TGAC
The Genome Analysis Centre™
BBSRC
Greater Norwich Development Partnership

John Innes Centre

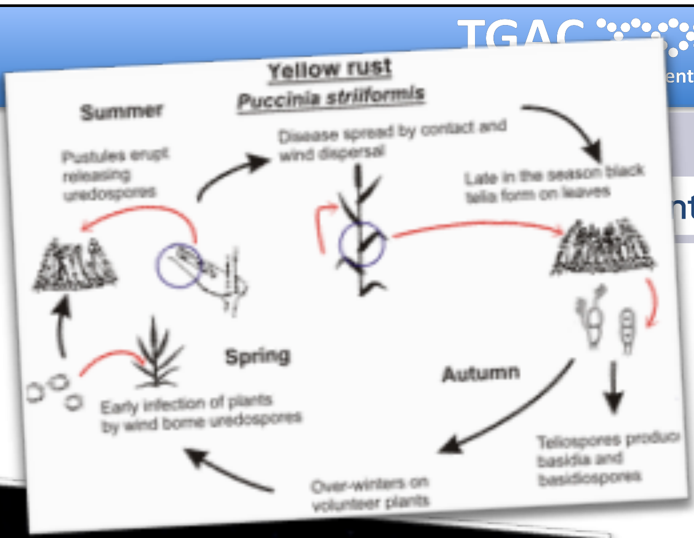
Using NGS-enabled genetics to improve marker selection and design in hexaploid wheat (*Yr15*)

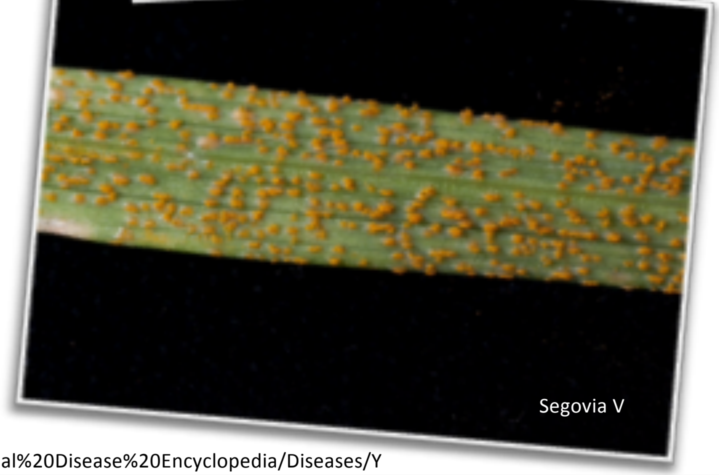
Ricardo H. Ramírez-González

Wheat Yellow Rust

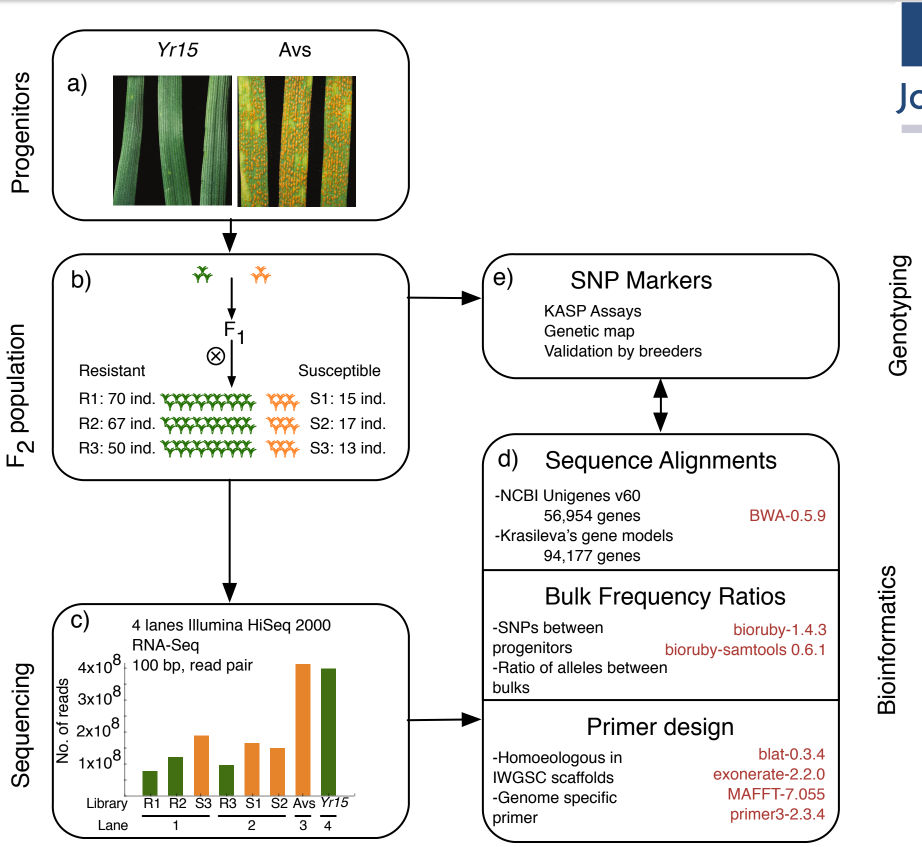
- *Puccinia striiformis.*
- Fungus.
- Traditionally controlled by resistance genes (for example, *Yr15*).

- *Yr15* locus is an introgression from *T. dicoccoides*





Cereal disease http://www.hgca.com/minisite_manager.output/3625/3625/Cereal%20Disease%20Encyclopedia/Diseases/Yellow%20(Stripe)%20Rust.mspx?minisiteId=26



Progenitors

a)

Yr15 Avs

F_2 population

b)

F_1

⊗

Resistant

Susceptible

R1: 70 ind.

R2: 67 ind.

R3: 50 ind.

S1: 15 ind.

S2: 17 ind.

S3: 13 ind.

Sequencing

c)

4 lanes Illumina HiSeq 2000

RNA-Seq

100 bp, read pair

Genotyping

e) SNP Markers

KASP Assays

Genetic map

Validation by breeders

Bioinformatics

d) Sequence Alignments

-NCBI Unigenes v60

56,954 genes

-Krasileva's gene models

94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors

-Ratio of alleles between bulks

bioruby-1.4.3

bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds

-Genome specific primer

blat-0.3.4

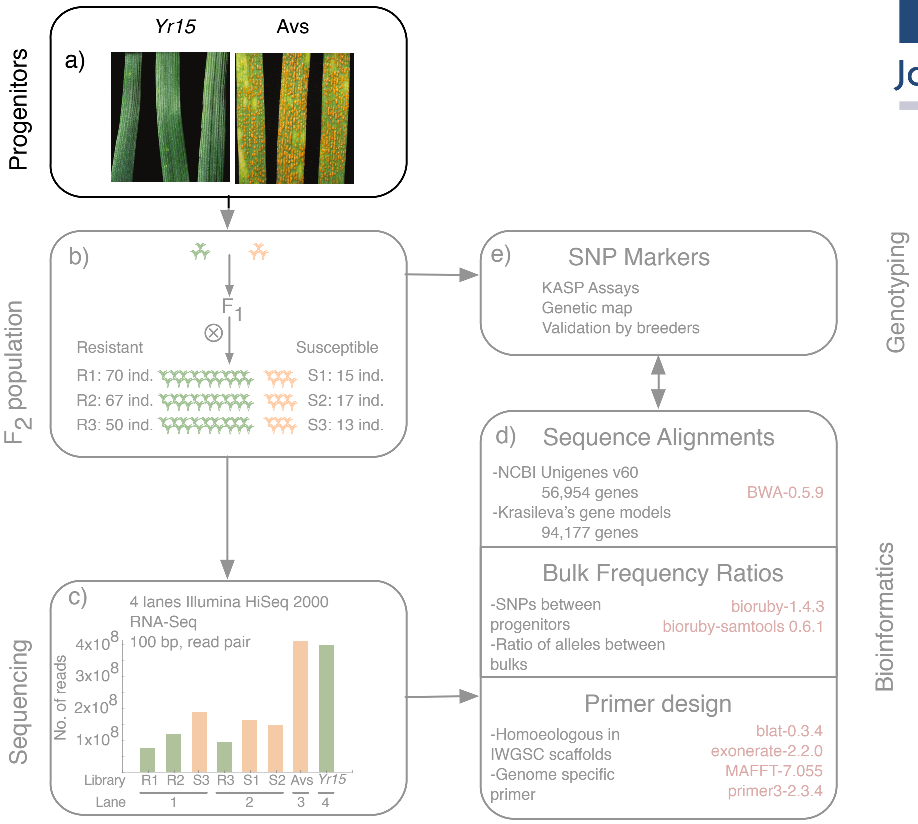
exonerate-2.2.0

MAFFT-7.055

primer3-2.3.4

TGAC
The Genome Analysis Centre™

John Innes Centre



Resistant
R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible
S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

SNP Markers

KASP Assays
Genetic map
Validation by breeders

Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

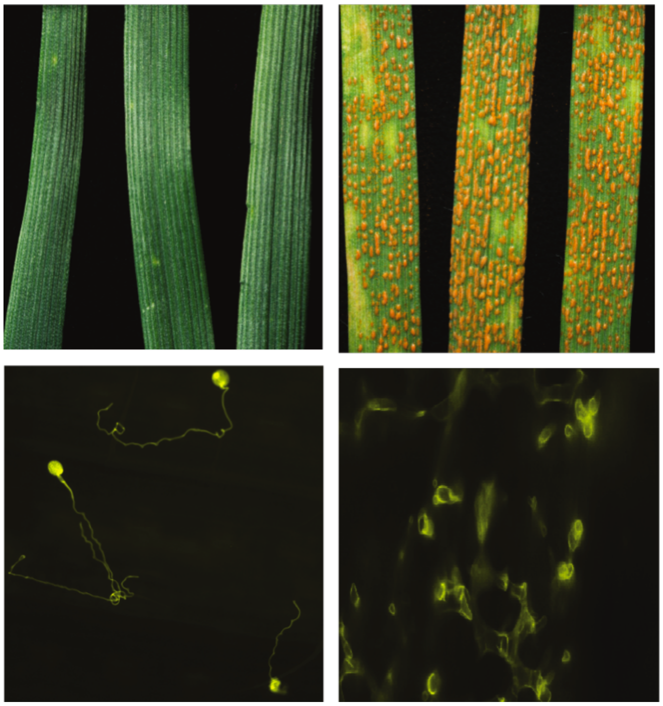
-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Parental Plants

TGAC The Genome Analysis Centre™

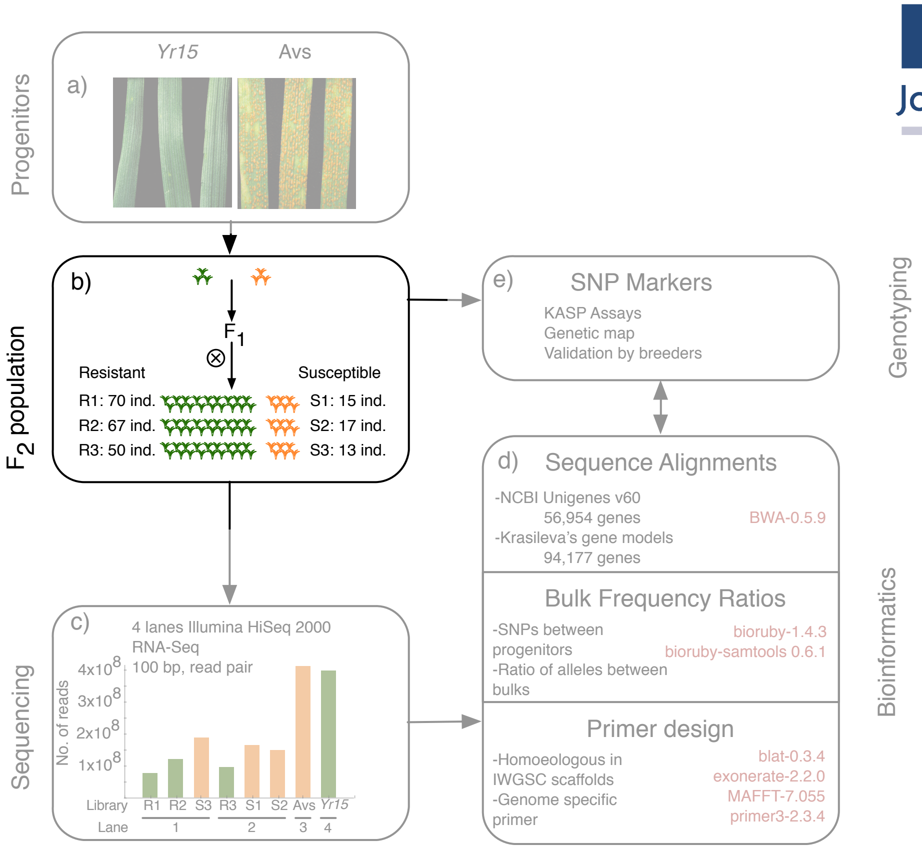
John Innes Centre



Avocet+*Yr15* *

Avocet

*Isogenic line developed by the University of Sydney



Resistant

R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible

S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

SNP Markers

KASP Assays
Genetic map
Validation by breeders

Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

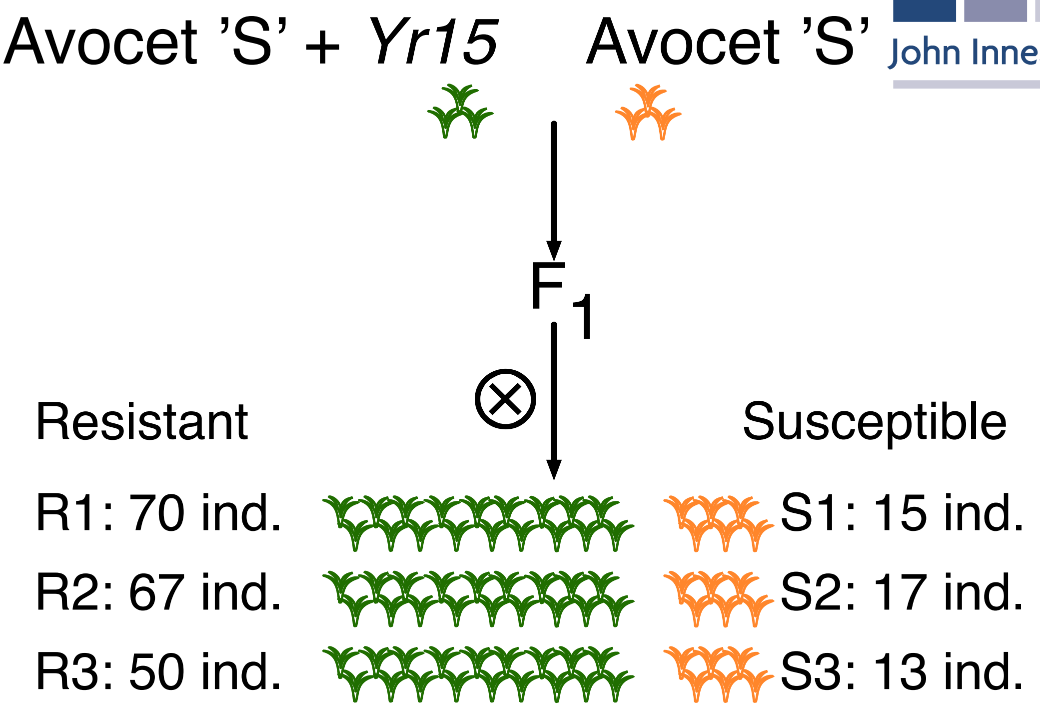
bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

F_2 population



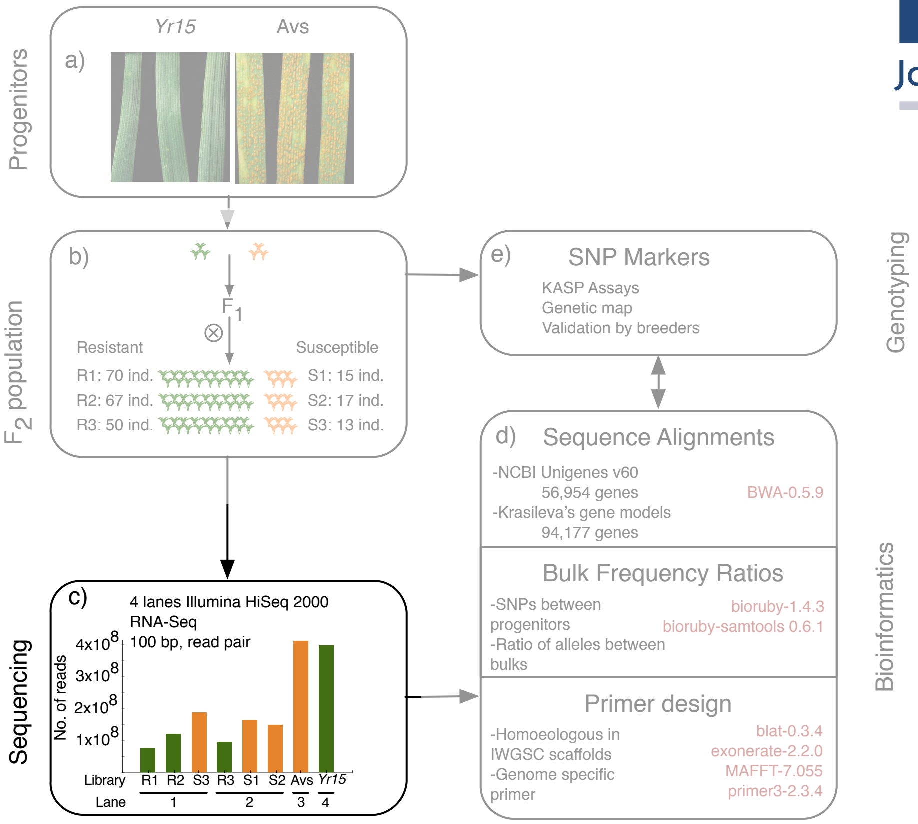
Avocet 'S' + *Yr15*   Avocet 'S'

F_1

⊗

Resistant	Susceptible
R1: 70 ind.	S1: 15 ind.
R2: 67 ind.	S2: 17 ind.
R3: 50 ind.	S3: 13 ind.

Expected segregation: 3 resistant : 1 susceptible
χ^2 P = 0.049; 187 resistant and 45 susceptible F_2 plants



F_2 population

b)

F_1

Resistant

Susceptible

R1: 70 ind.

R2: 67 ind.

R3: 50 ind.

S1: 15 ind.

S2: 17 ind.

S3: 13 ind.

Sequencing

c) 4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

Genotyping

e) SNP Markers

KASP Assays
Genetic map
Validation by breeders

Bioinformatics

d) Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

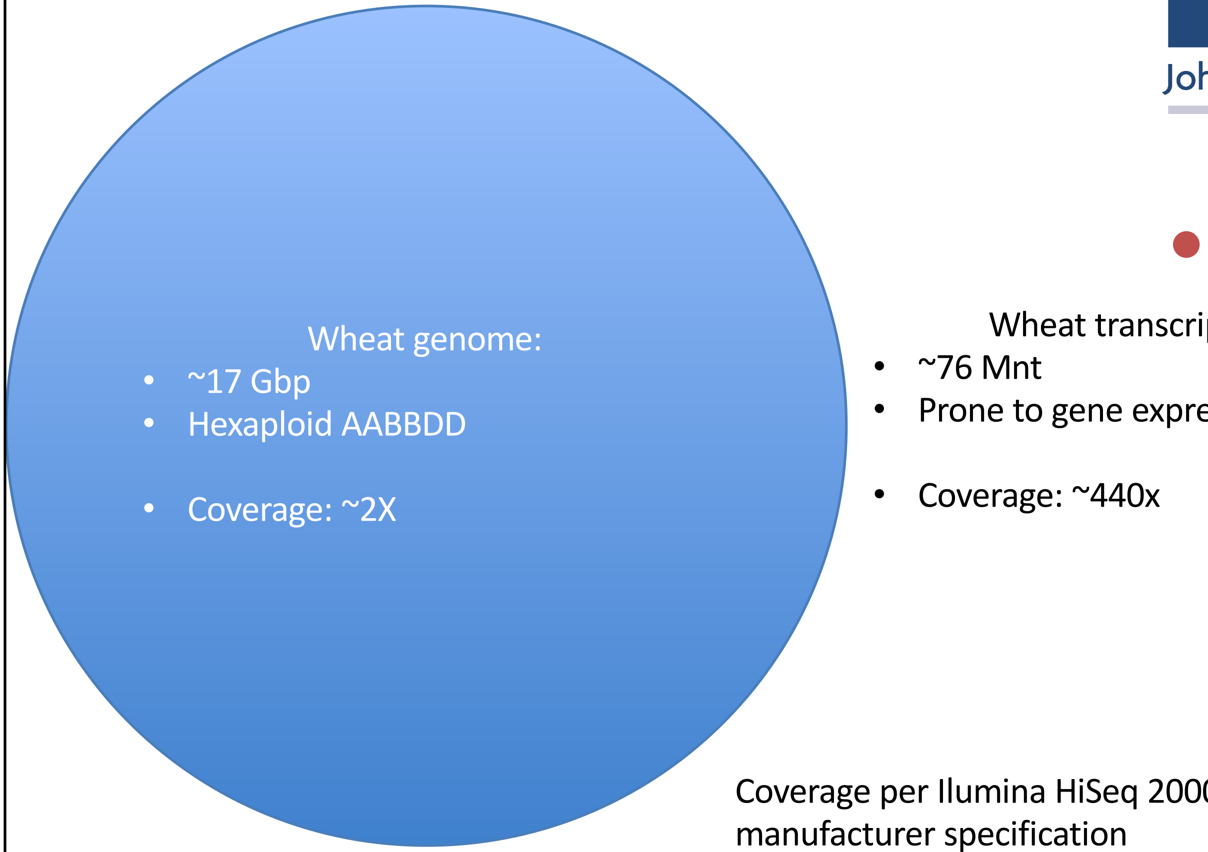
bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Transcriptome size



Wheat genome:

- ~17 Gbp
- Hexaploid AABBDD

- Coverage: ~2X

Wheat transcriptome:

- ~76 Mnt
- Prone to gene expression bias

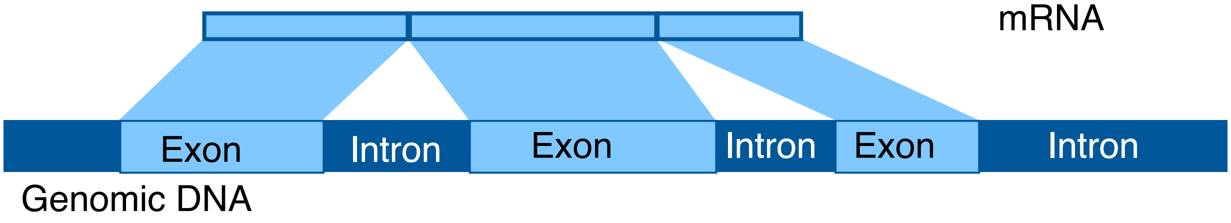
- Coverage: ~440x

Coverage per Ilumina HiSeq 2000 lane per manufacturer specification

RNA-Seq

TGAC
The Genome Analysis Centre™

John Innes Centre

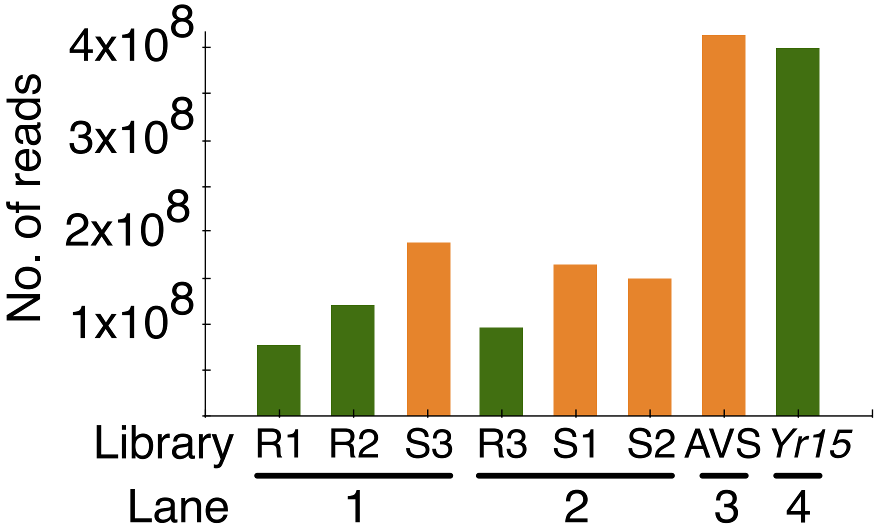


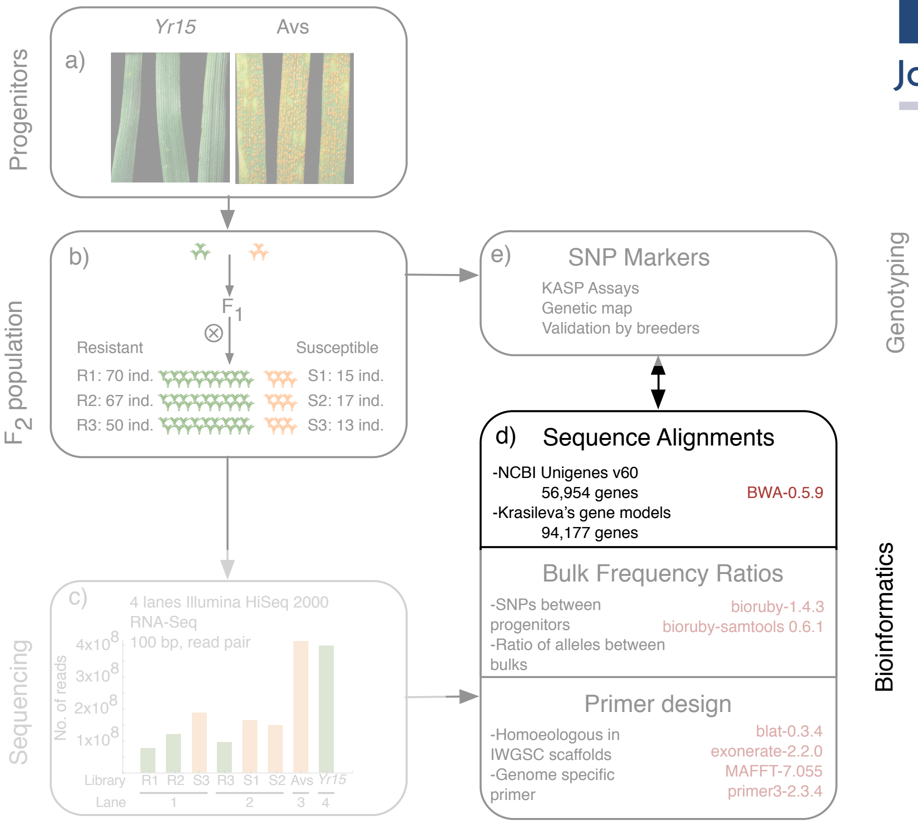
RNA-Seq

TGAC The Genome Analysis Centre™

John Innes Centre

4 lanes Illumina HiSeq 2000
100 bp, read pair





Resistant

R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible

S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

SNP Markers

KASP Assays
Genetic map
Validation by breeders

Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

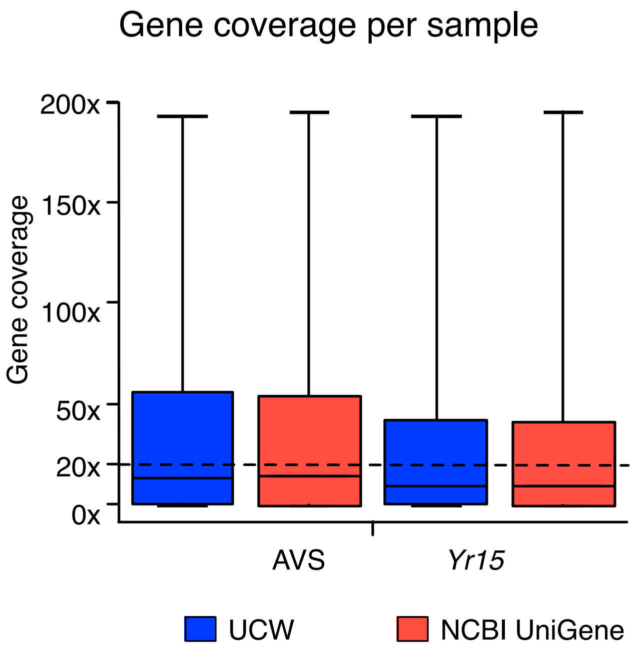
bioruby-1.4.3
bioruby-samtools 0.6.1

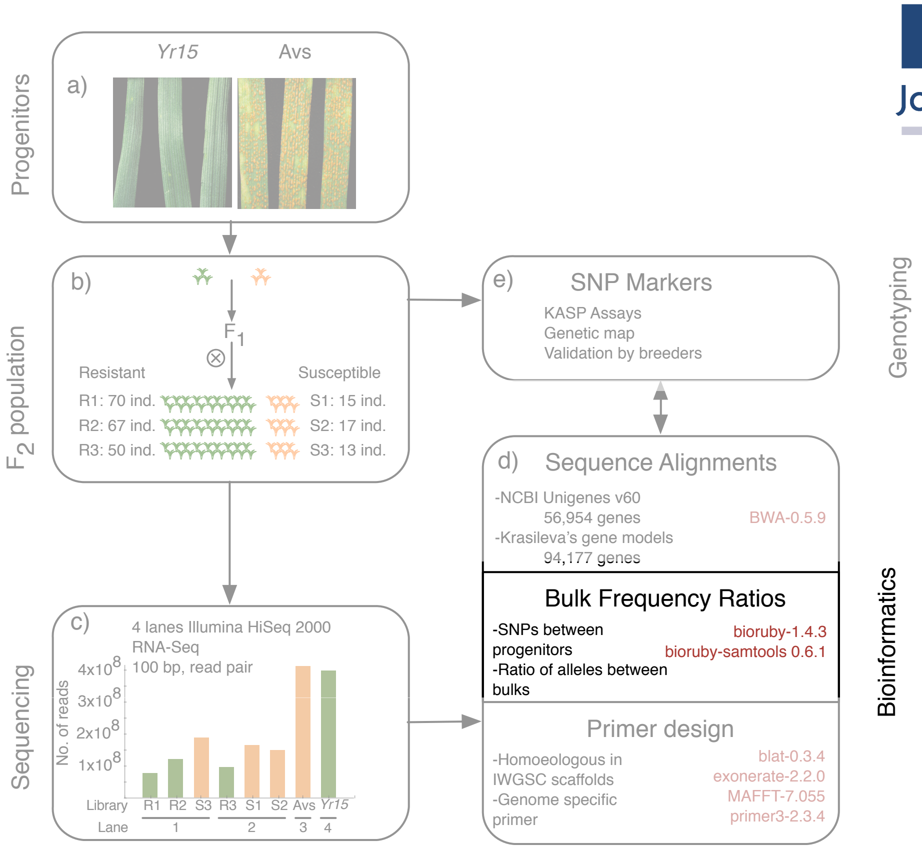
Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Sequence alignment





Resistant

R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible

S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

SNP Markers

KASP Assays
Genetic map
Validation by breeders

Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Bulk Frequency Ratios

Genomic Sequence

Avocet S ✗ ✓
1AS c t G t g G g a
1BS c t T t g G g a
1DS c t G t g G g a

Avocet S + *Yr15*
1AS c t G t g G g a
1BS c t T t g **A** g a
1DS c t G t g G g a

Bulk Frequency Ratios

Genomic Sequence

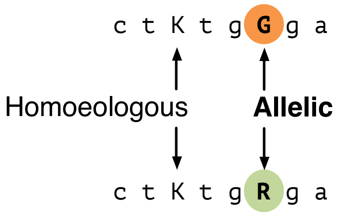
Avocet S ✗ ✓

```
1AS   c t G t g G g a
1BS   c t T t g G g a
1DS   c t G t g G g a
```

Avocet S + *Yr15*

```
1AS   c t G t g G g a
1BS   c t T t g A g a
1DS   c t G t g G g a
```

Consensus from parental



```
c t K t g G g a
```

Homoeologous **Allelic**

```
c t K t g R g a
```

Bulk Frequency Ratios

The Genome Analysis Centre™

John Innes Centre

Genomic Sequence

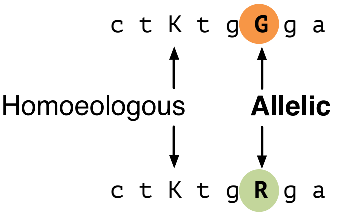
Avocet S ✗ ✓

```
1AS   c t G t g G g a
1BS   c t T t g G g a
1DS   c t G t g G g a
```

Avocet S + *Yr15*

```
1AS   c t G t g G g a
1BS   c t T t g A g a
1DS   c t G t g G g a
```

Consensus from parental



c t K t g **G** g a

Homoeologous

Allelic

c t K t g **R** g a

Susceptible bulk

```
Position       181   184
Reference  c t G t g G g a
           . . . . . . . .
           . . T . . . . .
           . . T . . . . .
           . . . . . . . .
           . . . . . A . .
           . . T . . . . .
           . . . . . . . .
           . . T . . . . .
```

Resistant Bulk

```
     181   184
 c t G t g G g a
 . . . . . . . .
 . . T . . A . .
 . . T . . A . .
 . . . . . . . .
 . . . . . A . .
 . . T . . A . .
 . . . . . A . .
 . . T . . A . .
```

SNP Index 184A:

$$\frac{1}{8} = 0.125 \qquad \frac{6}{8} = 0.75$$

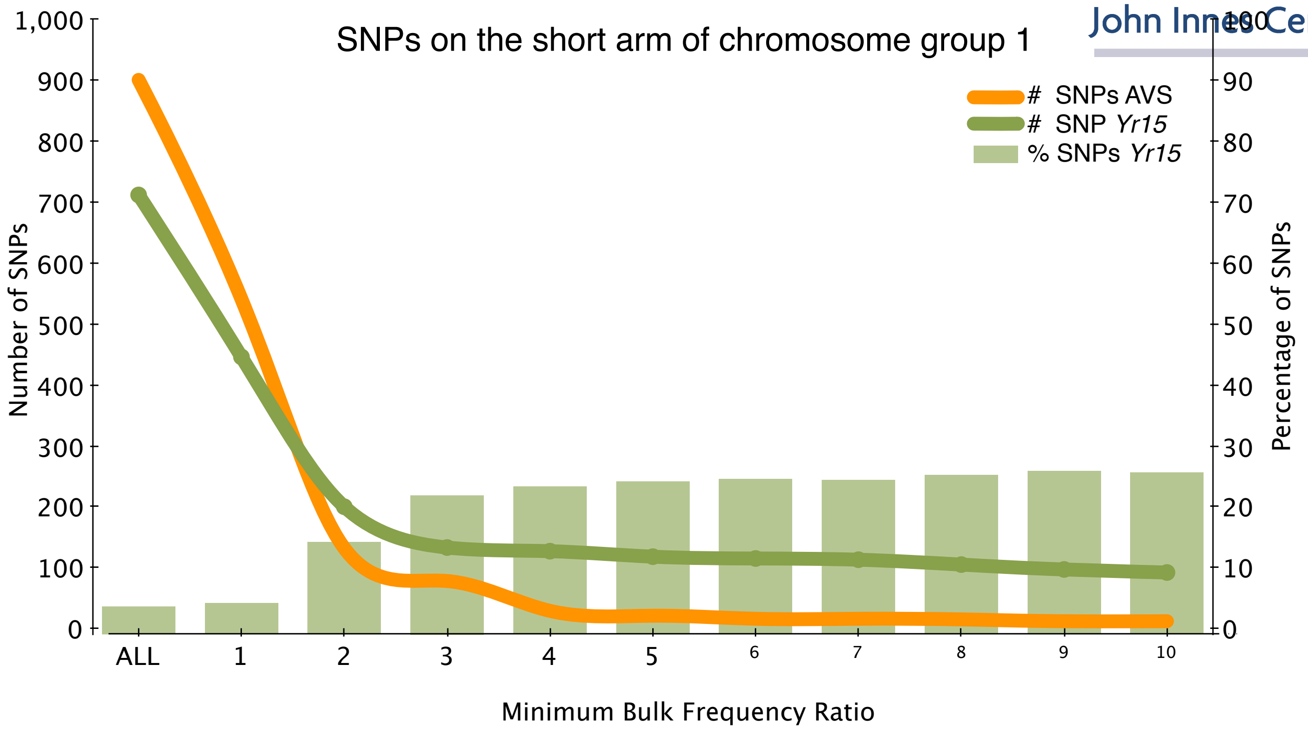
Bulk Frequency Ratio:

$$\frac{0.75}{0.125} = 6$$

Bulk Frequency Ratios

TGAC
The Genome Analysis Centre™

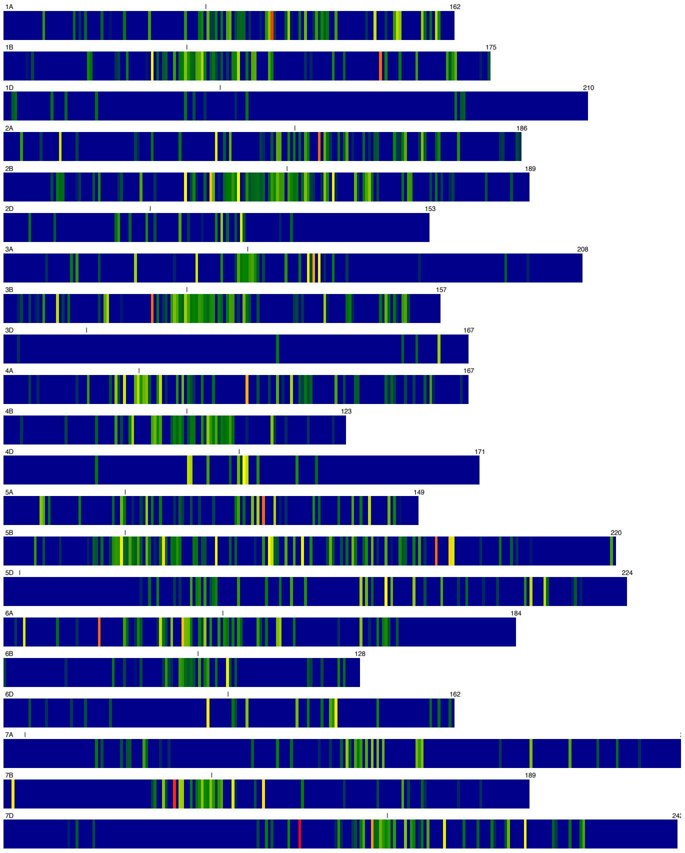
John Innes Centre

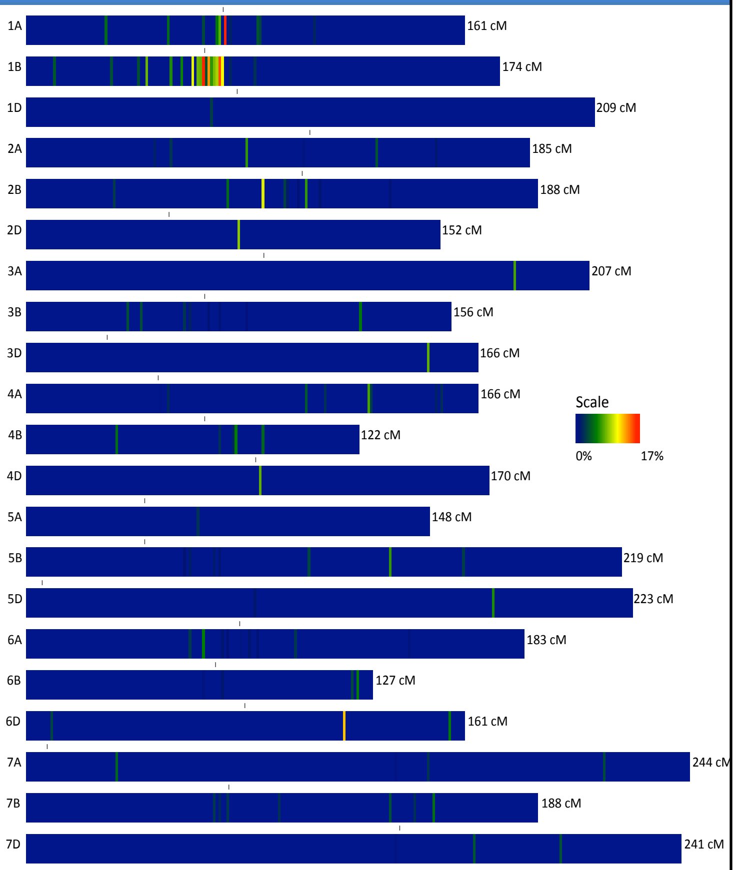


All SNPs

SNPs BFR>6

TGAC
The Genome Analysis Centre™



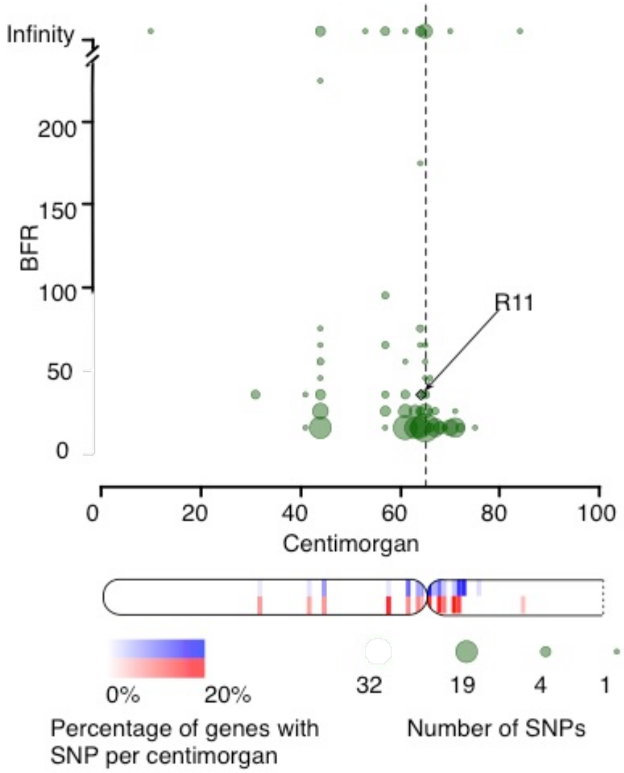


S. Wang, et al (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 SNP array. *Plant Biotechnology Journal.*

BFRs near the 1B centromere

TGAC The Genome Analysis Centre™

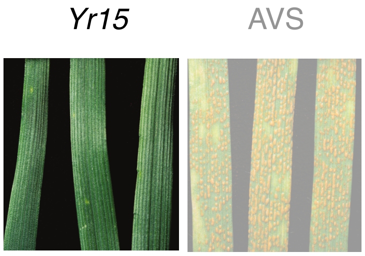
John Innes Centre



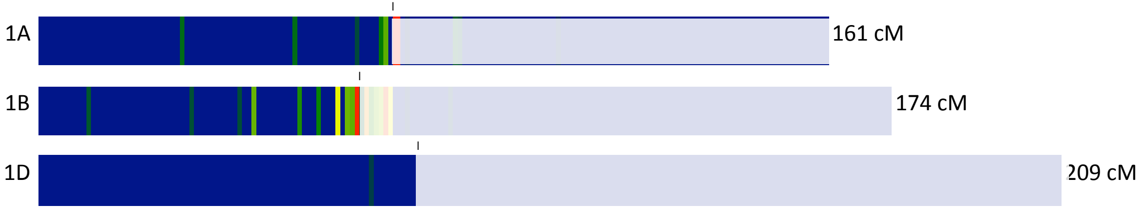
BFRs of SNPs mapping to chromosome 1B

Selection criteria

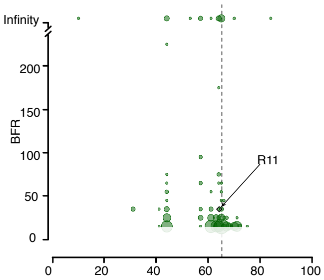
- Origin (*Yr15*)



- Short arm Chromosome group 1



- BFR > 6

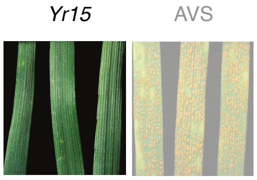


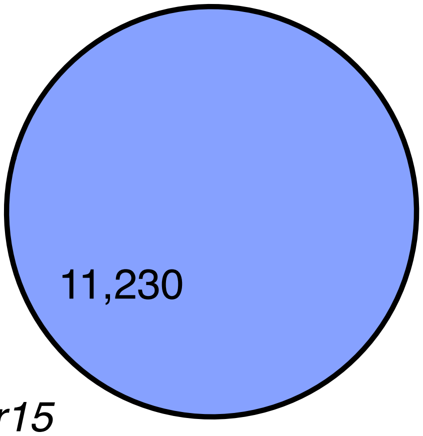
Candidate selection

TGAC
The Genome Analysis Centre™

John Innes Centre

UCW gene models





From *Yr15*
(11,230)

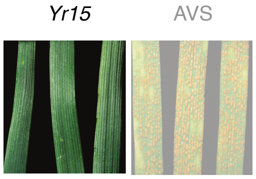
Putative genes with SNP: 16,022 (17.01%)

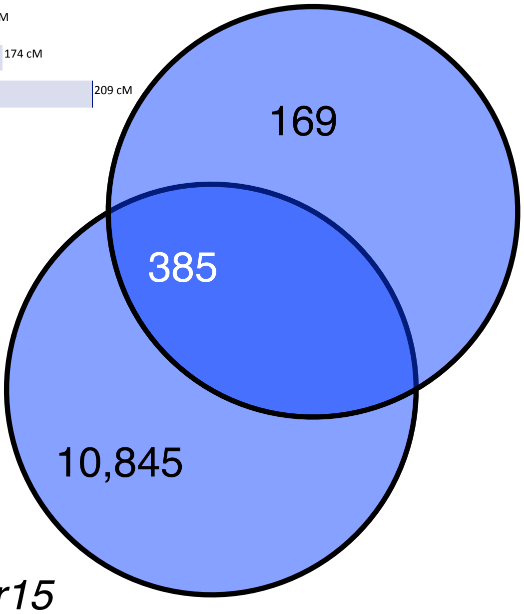
Candidate selection

TGAC
The Genome Analysis Centre™

John Innes Centre

UCW gene models





Group 1S (554)

169

385

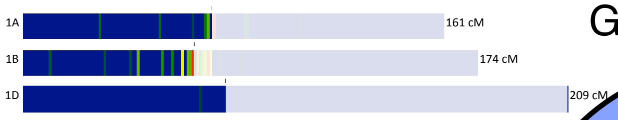
10,845

From *Yr15*
(11,230)

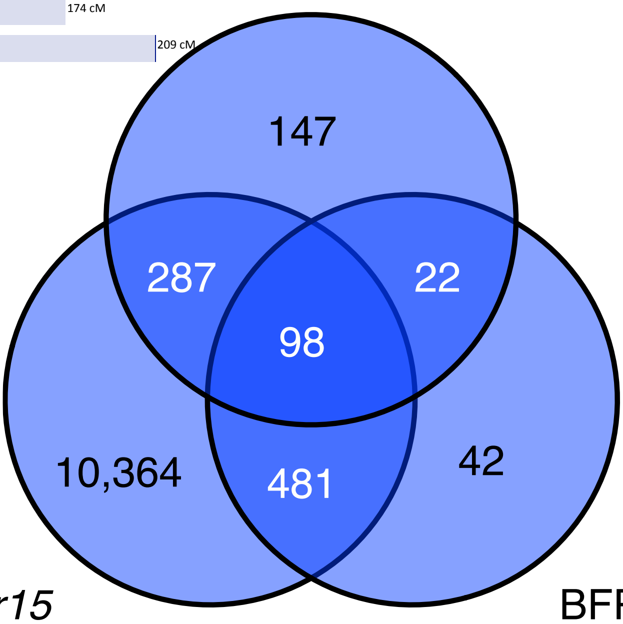
Putative genes with SNP: 16,022 (17.01%)

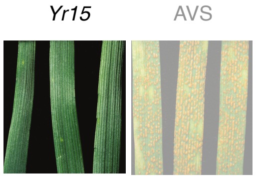
Candidate selection

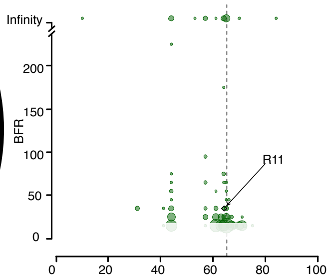
UCW gene models



Group 1S (554)



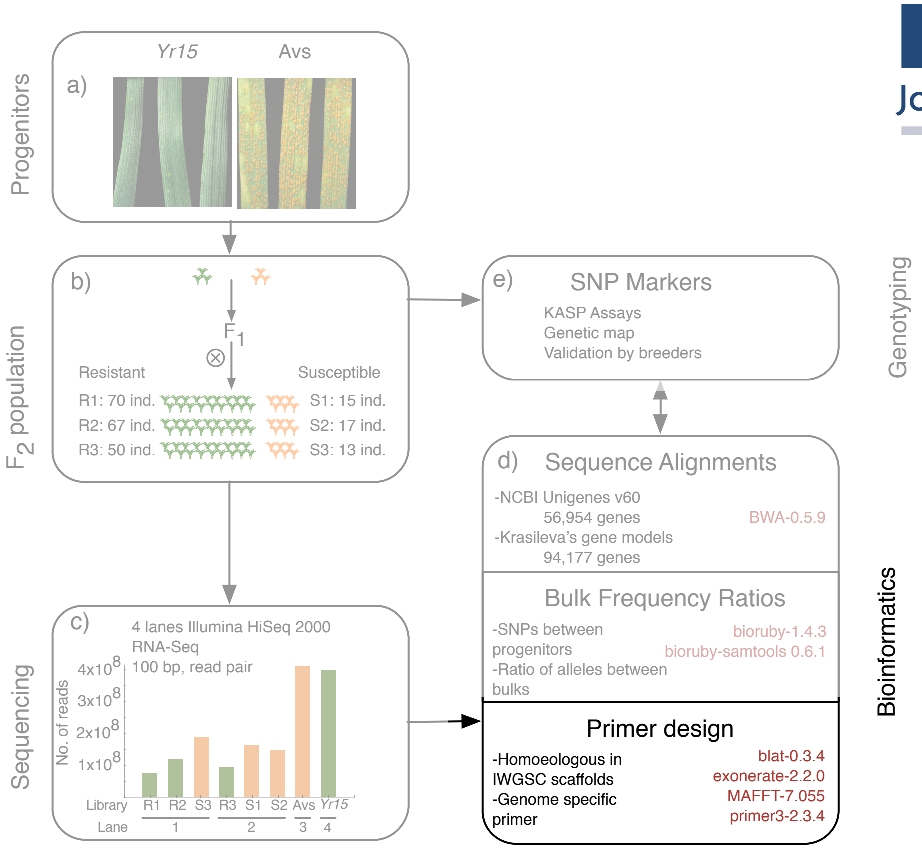




From *Yr15*
(11,230)

BFR > 6
(643)

Putative genes with SNP: 16,022 (17.01%)



Resistant
R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible
S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

e) SNP Markers
KASP Assays
Genetic map
Validation by breeders

d) Sequence Alignments
-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes
BWA-0.5.9

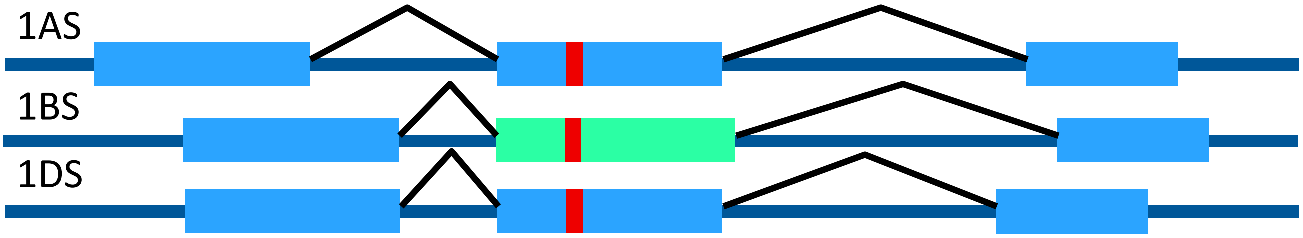
Bulk Frequency Ratios
-SNPs between progenitors
-Ratio of alleles between bulks
bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design
-Homoeologous in IWGSC scaffolds
-Genome specific primer
blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Target SNP in 1B

TGAC The Genome Analysis Centre™

John Innes Centre

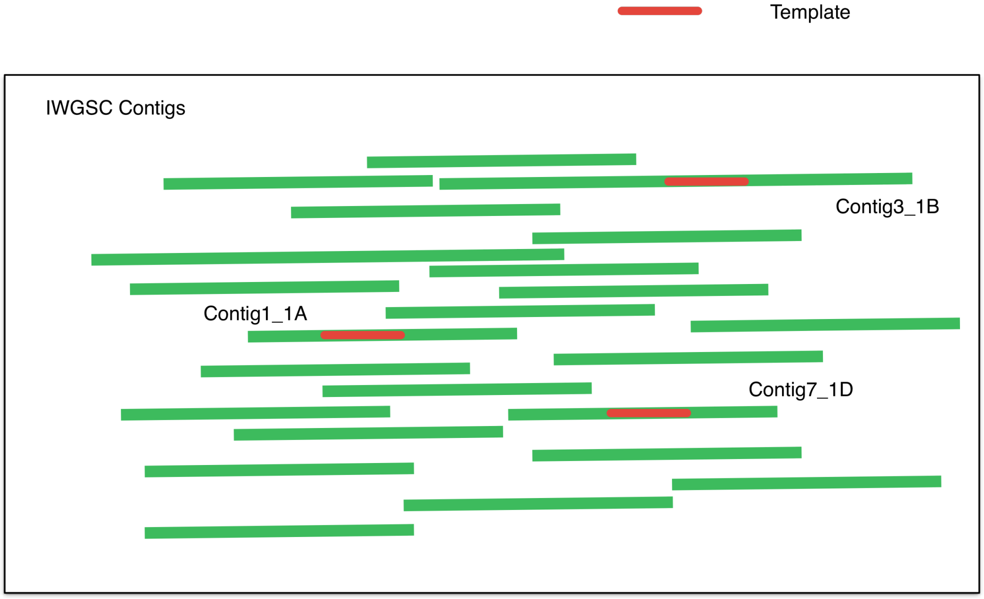


PolyMarker: Input

SNP-1 A cgcattt**G**cgcgYgcgataccggcgcctKtg**G**gaatatttgcagcgaaggcgtg

SNP-1 B cgcattt**A**cgcgYgcgataccggcgcctKtg**A**gaatatttgcagcgaaggcgtg

PolyMarker: Search sequence



PolyMarker: Local alignment

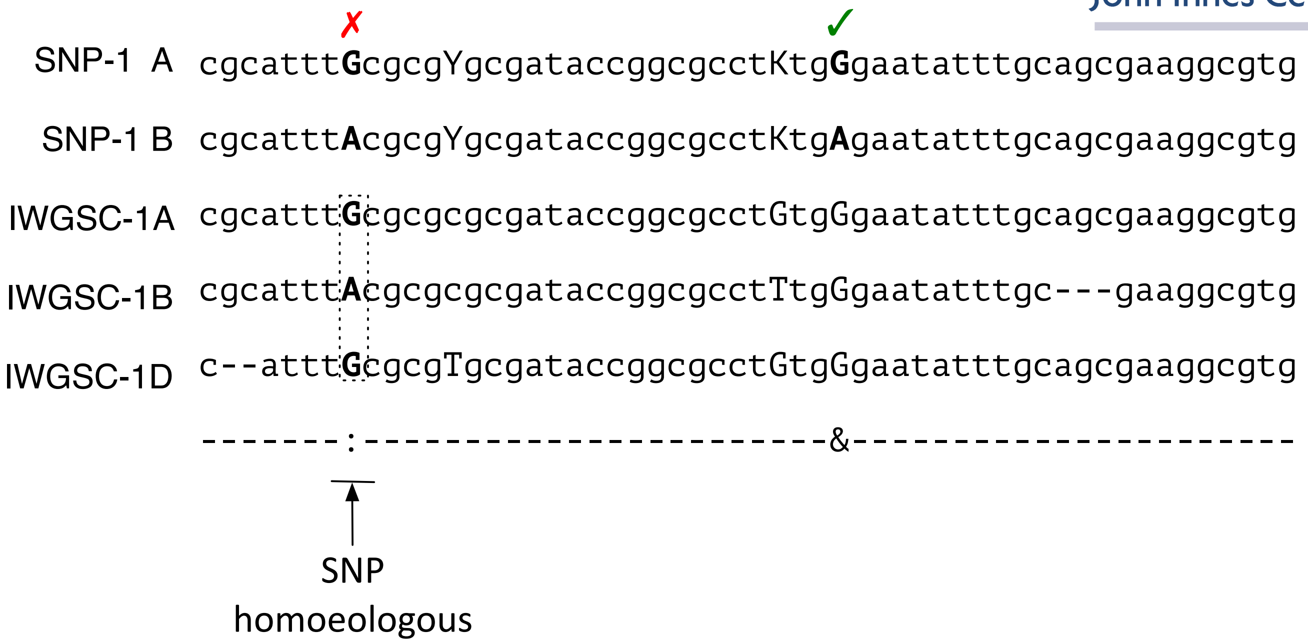
SNP-1 A	cgcattt**G**cgcgYgcgataccggcgcctKtg**G**gaatatttgcagcgaaggcgtg
SNP-1 B	cgcattt**A**cgcgYgcgataccggcgcctKtg**A**gaatatttgcagcgaaggcgtg
IWGSC-1A	cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B	cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D	c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg

PolyMarker: Candidate SNP

```
                                          ✓
SNP-1  A    cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B     cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A    cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B    cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D    c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
            -------------------------------&----------------------
```

SNP
non-homoeologous

PolyMarker: Candidate SNP



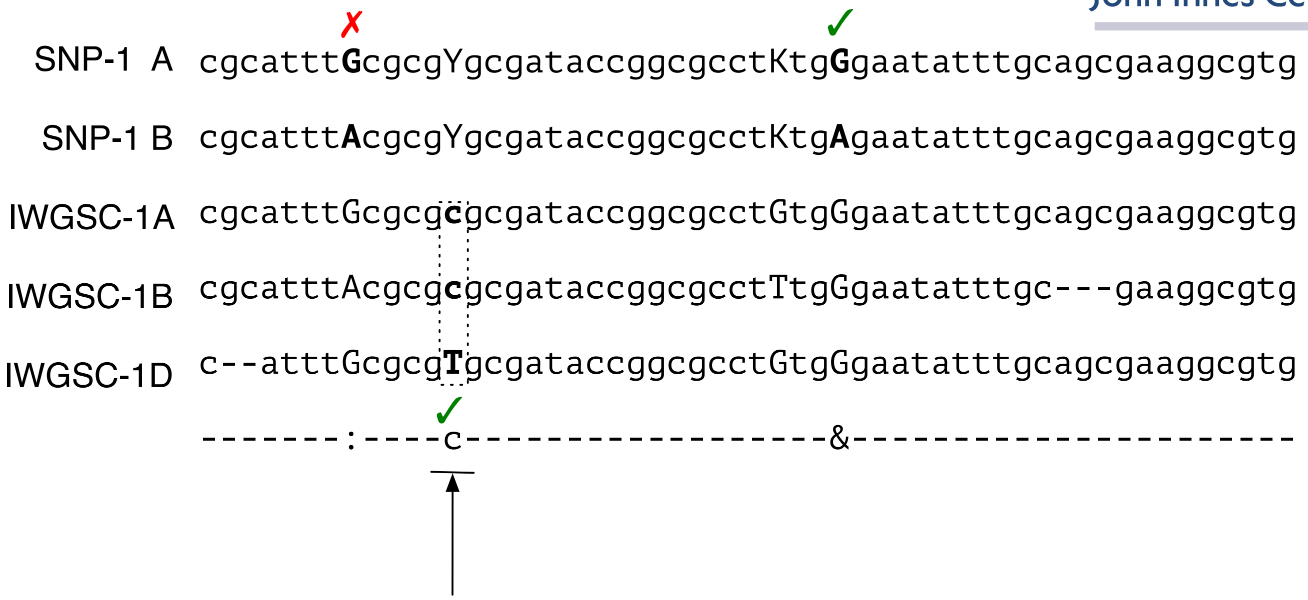
```
                   ✗                       ✓
SNP-1  A   cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B    cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A   cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B   cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D   c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
           -------:-----------------------&----------------------
```

SNP
homoeologous

PolyMarker: Genome Semi-Specific

TGAC The Genome Analysis Centre™

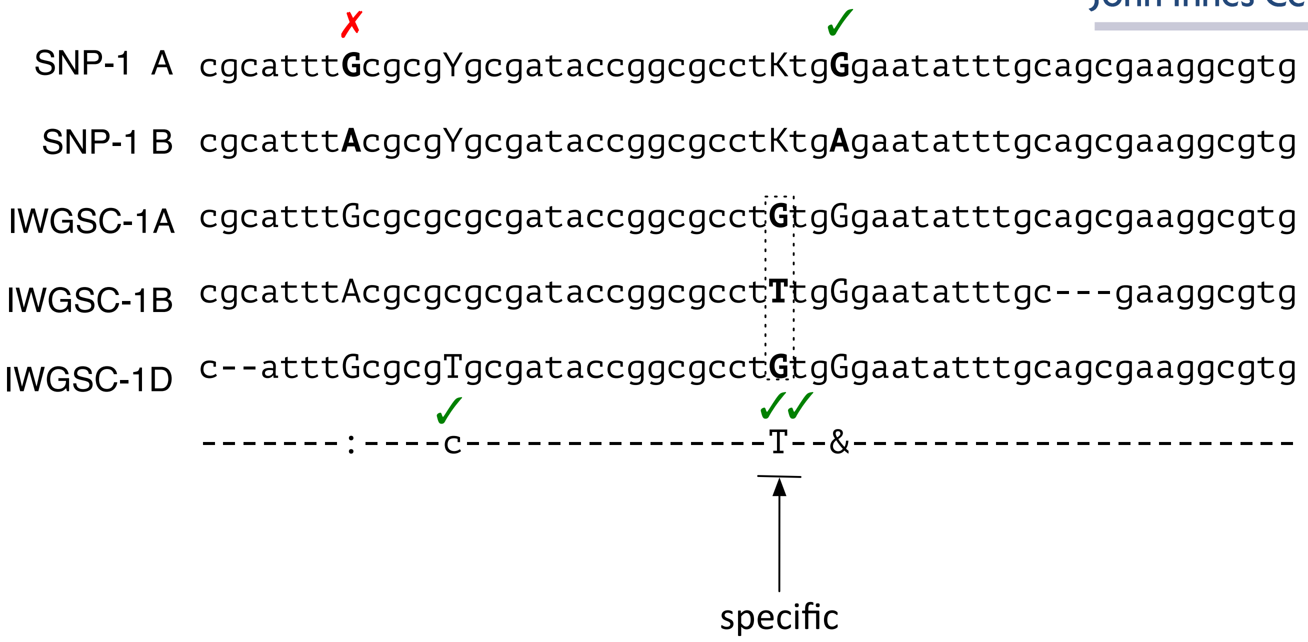
John Innes Centre



```
                 ✗                       ✓
SNP-1  A  cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B   cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A  cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B  cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D  c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
                      ✓
          -------:----c------------------&----------------------
```

semi-specific

PolyMarker: Genome specific

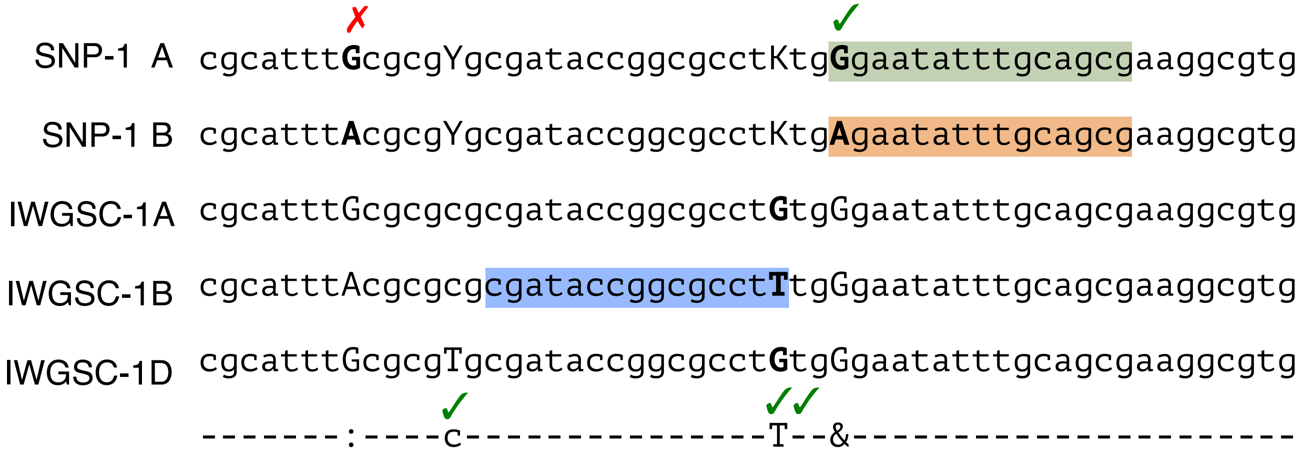


```
                 ✗                       ✓
SNP-1  A    cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B     cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A    cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B    cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D    c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
                        ✓               ✓✓
            -------:----c---------------T--&----------------------
```

specific

PolyMarker: Selected primer

Tested with Primer3



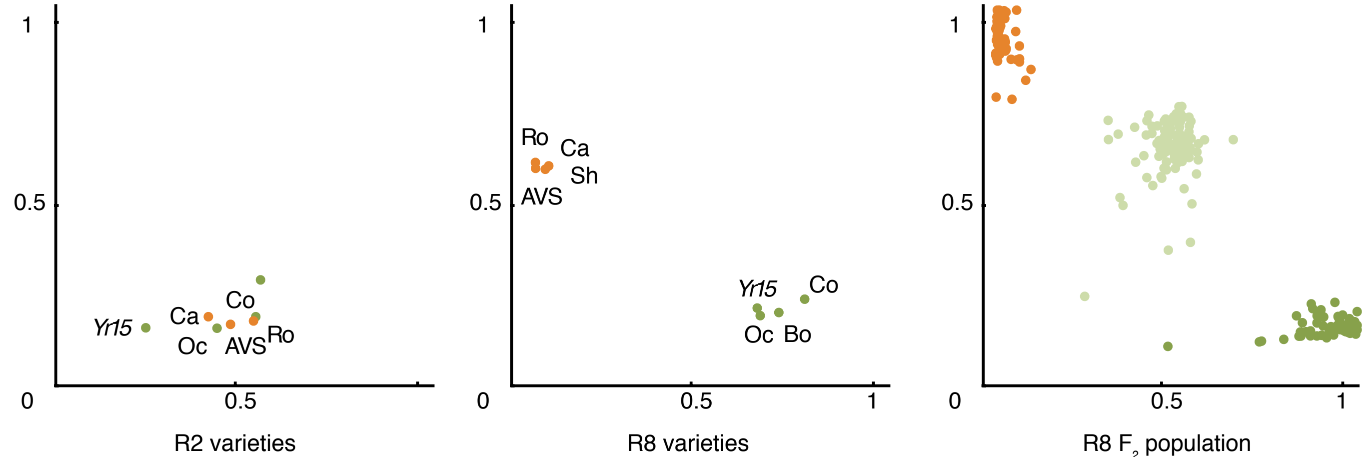
```
                   ✗                       ✓
SNP-1  A    cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B     cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A    cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B    cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgcagcgaaggcgtg
IWGSC-1D    cgcatttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
                        ✓               ✓✓
            -------:----c---------------T--&---------------------
```

Ramirez-Gonzalez, R, *et al*. (2014). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotech. J*, 12(9), 1–12.

Primer Validation

TGAC The Genome Analysis Centre™

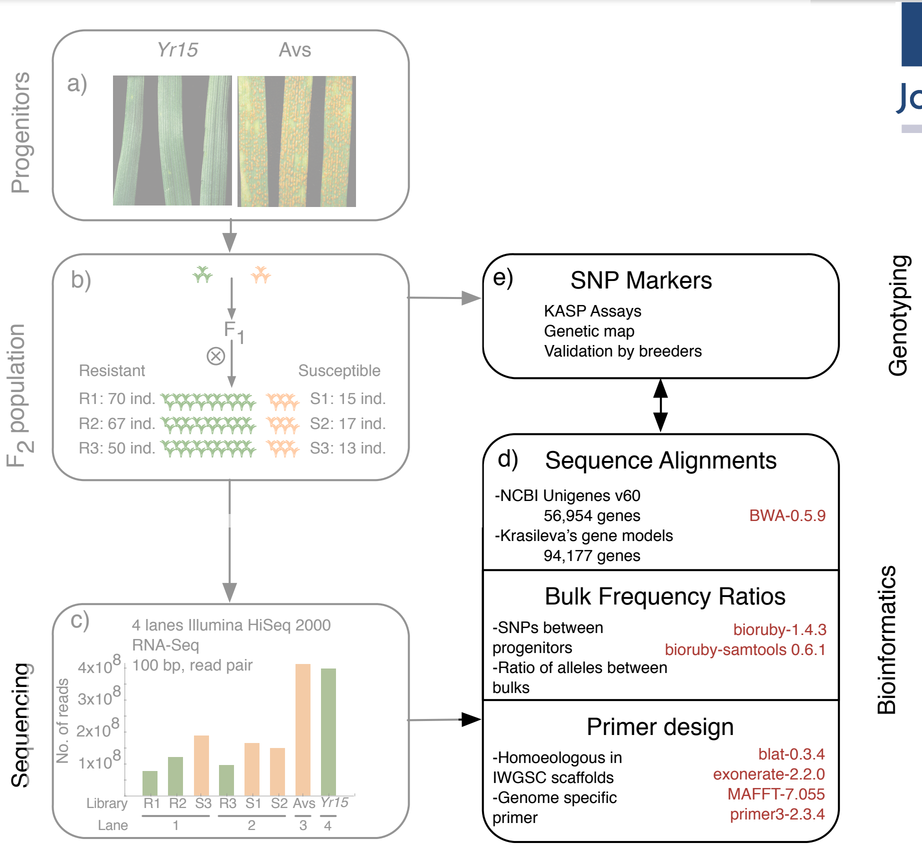
John Innes Centre



Request primers

PolyMarker is an automated bioinformatics pipeline for SNP assay development which increases the probability of generating homoeologue-specific assays for polyploid wheat. PolyMarker generates a multiple alignment between the target SNP sequence and the IWGSC chromosome survey sequences (IWGSC, 2014) for each of the three wheat genomes. It then generates a mask with informative positions which are highlighted with respect to the target genome.
See About to know how to prepare your input.
We have designed primers for the iSelect 90K chip

Upload File: Choose File no file selected
Email:
Upload



Resistant

R1: 70 ind.

R2: 67 ind.

R3: 50 ind.

Susceptible

S1: 15 ind.

S2: 17 ind.

S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

SNP Markers

KASP Assays
Genetic map
Validation by breeders

Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

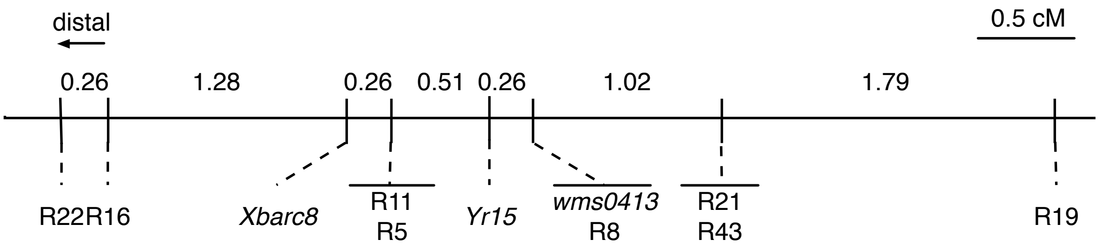
bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

Genetic Map



R22	R16	*Xbarc8*	R5	R11	Yr Score	R8	*wms0413*	R43	R21	R19	Individuals
B	B	B	B	B	R	B	B	B	B	B	51
B	B	H	H	H	R	H	H	H	H	H	1
B	B	B	B	B	R	B	B	B	B	H	6
H	H	B	B	B	R	B	B	B	B	H	1
H	H	B	B	B	R	B	B	B	B	B	1
H	H	H	B	B	R	B	B	B	B	B	1
H	H	H	H	H	R	H	H	B	B	-	1
H	H	H	H	H	R	H	H	H	H	H	88
A	A	A	A	A	S	A	A	A	A	A	37
H	H	H	H	H	S	A	A	-	A	A	1
H	H	H	H	H	R	H	H	A	A	A	2
A	H	H	H	H	R	H	H	H	H	H	1
A	A	H	H	H	R	H	H	H	H	H	2
A	A	A	A	A	R	H	H	H	H	H	1
A	A	A	A	A	S	H	H	H	H	H	1
A	A	A	A	A	S	A	A	H	H	H	1

Validation on breeding germplasm

TGAC The Genome Analysis Centre™

John Innes Centre

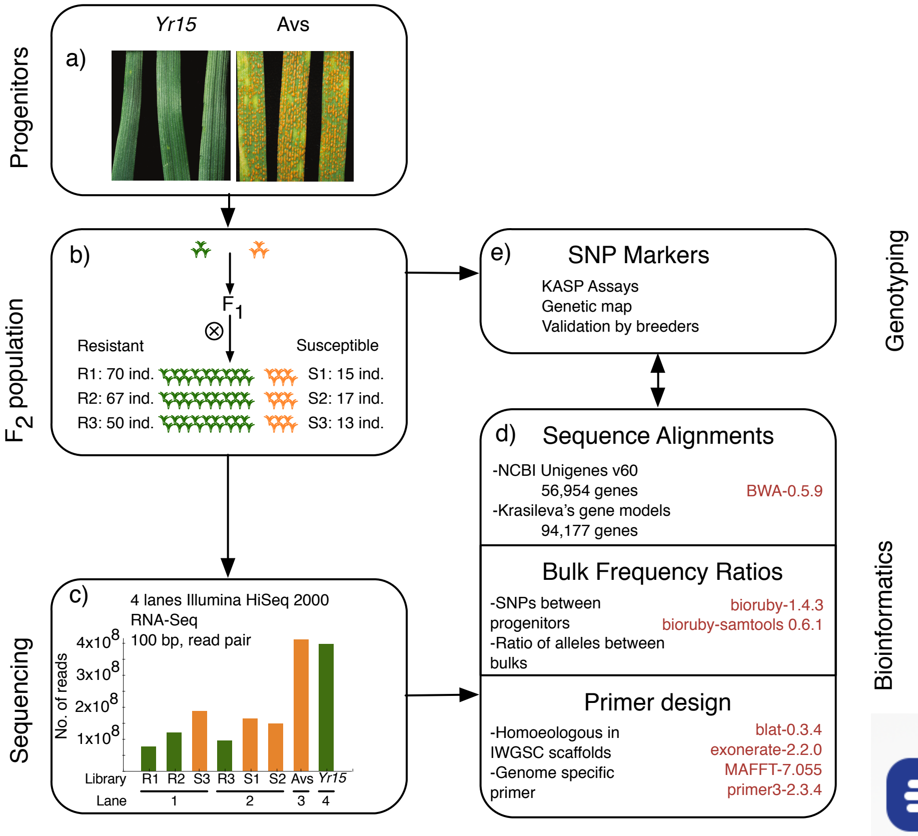
SNP haplotype			Reaction to *P. striiformis*		
R11	R5	R8	Resistant	Intermediate	Susceptible
C	A	T	-	6	16
T	A	T	-	11	-
T	G	C	79	1	-

Limagrain United Kingdom

R·A·G·T SEEDS

Validation on 113 UK varieties

John Innes Centre



Resistant

R1: 70 ind.
R2: 67 ind.
R3: 50 ind.

Susceptible

S1: 15 ind.
S2: 17 ind.
S3: 13 ind.

4 lanes Illumina HiSeq 2000
RNA-Seq
100 bp, read pair

e) SNP Markers

KASP Assays
Genetic map
Validation by breeders

d) Sequence Alignments

-NCBI Unigenes v60
56,954 genes
-Krasileva's gene models
94,177 genes

BWA-0.5.9

Bulk Frequency Ratios

-SNPs between progenitors
-Ratio of alleles between bulks

bioruby-1.4.3
bioruby-samtools 0.6.1

Primer design

-Homoeologous in IWGSC scaffolds
-Genome specific primer

blat-0.3.4
exonerate-2.2.0
MAFFT-7.055
primer3-2.3.4

PolyMarker

Acknowledgments

- TGAC
  - Mario Caccamo
  - Sarah Ayling
  - Paul Bailey
  - Jon Wright
- JIC
  - Cristobal Uauy
  - Nick Bird
  - Vanesa Segovia
  - Martin Trick
- Limagrain
  - Paul Fenwick
  - Simon Berry
- RAGT Seeds
  - Sarah Holdgate
  - Peter Jack
- University of Sidney
  - Robert McIntosh

BBSRC

International Wheat Genome Sequencing Consortium

Thank you for listening.

Ramirez-Gonzalez, R, *et al*. (2014). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotech. J*, 12(9), 1–12.