



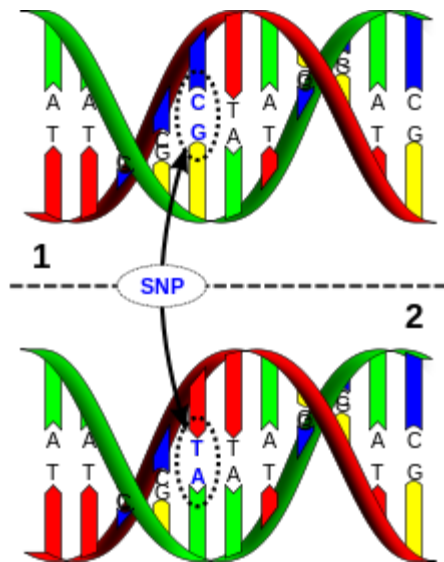
Development of genome-wide SNP-based markers in hexaploid wheat

Sacha Allen



🔥 Background

Single nucleotide polymorphism (SNP)

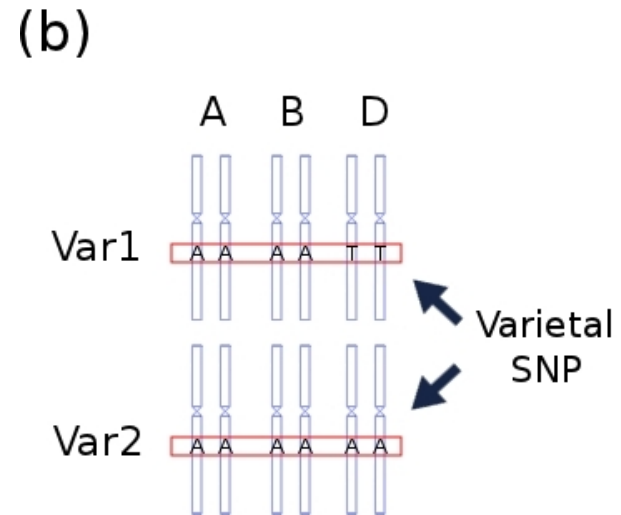
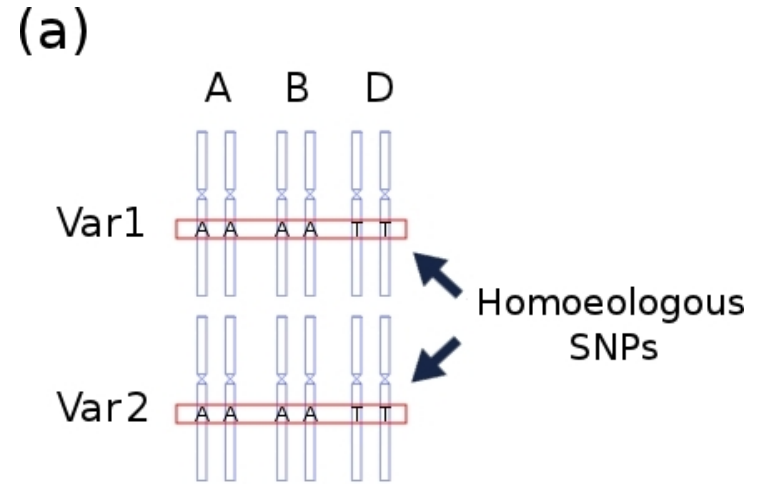
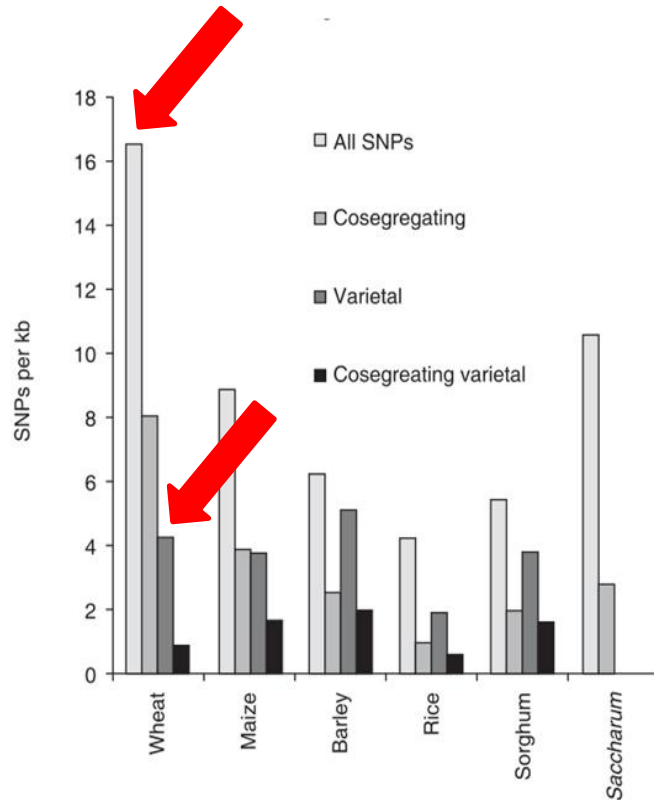


- Occur frequently throughout the genome
- Genetically relatively stable
- Enable tracking of traits through populations
- Low cost, high throughput genotyping



Background

- SNPs in hexaploid wheat



Background

1. Mining the allohexaploid wheat genome for useful sequence polymorphisms

- 5 x Chinese Spring shotgun sequence
- SNP discovery
- Validation of first KASPar wheat panel

2. Development and validation of a flexible genotyping platform for wheat (CIRC)

- 5500 KASPar SNP assays
- Targeted design of assays

3. Wheat LoLa/WISP

- Develop further genotyping platforms for wheat
- Screen more diverse lines



🌾 Wheat Institute strategic program (WISP)



🌿 Exome sequencing



- With Anthony Hall *et al.* and Roche we generated a **132K** feature Nimblegen capture array based on wheat cDNAs
- Capture size = 56 Mb (~ 1 coding genome)
- Hybridised multiple UK lines to the array and Illumina sequenced for SNP discovery



Exome sequencing



Hexaploid

Avalon
Cadenza
Rialto
Savannah
Xi19
Alchemy
Robigus
Hereward
Paragon (JIC)
Paragon (NIAB)
Pavon 76
Chinese Spring L42
Chinese Spring Phi

Tetraploid

KU37
Ttd140

Landraces

Watkins 34
Watkins 141
Watkins 209
Watkins 292
Watkins 352
Watkins 468
Watkins 729
Watkins 126
Watkins 199

Diploid

Triticum urartu
Thinopyrum
bessarabicum
Rye (*Secale cereale*)
Aegilops mutica
Thinopyrum elongatum
Ae. caudata
Th. intermedium
Th. ponticum
Ae. varabilis
Ae. speltoides
Ae. tauschii 232
Ae. tauschii 320
Ae. tauschii JIC2220007
Ent-392
Ent-414
Ent-336
Ent-o88

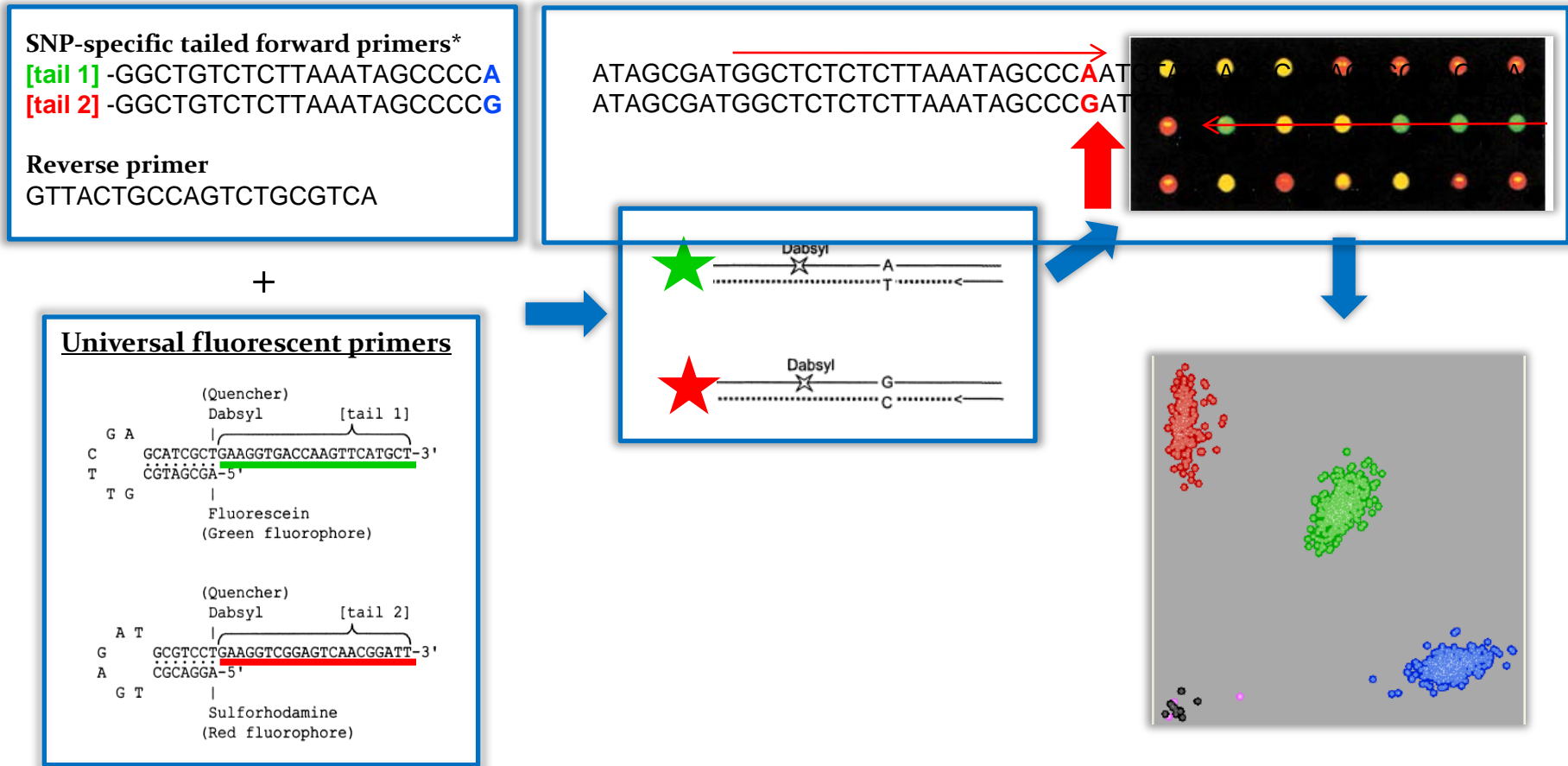


SNP discovery

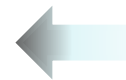
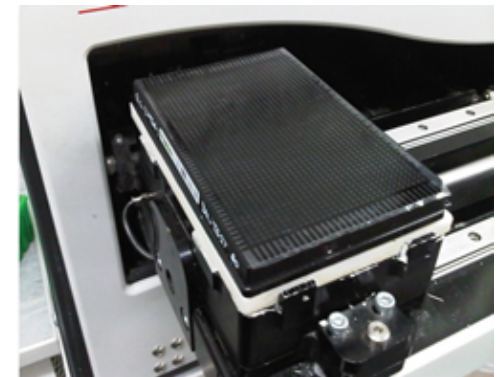
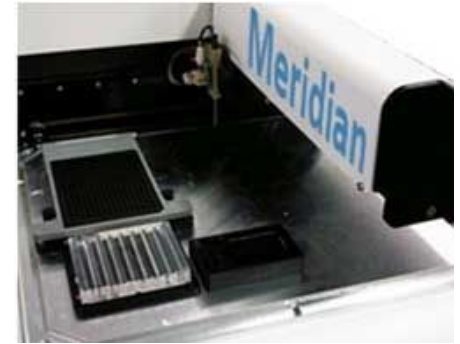
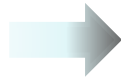
- Mean coverage per UK variety = 48X
- 500,000 SNPs identified in sequence data from 8 elite cultivars
- 1 SNP per 67bp / ~15 SNPs per kb
- ~ 95,000 = varietal SNPs



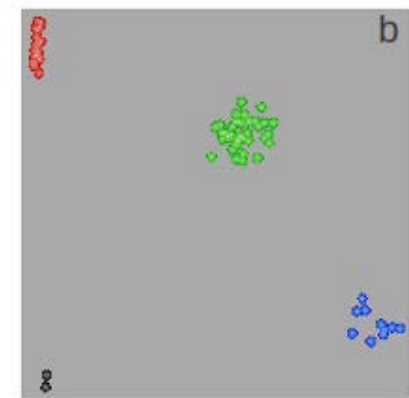
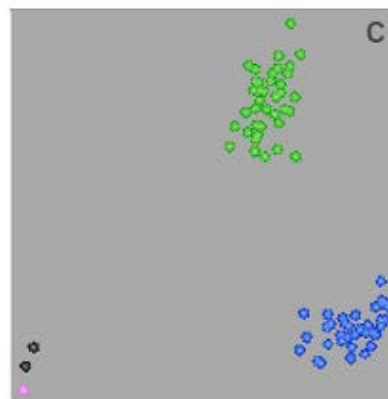
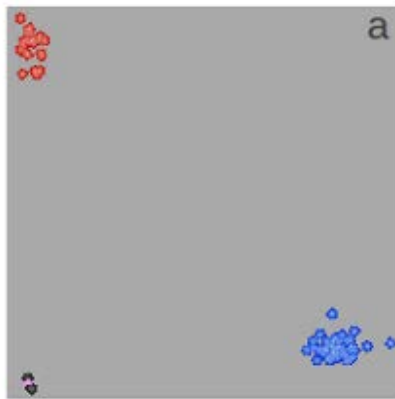
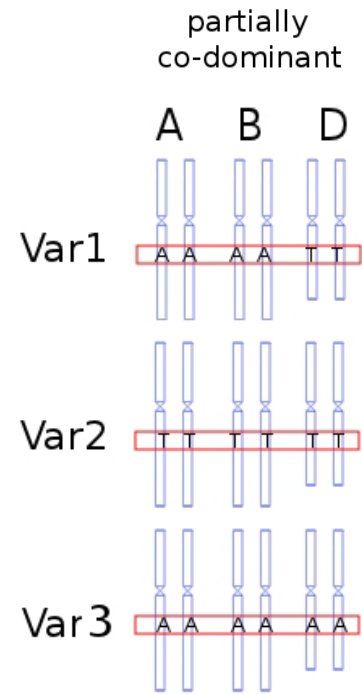
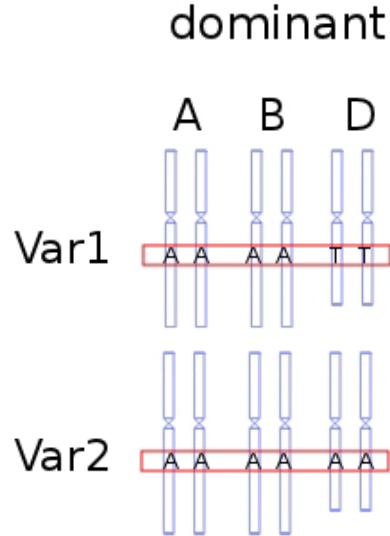
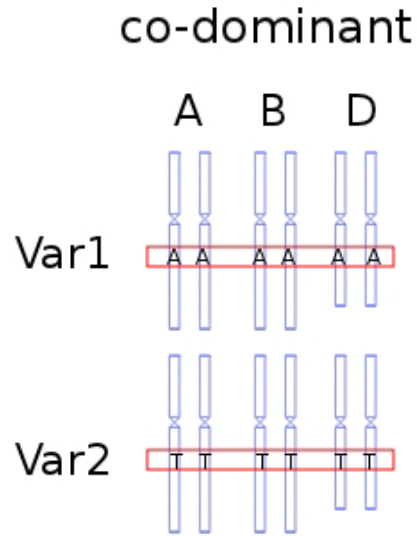
SNP validation (KASPar)



🌟 SNP validation - KASPar



SNP validation

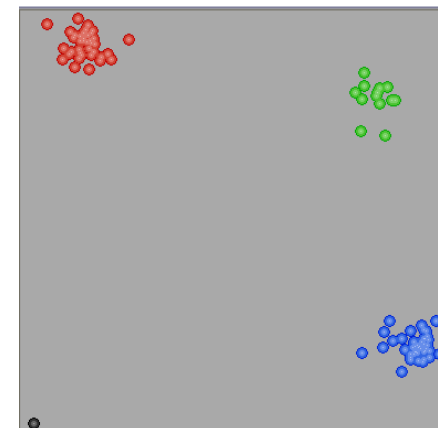
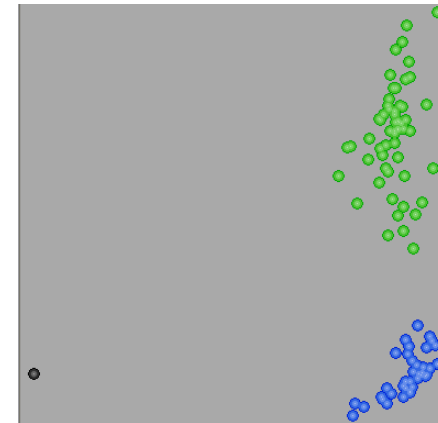


SNP validation

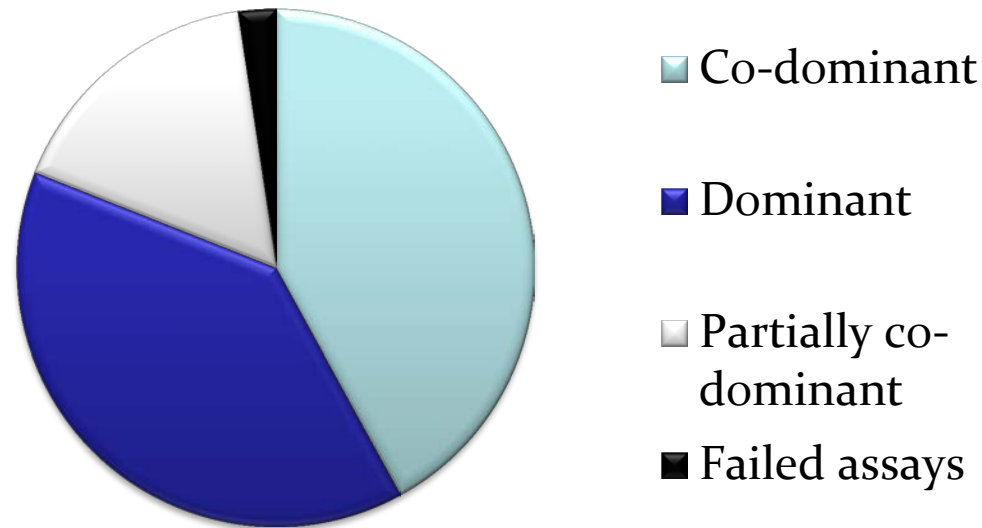
KASPar scoring on an F₄ population

	Dominant example	
	Genotype	KASPar call
Paragon	AA AA GG	A/G
Watkins 34	AA AA AA	A
Paragon x Watkins 34	AA AA AG	A/G

	Co-dominant example	
	Genotype	KASPar call
Paragon	TT ?? ??	T
Watkins 34	CC ?? ??	C
Paragon x Watkins 34	TC ?? ??	T/C



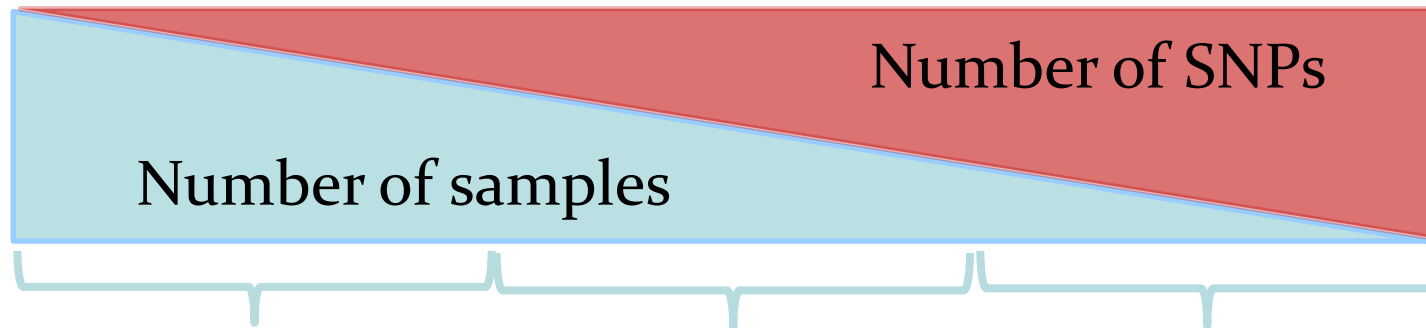
🔥 Targeted co-dominant SNP development



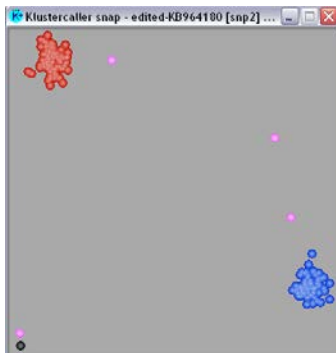
3239 out of 5392 SNP assays (60%) of those designed to date are co-dominant in some or all varietal comparisons.



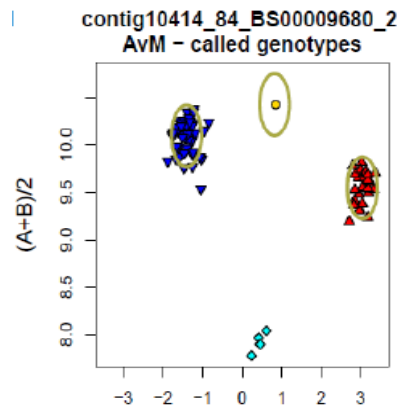
🔥 Genotyping platforms



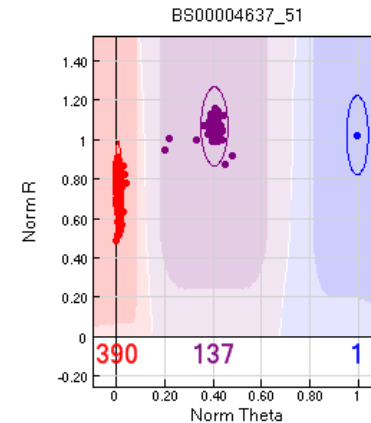
Kbioscience KASPar
~5000 SNPs



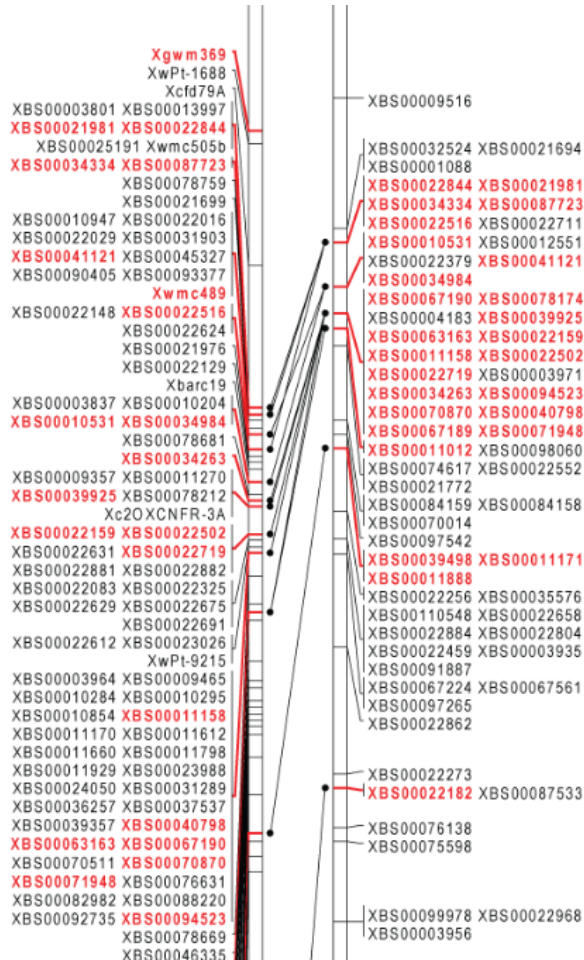
Affymetrix Axiom
~12000 SNPs



Illumina iSelect
~81000 SNPs



Genetic Maps



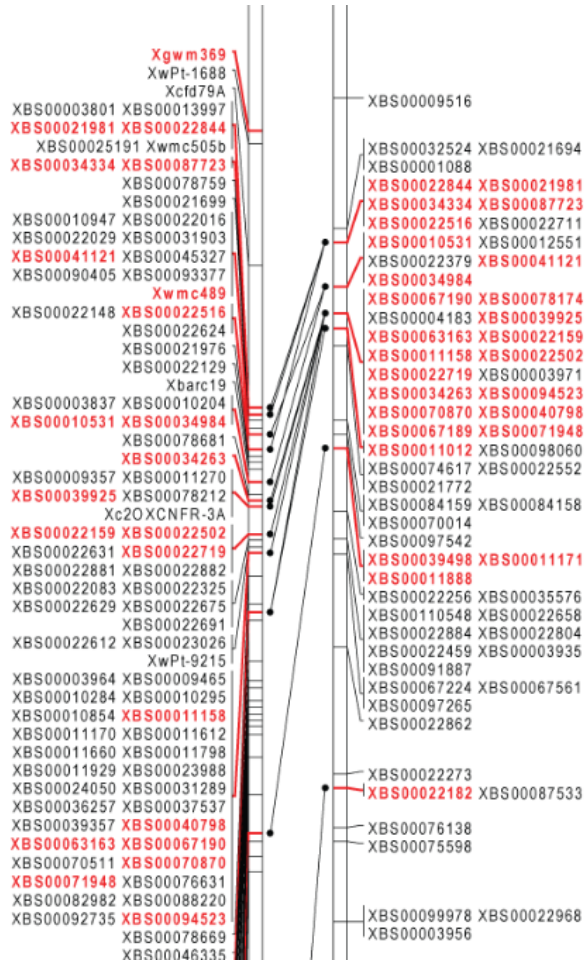
Numbers of markers on maps:

	SSR	KASPar	iSelect	Total
AxC	331	3120	9420	12871
SxR	337	1512	8082	9931

Total no. mapped markers = 18905



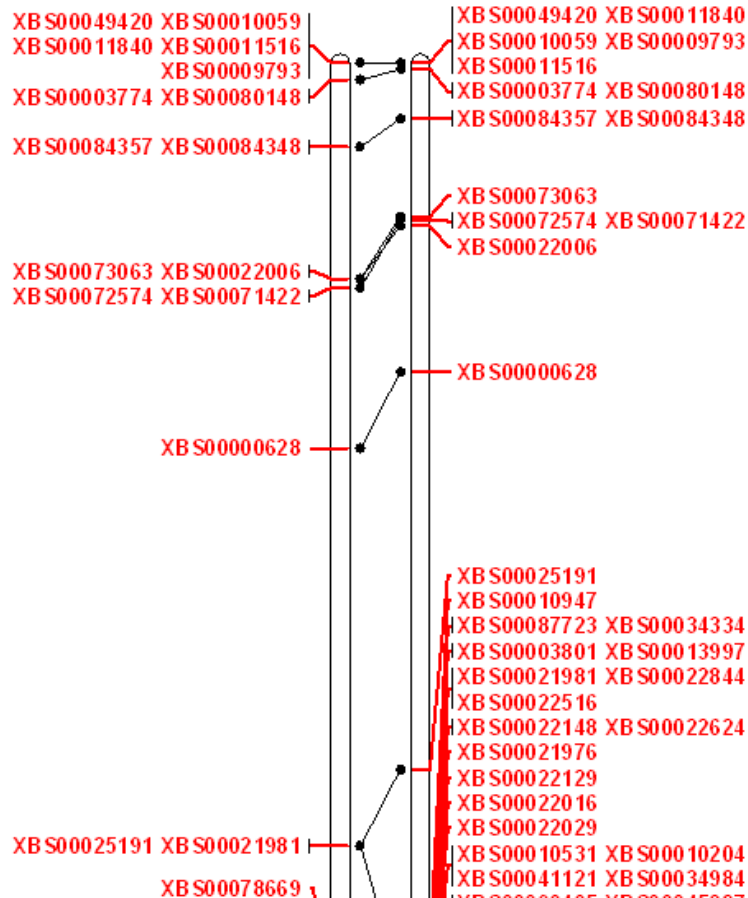
Genetic Maps



Avalon x Cadenza map:

	A	B	D	Total
1	682	1367	479	2528
2	758	929	282	1969
3	695	994	51	1740
4	473	495	105	1073
5	657	1342	242	2241
6	1035	829	118	1982
7	758	441	139	1338
Total	5058	6397	1416	12871

Genetic Maps

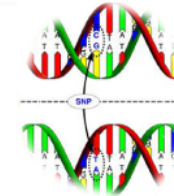


Extended Avalon x Cadenza map:

<i>Bin (cM)</i>	Standard	Extended
0	64	55
0.1-1.0	18	32
1.1-2.0	7	4
2.1-5.0	5	2
5.1-10.0	3	3
10.1-20.0	1	1
20.1-40.0	1	1

SNP database

CerealsDB



SITE PORTAL

Wheat KASPar SNP database

CEREALSDB HOME

WHEATBP HOME

FAQS

CONTACT US

Important Note

Via this web site and associated publications, we have placed all the data related to each and every SNP, including available flanking sequences (usually 120 base pairs or greater), into the public domain. We hope that by making this information freely available without any restrictions, it will encourage other wheat geneticists to do likewise so ensuring that all such markers are made public and free of use. Only by following this principle can we hope to ensure that wheat breeders across the world are in possession of the tools they require to create the new varieties that will be needed to help feed the World beyond 2050.

KASPAR SNP DATABASE

[SELECT PRIMERS](#)

[OBTAIN HAPLOTYPES](#)

[DOWNLOAD ALL DATA](#)

[DOWNLOAD AXC MAPPING DATA](#)

[BLAST CONTIGS](#)

[CONTIG INFORMATION](#)

[ALIGN TO BRACHY](#)

[BLAST WHEAT GENOMIC](#)

SNP database last updated October 1st 2012

This site has been designed with breeders in mind, and we hope that it will be easy and straightforward to use. Principally the site is aimed at those who wish to obtain information about SNP markers; e.g., the sequence upon which they are based, obtain primers used for their identification, identify the haplotypes of common UK varieties. However, there are pages that describe, in simple terms, what SNPs are and how they can be a wonderful tool for the breeder. To gather such information, please follow the links on the menu bar to the left.

Information on genetic relationships among individuals is of importance to plant breeders for line and hybrid development.

The KASPar SNP database contains

Varietal SNPs:	99945
Validated SNPs:	5388
Mapped SNPs (Avalon x Cadenza):	3029
Mapped SNPs (Savannah x Rialto):	1463
Wheat varieties studied:	169

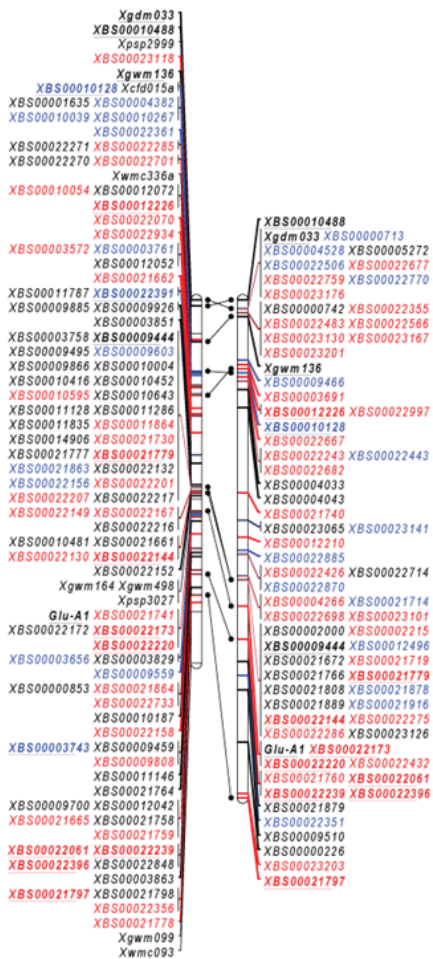
SNP database

- Primer sequences
- Contig sequences
- Map locations
- Genotype information – 101 wheat lines
- Haplotype information
- BLAST information



SNP database

Chromosome 1A



SNP name	cM	Ax cM	SxR	Forward primer (FAM)	Forward primer (VIC)	Reverse primer
BS00010488	1.7		0	CATAGTTCACACTTAAATCACAAACGTAC	ACATAGTTCACACTTAAATCACAAACGTAT	CGACGGACCCCTAAAACG
BS00023118	11.4		0	GAATGCCGTATGGCTCAGAGTAG	ATGAATGCCGTATGGCTCAGAGTAA	CCGCACTGAGTTGAACTC
BS00010128	20.5	20.99	1	CACCTCCATCACAAAGGTATACAC	CACCTCCATCACAAAGGTACAG	CCGTTGACAGCATCGTGC
BS00004382	21.1		0	CCGTTGCAATTTCTCAAGATCCC	CCGTTGCAATTTCTCAAGATCCG	CTAAAACCTCTCAGGGTGT
BS00010267	21.1		0	CTGGATACGGAACTTTTCAGAAC	AATCTGGATACGGAACTTTTCAGAACT	GGGTACGGATACATGGCC
BS00001635	21.1		0	ACCATAACACCAAGTGTGCCGT	CTACCATAACACCAAGTGTGCCGA	CAAGGGCTGTTCCTCTCA
BS00010039	21.1		0	GGTTGAATCAAGCTTTACTCTGGGA	GGTTGAATCAAGCTTTACTCTGGG	AATCTGAACTTATGATG
BS00022361	21.7		0	CAAGACCTGCGACAAGACCAG	CAAGACCTGCGACAAGACCAC	GACCGGCTCTTGCACCTG
BS00022285	24.6		0	CCAGAAGATCTCCAGCATTGGAT	CCAGAAGATCTCCAGCATTGGAA	TTGATAAGATGATGAACT
BS00022271	24.6		0	GAGGTTTTGTCATGGTCCAGGT	GAGGTTTTGTCATGGTCCAGG	GTGCGTCGCAAACTGATC
BS00022701	25.2		0	AATGTATGACTGTCCAGTCTGTG	CAATGTATGACTGTCCAGTCTGTG	GGGCTATCCACTCATAT
BS00022270	25.2		0	GGCTTGATGCGAGGATTTCTACGA	GCTTGATGCGAGGATTTCTACGG	GGACACCTTAAAACCTC
BS00012072	27.5		0	AATCATTGTTGTGCAAAAATGGCCCTG	AATCATTGTTGTGCAAAAATGGCCCTC	CCTGCAGCGGCTCTCCAT
BS00010054	27.5		0	ACACCCAGGCTGAACGCAACT	ACACCCAGGCTGAACGCAACTA	GGAGCTTTGCCGAGGGCC
BS00012226	27.5	19.81	1	GCAGCACGATGATAGTACCATCAAT	GCAGCACGATGATAGTACCATCAAC	CTGGATGATGAGAGGGG
BS00022070	31.2		0	AATGTCTCCGATGAACACGTGC	CTAATGTCTCCGATGAACACGTCA	GCCGGACAGCAAACTCTA
BS00022934	32.8		0	GATTCGTATGTTGTGCTCCG	CGATTCGTATGTTGTGCTCCT	CTCACAGCGAGAGTAGA
BS00003761	33.4		0	GCTACCAACAACAAGGGAG	GCTGCTACCAACAACAAGGGAA	TGATGGCTCTGGGCGA
BS00003572	33.4		0	CACCTGCTCTCCGCCGA	CACCTGCTCTCCGCCGT	GGAGGGAAACAGGGGG
BS00012052	33.4		0	GAAAGTTGAGCGAATTGACAGAGAAG	GAAAGTTGAGCGAATTGACAGAGAAA	TTCAAGTTGACAGGGAA
BS00021662	38.2		0	GAGAAAGACAAGCCAGTCAACC	GAGAAAGACAAGCCAGTCAACC	GCTCCCAACTCAGTCCAA
BS00022391	44.7	47.99	1	GCAAAGGTAGCAAAGACGATGATA	GCAAAGGTAGCAAAGACGATGATA	GGTGTGGGTGGGCAAT
BS00011787	44.7		0	CGTGGCAGGACAGCACT	GTGGCAGGACAGCAGC	CCCTTCTGTTTTTAAAC
BS00009926	47.1		0	CAGAAAGTTGCGAGCTTGAGGAC	CAGAAAGTTGCGAGCTTGAGGAG	GCTGGTCCGCGCAAACT
BS00009885	47.1		0	ACCGTGTGGCCCTCG	GGCTACCGTGTGGCCCTCA	CAGAGACAACAGGGCCA
BS00003851	53.2		0	TTCACTGGAAGAATCTGCAGCCTA	CACCTGGAAGAATCTGCAGCCTC	GACCTCGGAACCCAGCC
BS00010416	53.9		0	CTCCCACCATCTACCTCAGCA	CTCCCACCATCTACCTCAGCC	GGCATATTTGGCGGGCCG
BS00022207	53.9		0	CATGTATACACTTTTATTTCAACCATCC	CCATGTATACACTTTTATTTCAACCATCA	CGTGTGTTGCTCTAGTG
BS00021730	53.9		0	GTTGGTAACTCACTCCCACTCA	GGTAACTCACTCCCACTCA	GGCTGGAAGGGCAAAA
BS00010595	53.9		0	CCGCTGATGCGTAGGCC	CCGCTGATGCGTAGGCCA	AGCAGCACCAAGCAGCA
BS00009444	53.9	80.61	1	ATCTCTCTCTGACTCTCCAT	CCTCTCTCTGACTCTCCAG	CACAGTATGAGGACA
BS00021779	53.9	80.61	1	CACATGGATGACTGGCTATTAATACT	ACATGGATGACTGGCTATTAATACT	AACAACACTTTAGGAAGT
BS00011128	53.9		0	GCATATCTGCAAGCCAGGAGT	GCATATCTGCAAGCCAGGAGG	TATGGAATGATAACTCTT
BS00009603	53.9		0	CCATAATCCCTCAAACCTCAATGT	CCTAATCCCTCAAACCTCAATGT	GGTGGCTATTACAGAACT
BS00022132	53.9		0	GCAGCGCAACCTAAAACATAGGC	AGCAGCGCAACCTAAAACATAGGT	CACATACGCATAGAGCT
BS00011835	53.9		0	GAAACATTTGAGTGGTGGCAATAATC	GAAACATTTGAGTGGTGGCAATAATA	TCTCCAAATGGTGGCTGT
BS00010004	53.9		0	CACATTTCTCTCGAGTGTCC	CCACATTTCTCTCGAGTGTCT	AACAGCTCCAGTGCAT
BS00022201	53.9		0	GATTATTTGTAATCAGTACTATGCCCT	AGATTATTTGTAATCAGTACTATGCCCT	CAAAGGGTAGTAGTCTTA
BS00014906	53.9		0	GGATAGAACCAACAACCTTGATCAATTTAA	GGATAGAACCAACAACCTTGATCAATTTAG	TTGTCCCAAAGATTTGCT
BS00010452	53.9		0	GTTAAAGGGATCAATGCTCGCTGG	GTTAAAGGGATCAATGCTCGCTGG	GCACATCTTCATCTGAA
BS00022217	53.9		0	GGCTTCTGCTCGCTCAAATCTC	GGCTTCTGCTCGCTCAAATCTC	TAACGCCCTTCCCTCTT
BS00003758	53.9		0	GTCACGACATCAGAGGACATCC	GTGTCACGACATCAGAGGACATCA	TTGCAAGTAAGATCTTTG
BS00021777	53.9		0	CCATGCTTCTATACCAACCAACA	CATGCTTCTATACCAACCAACG	AGTAGCTACTTAAATC
BS00010643	53.9		0	CCTTGTGATTAACCTGGTGGG	CCTTGTGATTAACCTGGTGGGT	TGGCATATGATATGCCTG
BS00009495	53.9		0	AACTTAGCCATTTTCAGAAACAAGATG	CAACTTAGCCATTTTCAGAAACAAGATA	CTCTGATTTGGGAGCAG
BS00022351	53.9		0	GTCAATTATAGTCAAGTTCGGTGGCC	CACTAATTATAGTCAAGTTCGGTGGCA	AGGCTTATCCGGCAACT
BS00011286	53.9		0	ACCAAGCTCAATTTGGATGGCAT	CCAAGCTCAATTTGGATGGCAC	GTGCTGGAACCTCATATG
BS00009866	53.9		0	CGGAGGAGGTA AAAAGAGAG	CGGAGGAGGTA AAAAGAGAA	CTGACTGATCGGCTGTG
BS00022156	53.9		0	GGTTGTGAAAGCAATTGAGAAAGGCA	GGTTGTGAAAGCAATTGAGAAAGGCT	GATAGCCCTCGGGAATC
BS00011864	53.9		0	AGAGCAAGCTATTGTGAACCTACTA	GAGCAAGCTATTGTGAACCTACTG	CTGTTATCTTAAAACTC
BS00022149	55		0	GATCGCGAAAAGAGATCATGAGCC	GATCGCGAAAAGAGATCATGAGCC	TGAGGCGGCTCACTTGC
BS00022216	55		0	AGATCGCGGAAAAGAGGGCTT	GATCGCGGAAAAGAGGGCTC	CGGCGGGGAGATGTCT
BS00022167	55		0	GTGCATAATCTGGGCTGCCGA	GCATAATCTGGGCTGCCGG	GACACATCAACCCCAAC

Future developments

- D-genome SNPs
- Alien/progenitor SNPs
- ITMI mapping population
- Identify core set of ~ 1,000 evenly distributed SNPs
- Add current recommended list varieties to screening panels



Acknowledgements

Cereal Genomics Group:

Keith Edwards, Amanda Burridge, Gary Barker, Paul Wilkinson, Mark Winfield

Bristol Genomics Facility:

Jane Coghill, Christy Waterfall

LoLa/Crop club partners:

Graham Moore, Simon Griffiths, Cristobal Uauy (JIC); Peter Shewry, Malcolm Hawkesford (Rres); Ian King, Julie King, John Faulkes (Nottingham); Andy Greenland (NIAB)

Breeders:

Peter Jack (RAGT), Simon Berry (Limagrain), Peter Werner (KWS), James Melichar (Syngenta)

Kbioscience:

Jane McDougall, Rhian Gwilliam, Phil Robinson

BBSRC for funding

