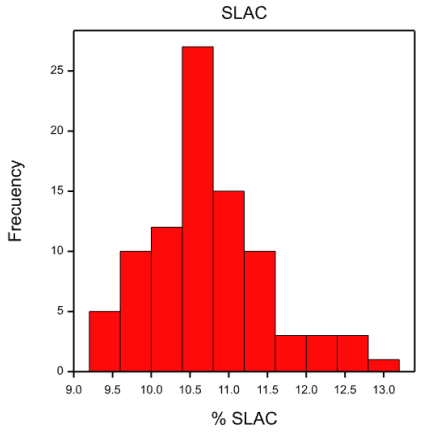
VI) Linkage map analysis for the SNP markers identified using the cQTL approach (validation assay) and characterization of SNP markers showing significant associations.

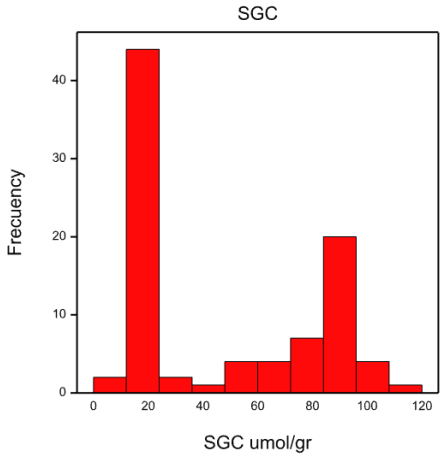
Results

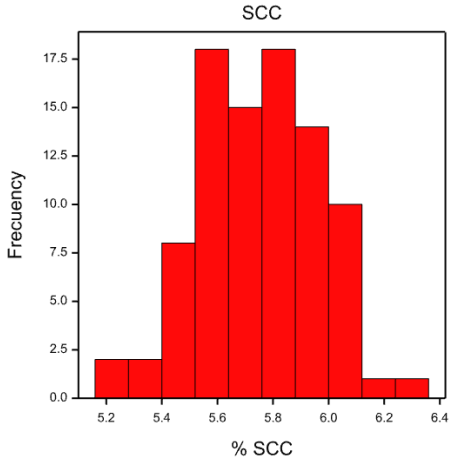
Seed quality histogram distribution for the six traits evaluated in the diversity panel.
(average from the four locations evaluated)

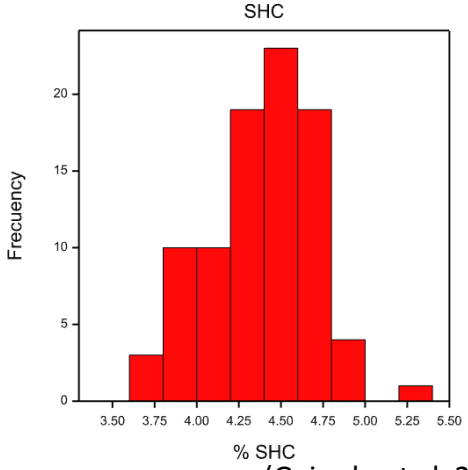












(Gajardo et al. 2014, submitted)

Results

Analysis of Variance and descriptive statistics.

Source of variation	SOC	SPC	SLAC	SGC	SCC	SHC
Genotype	9.87**	8.04**	9.63**	267.38**	3.31**	14.60**
Environment	3.27**	1.45**	1.17**	1.04**	1.86**	0.78**
REP x E	97.72**	46.49**	255.87**	8.92**	244.0**	53.42**
G x E	1.13ns	1.16ns	0.74ns	0.47ns	1.30ns	0.41ns
Mean	46.775	21.765	10.73	47.195	5.755	4.39
Range	39.93-52.4	18.9-25.21	9.28-12.98	10.86-112.7	5.28-6.26	3.66-5.21
H	0.89	0.87	0.86	0.99	0.69	0.84
CV (%)	2.63	4.57	5.37	12.75	3.59	5.8

Phenotypic correlation coeficient for the six traits evaluated.

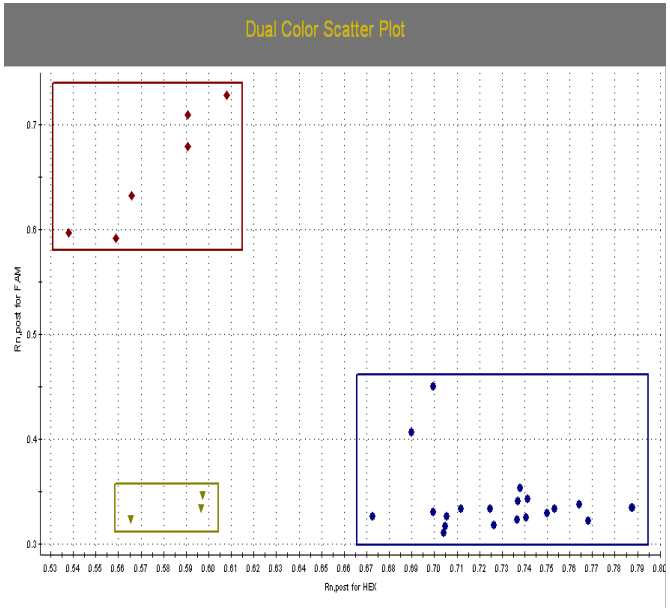
	SOC	SPC	SLAC	SGC	SCC	SHC
SOC	-					
SPC	-0.530 **	-				
SLAC	0.091 ns	-0.014 ns	-			
SGC	-0.006 ns	0.562 **	-0.023 ns	-		
SCC	-0.650 **	0.233 **	-0.563 **	-0.159 **	-	
SHC	-0.782 **	0.139 **	0.017 ns	-0.271 **	0.537 **	-

(Gajardo et al. 2014, submitted)

Results

Selection and genotyping of the association panel (89 lines) using a PCR allele specific method (cQTL) and a 6K Illumina array.

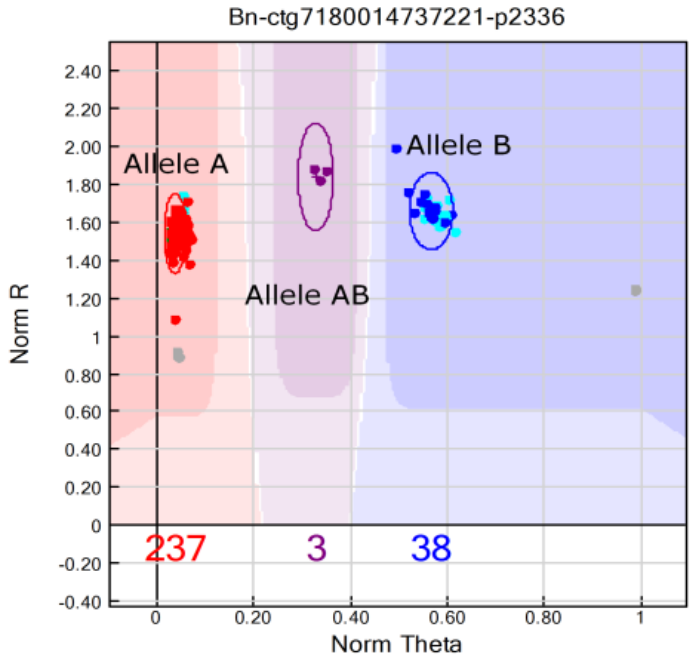
cQTL



MxPro qPCR

100/140 SNP were chosen for further analysis (highly polymorphic)

GWAS



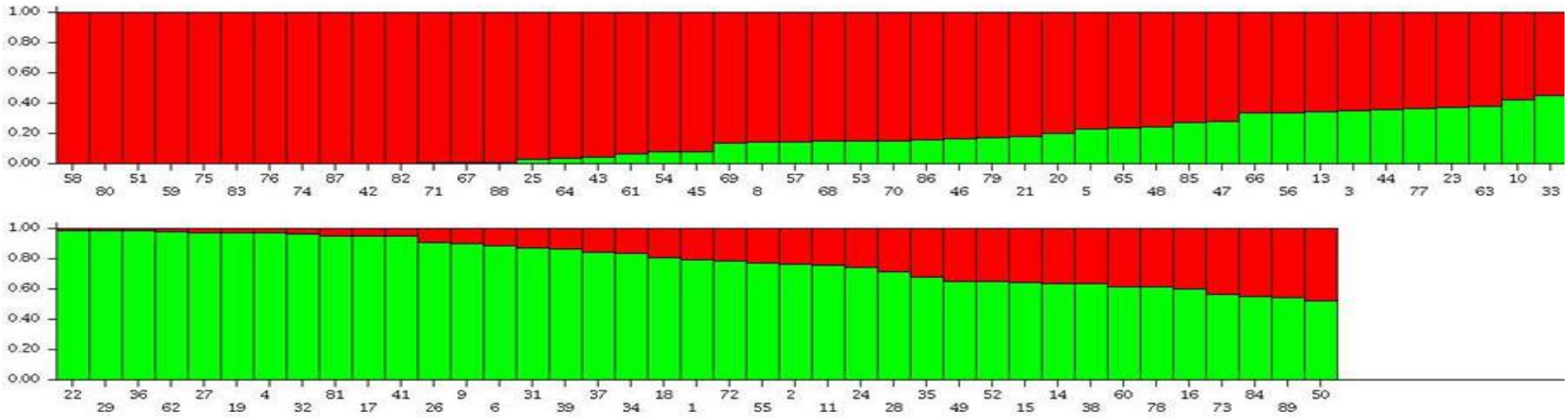
Genome Wide Study

Filter GWAS polymorphic SNPs: 4200

(Gajardo et al. 2014, submitted)

Results

Population Structure determination.

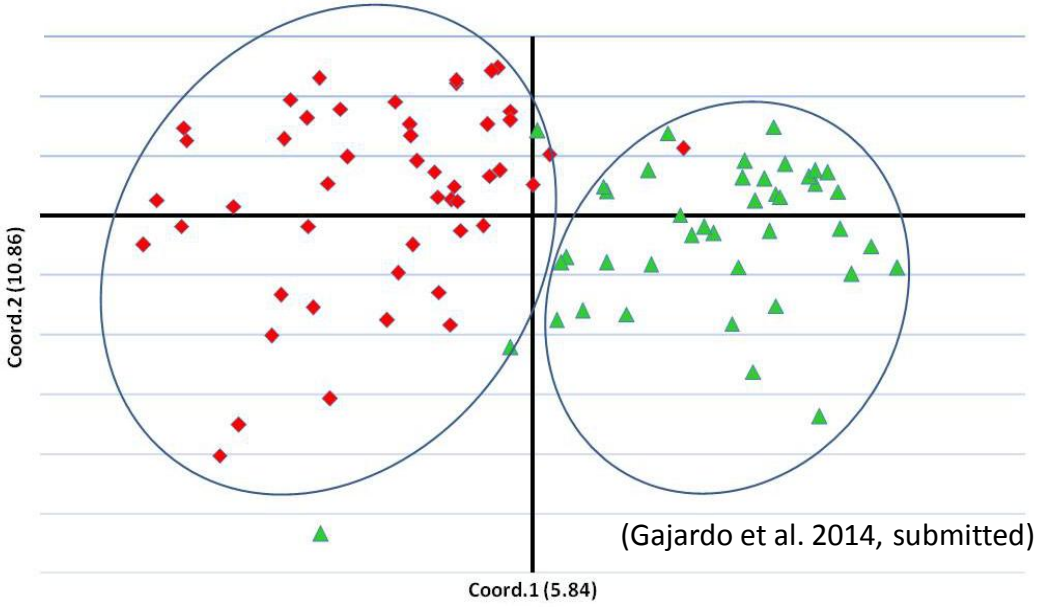


STRUCTURE v2.3.4 (Pritchard et al. 2000)

Group 1: red
Group 2: green

***Fst*: 0.037**
(Genetic Differentiation Coef.)

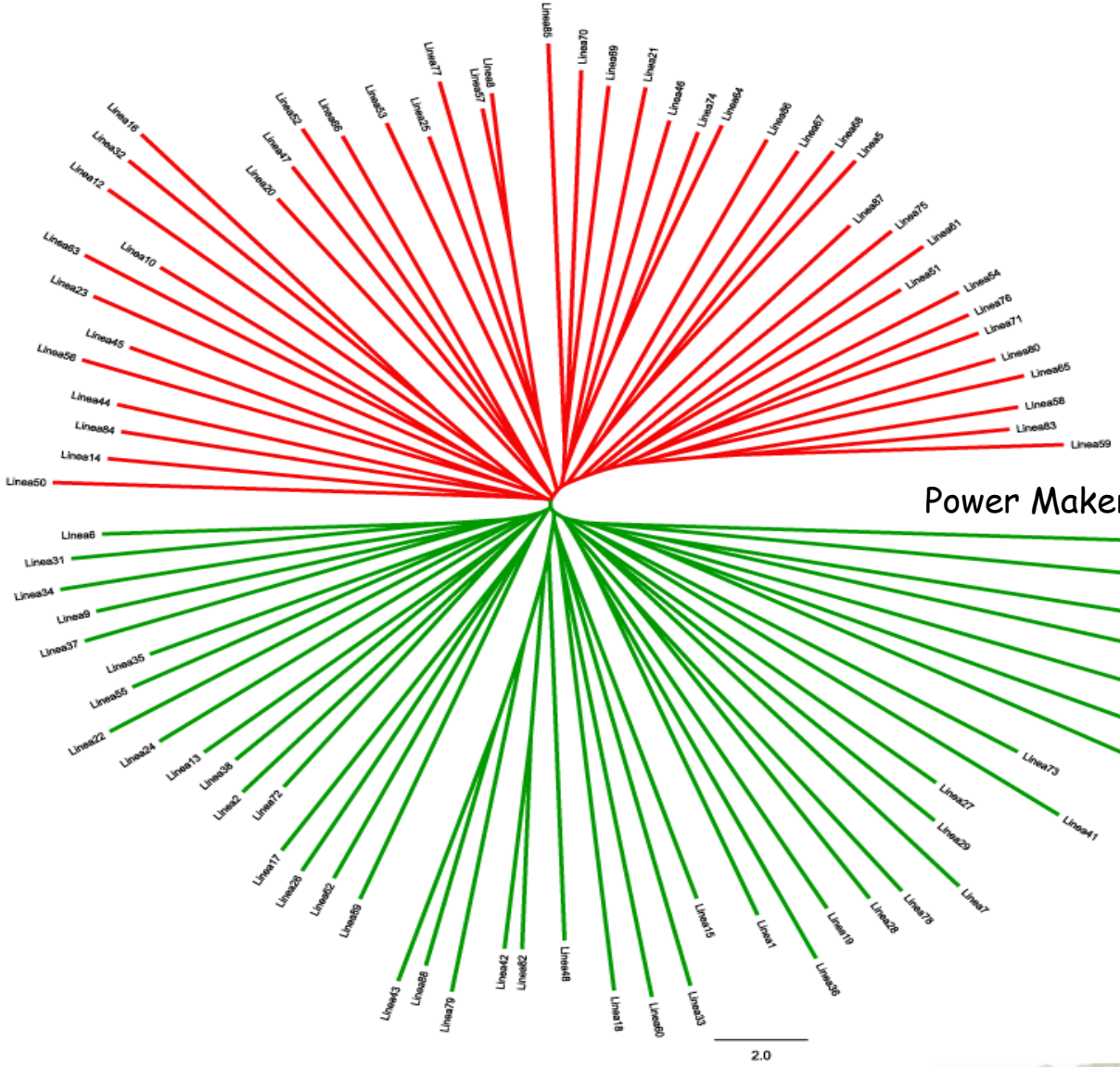
GENALEX v.6.5 (Peakall and Smouse, 2006)



(Gajardo et al. 2014, submitted)

Results

Population Structure determination.



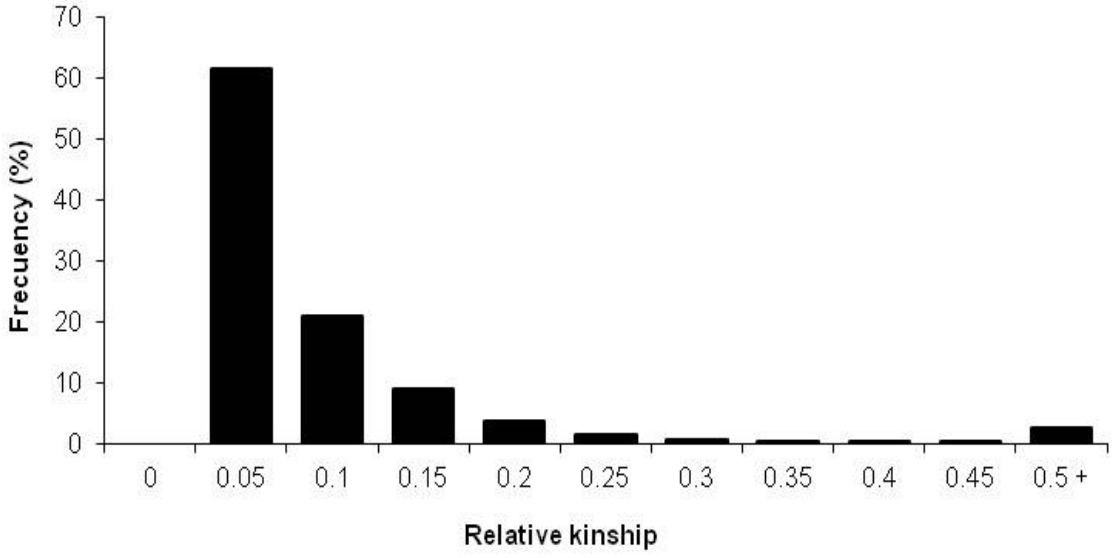
Power Maker v3.25 (Liu and Muse, 2005)

(Gajardo et al. 2014, submitted)

Results

Relative kinship determination for the *B. napus* lines in the diversity set studied.

SpaGEdi v1.4 (Loiselle et al. 1995; Hardy and Vekemans 2002)



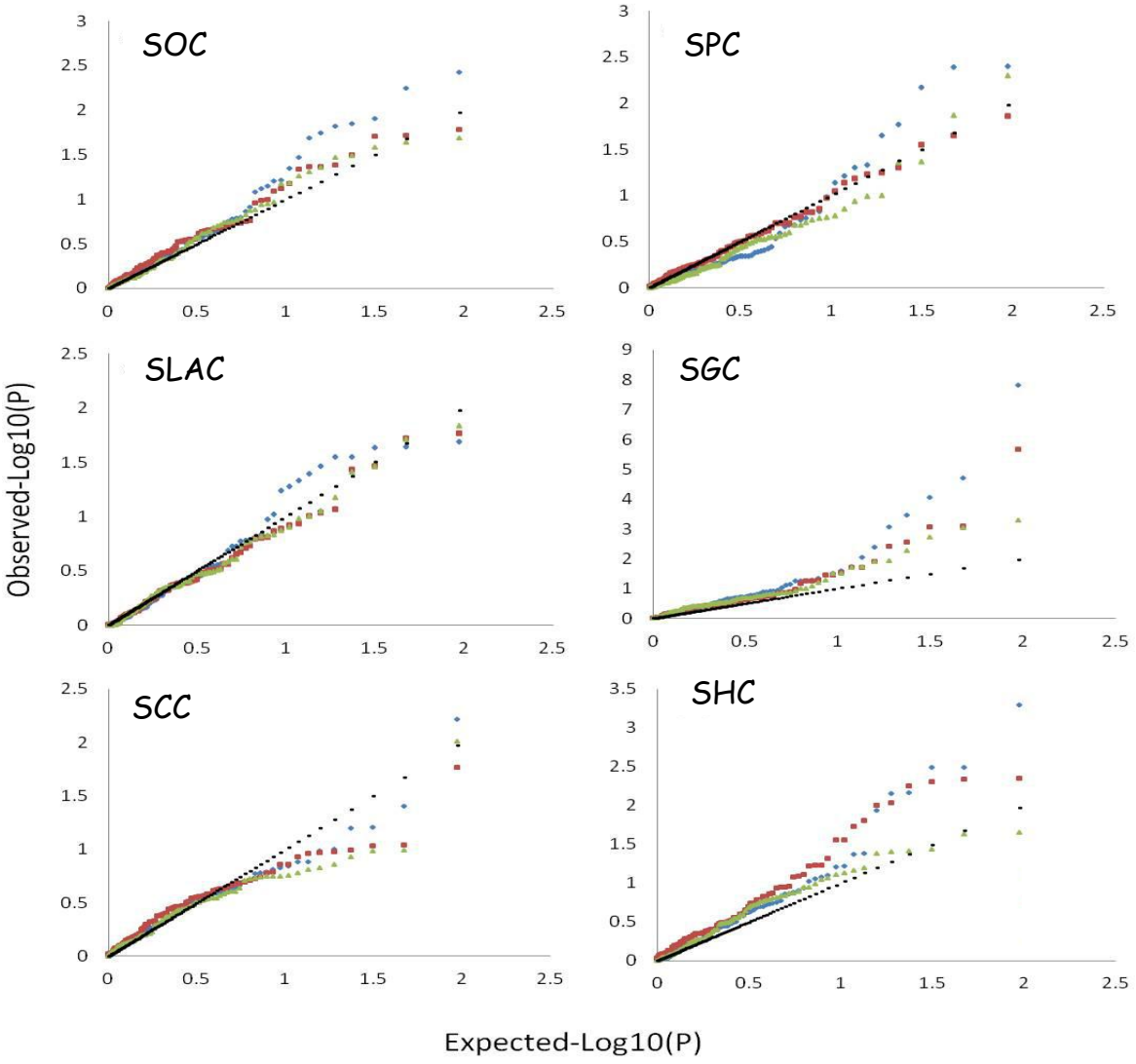
- Average relative kinship between any genotype pair tested was low (0.04).

- Approximately 82% of all kinship comparisons ranged between 0 and 0.1.

(Gajardo et al. 2014, submitted)

Results

cQTL approach: association mapping analysis using three different stringency levels.



- GLM+Q
- MLM+K
- MLM+Q+K

MLM+Q+K: para SOC, SGC, SLAC and SCC.

MLM+K para SPC

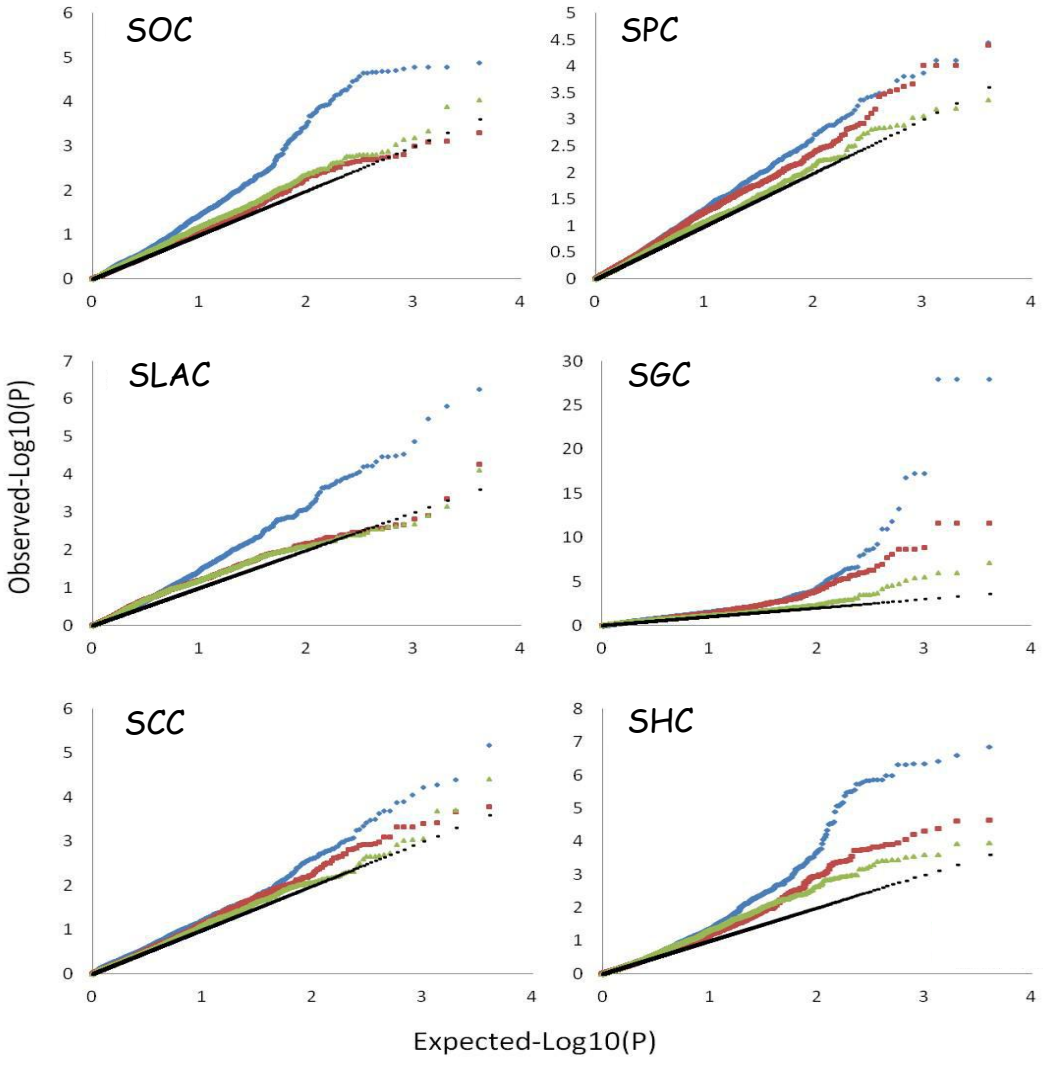
GLM+Q para SHC

P-P graphs for the three association mapping models.

(Gajardo et al. 2014, submitted)

Results

GWAS approach: association mapping analysis using three different stringency levels.



- GLM+Q
- MLM+K
- MLM+Q+K

MLM+Q+K: para SOC, SPC, SLAC, SGC, SCC y SHC.

P-P graphs for the three association mapping models.

(Gajardo et al. 2014, submitted)

Results

SNP markers significantly associated with SGC and SHC identified by GWAS and cQTL approaches.

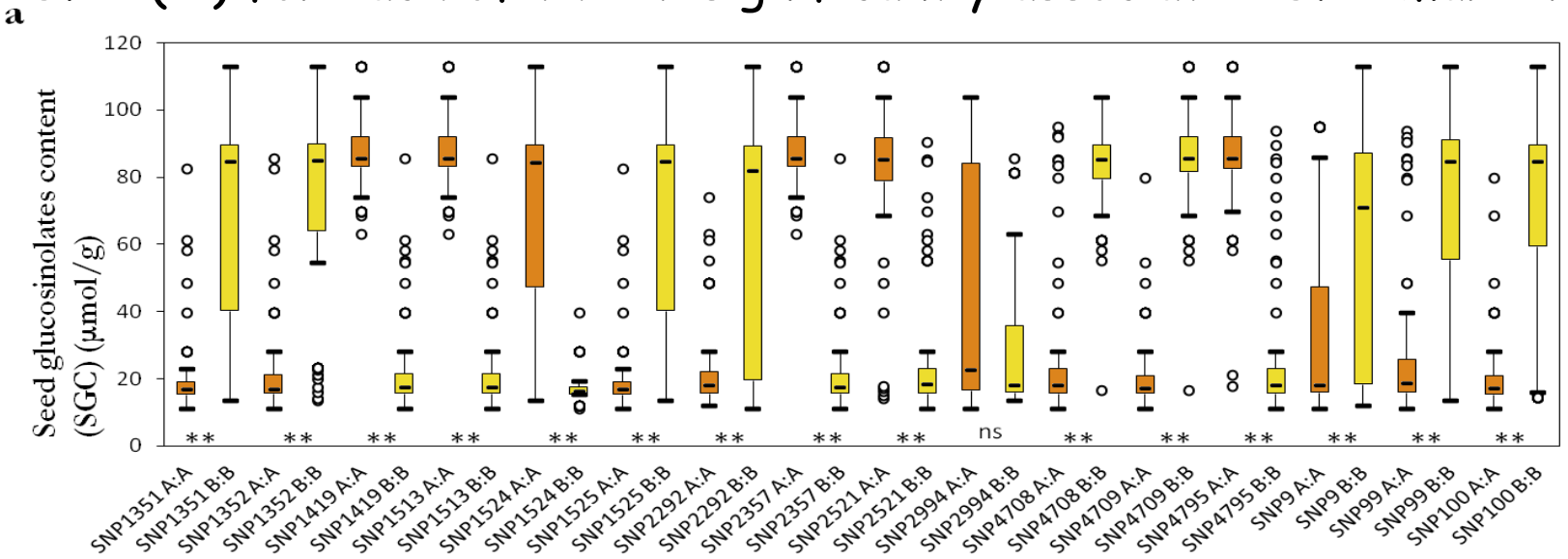
	SNP	GG2010	GG2011	R2010	R2011	Average	R^2
		p-value	p-value	p-value	p-value	p-value	
SGC:GWAS	1351	0.00000146**	0.00000295**	0.00000169**	0.00000146**	0.000000981**	0.149
MLM+Q+K	1352	0.00000579**	0.00000849**	0.00000846**	0.00000358**	0.00000346**	0.150
	1419	0.0000133**	0.0000104**	0.000161*	0.00031*	0.000314^{ns}	0.064
	1513	0.0000133**	0.0000104**	0.000161*	0.00031*	0.000314^{ns}	0.064
	1524	0.0000061**	0.0000125**	0.00000681**	0.0000063**	0.00000426**	0.148
	1525	0.00000146**	0.00000295**	0.00000169**	0.00000146**	0.000000981**	0.149
	2292	0.0000758*	0.000043*	0.0000395*	0.0000238**	0.0000259**	0.109
	2357	0.0000133**	0.0000104**	0.000161*	0.00031*	0.000314^{ns}	0.064
	2521	0.0002*	0.0000932*	0.000106*	0.0000668*	0.0000649*	0.080
	2994	0.0000687*	0.00000265**	0.0000928*	0.0000197**	0.0000173**	0.114
	4708	0.000102*	0.0000761*	0.00003**	0.0000463*	0.0000341**	0.122
	4709	0.00000024**	0.000000128**	0.0000000787**	0.000000106**	0.0000000788**	0.202
	4795	0.0000466*	0.00000496**	0.0000169**	0.00000802*	0.00000768**	0.140
SGC:cQTL	9	0.000984*	0.000481*	0.002070579^{ns}	0.002150006*	0.000943*	0.069
MLM+Q+K	99	0.005803458^{ns}	0.001284629*	0.007207239^{ns}	0.001127141*	0.001900969*	0.074
	100	0.000963*	0.000574*	0.001147478^{ns}	0.000272*	0.000522*	0.089
SHC:cQTL	97	0.006035472^{ns}	0.0513845^{ns}	0.000209*	0.02708079^{ns}	0.000506*	0.122
GLM+Q							

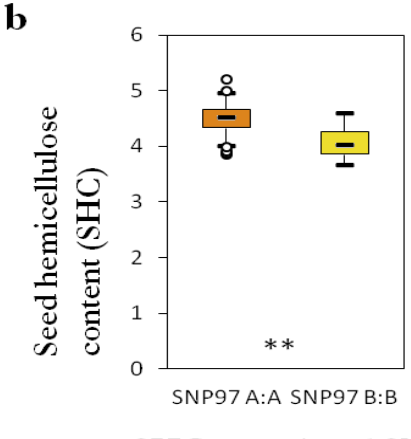
[a]GG2010: Groß Gerau season 2010; GG2011: Groß Gerau season 2011; R2010: Rauischholzhausen season 2010; R2011: Rauischholzhausen season 2011. [b]R^2: phenotypic variation explained for the marker. Significance after multiple comparison adjustments qFDR (Software Q-value): *: $q<0.05$, **: $q<0.01$, ^{ns}: not significant.

(Gajardo et al. 2014, submitted)

Results

Allele effect assessment observed across the distribution for SGC (µmol/g) and SHC (%) for each of the 17 significantly associated SNP markers.





Box plot illustrating the distribution of SGC in the B. napus association panel for: (a) 16 SNP markers significantly associated with SGC (µmol/g) and (b): 1 SNP marker associated with SHC (%).

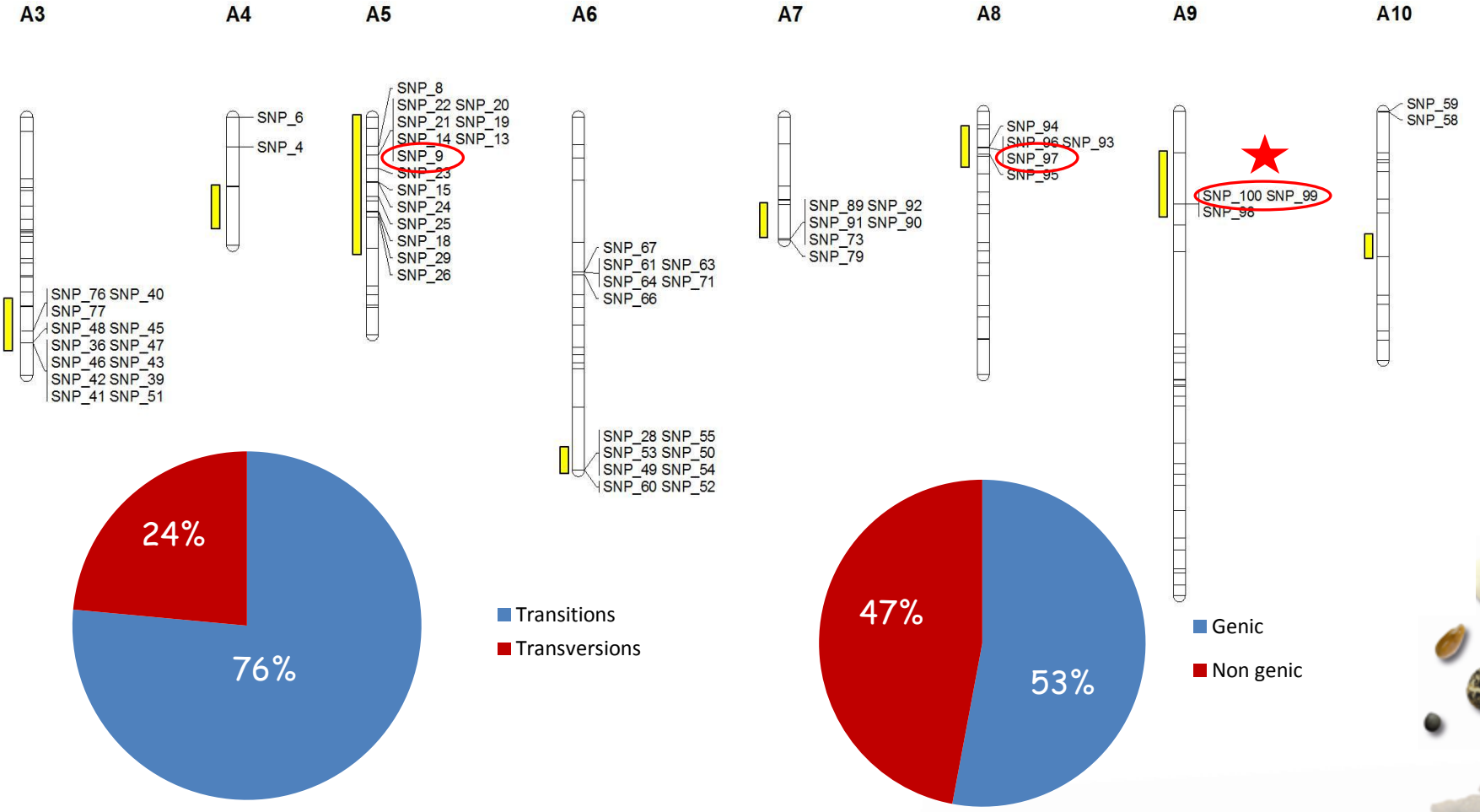
The effect of both SNP markers alleles (A:A=orange bars or B:B=yellow bars) over the trait distribution is shown.

Statistical significance was obtained using the Kruskal Wallis non-parametric test: **: p<0.01, ns: not significant.

(Gajardo et al. 2014, submitted)

Results

Linkage map analysis for the SNP markers identified using the cQTL approach (validation assay) and characterization of SNP markers showing significant associations.



(Gajardo et al. 2014, submitted)

Conclusions SNP/Phenotype Associations

I) The SNP-trait associations identified in this study were highly significant and consistent across environments and seasons evaluated.

II) The targeted cQTL approach resulted in a more efficient methodology to identify SNP associations with seed quality traits compared to GWAS.

III) Most SNP-trait associations displayed a significant allele effect over the associated trait.

IV) Therefore, these SNP markers could assist in the selection of lines with reduced levels of seed glucosinolates and lower hemicellulose content (contributing to improved oil content) in *B. napus* breeding programs.

Acknowledgements

CGNA
Centro de Genómica Nutricional Agroacuícola
INNOVANDO DESDE LA ARAUCANÍA

Dr. Haroldo Salvo
Dr. Maria Laura Federico
Dr. Braulio Soto-Cerda
Mr. Huberto Gajardo
Mr. Fernando Westermeyer
Mrs. Jimena Caril
Ms. Jacqueline Rilling
Ms. Daniela Quesada

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Dr. Rod Snowdon
Dr. Benjamin Wittkop

Agriculture and Agri-Food Canada
Agriculture et Agroalimentaire Canada

Dr. Isobel Parkin
Mr. Wayne Clarke
Ms. Erin Higgins

NRC·CNRC

Dr. Andrew Sharpe
Mrs. Christine Sidebottom

CGNA Rapeseed Breeding Team

Funding

GOBIERNO REGIONAL DE LA ARAUCANÍA
CONICYT Ministerio de Educación Gobierno de Chile

CONICYT Regional Program and the Araucania Regional Government
CGNA/R10C100

FONDECYT
Fondo Nacional de Desarollo Científico y Tecnológico
N°: 1100732