

IWGSC protocols

Second generation sequencing for
marker discovery

1. Coordination of efforts

Gather information on who is sequencing which species/varieties
Possibly collection of all gene sequences in a combined database

2. Methods for library production

Determine if efforts should be made to remove mt and chloroplast DNA

3. Standard QC for data evaluation and comparison between datasets

It would be important to have a discussion what people use as criteria.
Sequence quality for assemblies,
Cutoff for BLAST searches,
etc.

4. Approaches to SNP discovery

Consider whether (and which) different varieties or species should be used
Identification of repeats and genes
Identification and comparison of true orthologs from different varieties/species

Who is doing what?

Country	Target	Coverage	Technique
UK	WGS	5x	454 Titanium
UK	3DL	75x	GAI PE
Triticeae Genome	Group 1	1.5x	454 Tit
France	3B	2x	GAI
Australia	WGS	0.2x	GAI PE
Australia	7DS	16x	GAI PE
US	3A	2x	454 Tit
Italy	5A	2x	454 Tit
Switzerland	A/B/D ancestors and R(ye)	0.1 x each	454 Tit
China	D (Ae. tauschii)	40x	GAI PE
China	A (T. Urartu)	40x	GAI PE

1. Coordination of efforts

Gather information on who is sequencing which species/varieties
Possibly collection of all gene sequences in a combined database

2. **Methods for library production**

Determine if efforts should be made to remove mt and chloroplast DNA

3. Standard QC for data evaluation and comparison between datasets

It would be important to have a discussion what people use as criteria.
Sequence quality for assemblies,
Cutoff for BLAST searches,
etc.

4. Approaches to SNP discovery

Consider whether (and which) different varieties or species should be used
Identification of repeats and genes
Identification and comparison of true orthologs from different varieties/species

1. Coordination of efforts

Gather information on who is sequencing which species/varieties
Possibly collection of all gene sequences in a combined database

2. Methods for library production

Determine if efforts should be made to remove mt and chloroplast DNA

3. **Standard QC for data evaluation and comparison between datasets**

It would be important to have a discussion what people use as criteria.
Sequence quality for assemblies,
Cutoff for BLAST searches,
etc.

4. Approaches to SNP discovery

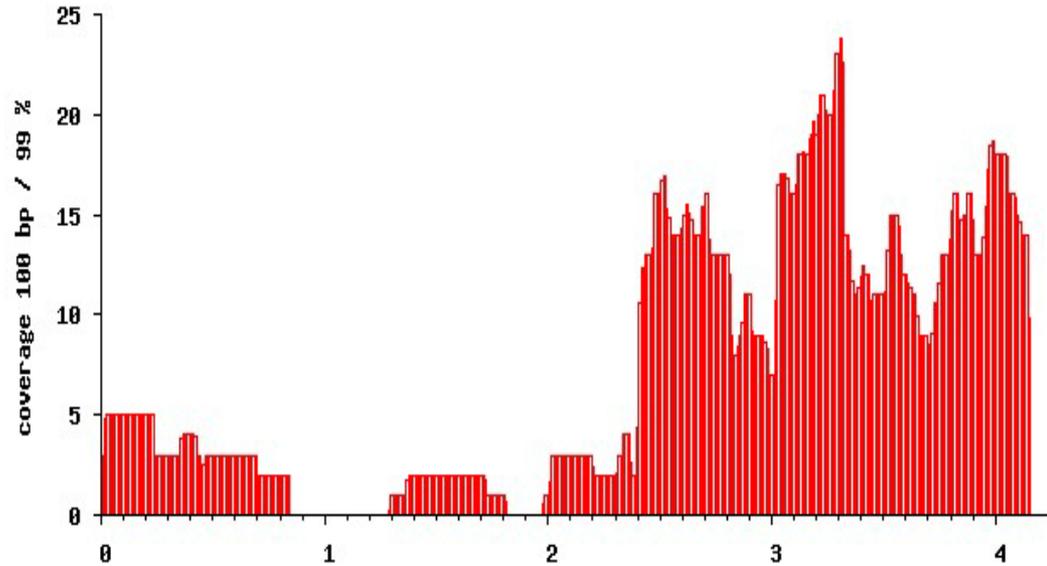
Consider whether (and which) different varieties or species should be used
Identification of repeats and genes
Identification and comparison of true orthologs from different varieties/species

Coverage of 1AS genes

Pm3

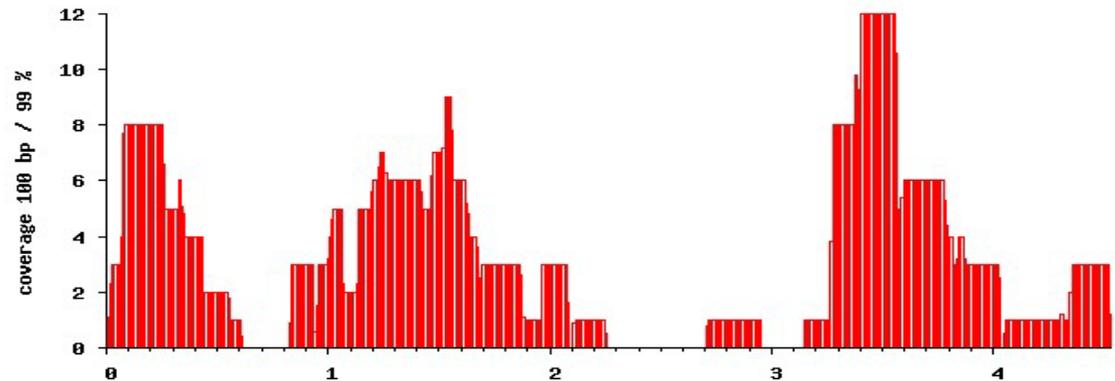
High coverage in LRR region

➔ Maybe partially duplicated elsewhere



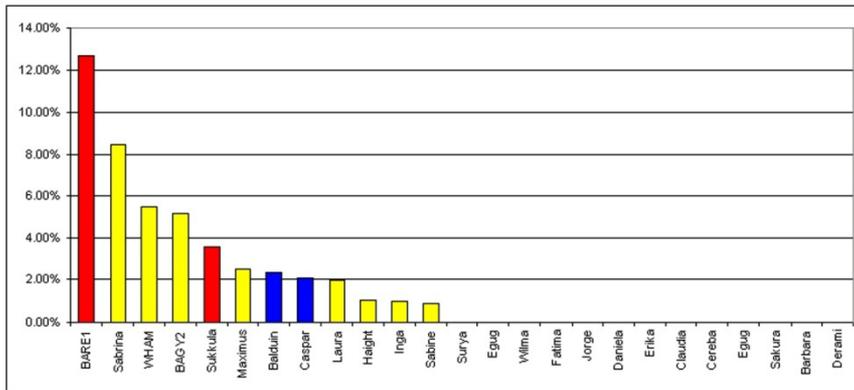
RGL9

Many reads at exactly the same positions
Uneven amplification?

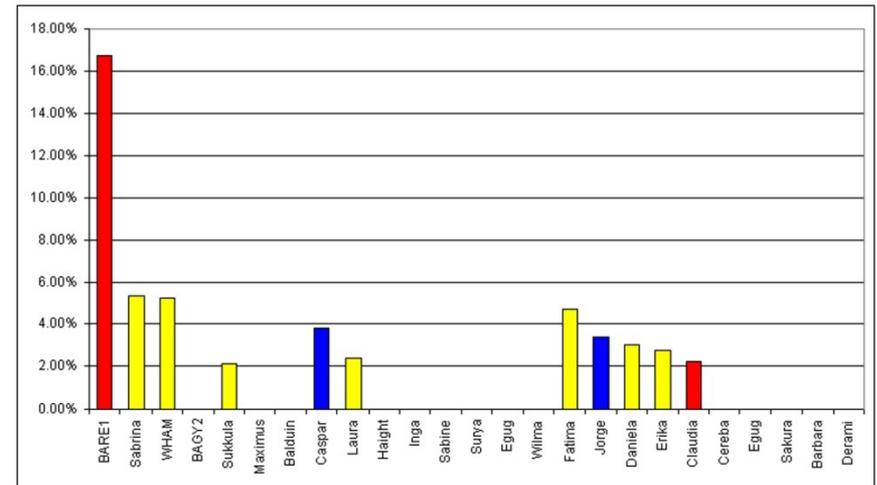


Amplified vs. non-amplified 454

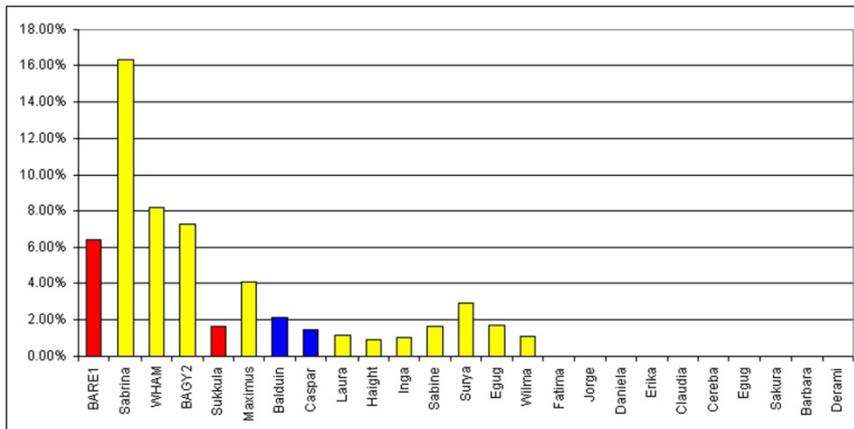
Barley whole genome



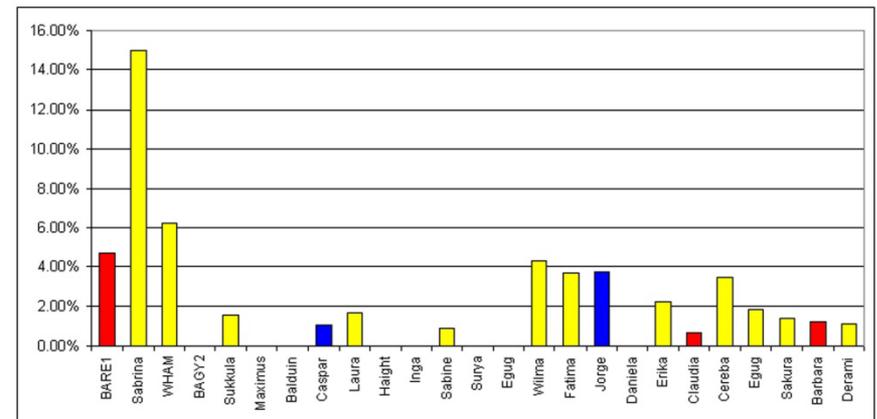
Wheat whole genome



Barley 1H amplified

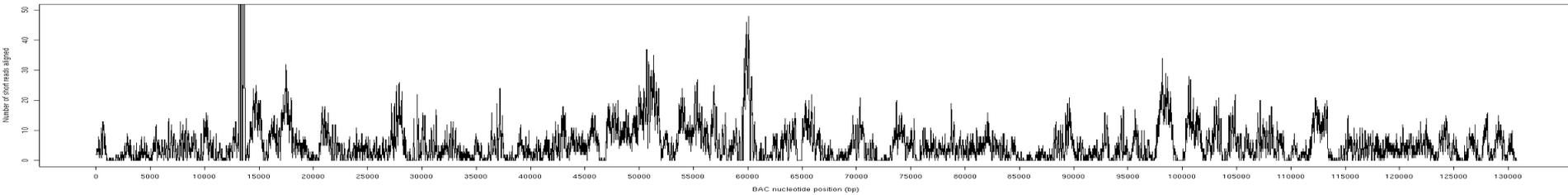


Wheat 1AS amplified

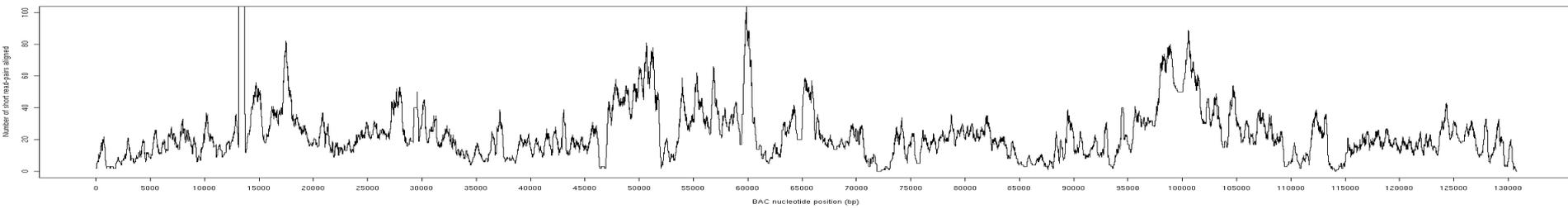


Data visualisation

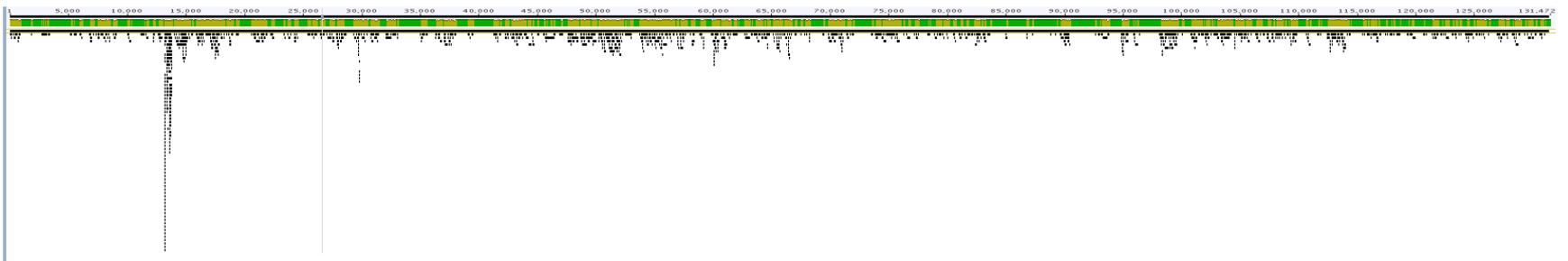
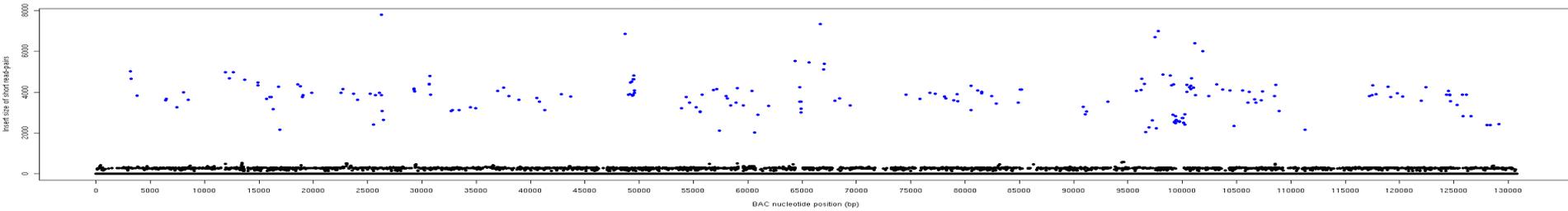
Coverage Plot of Brassica rapa BAC AC189312 vs Brassica napus paired-end short-read libraries



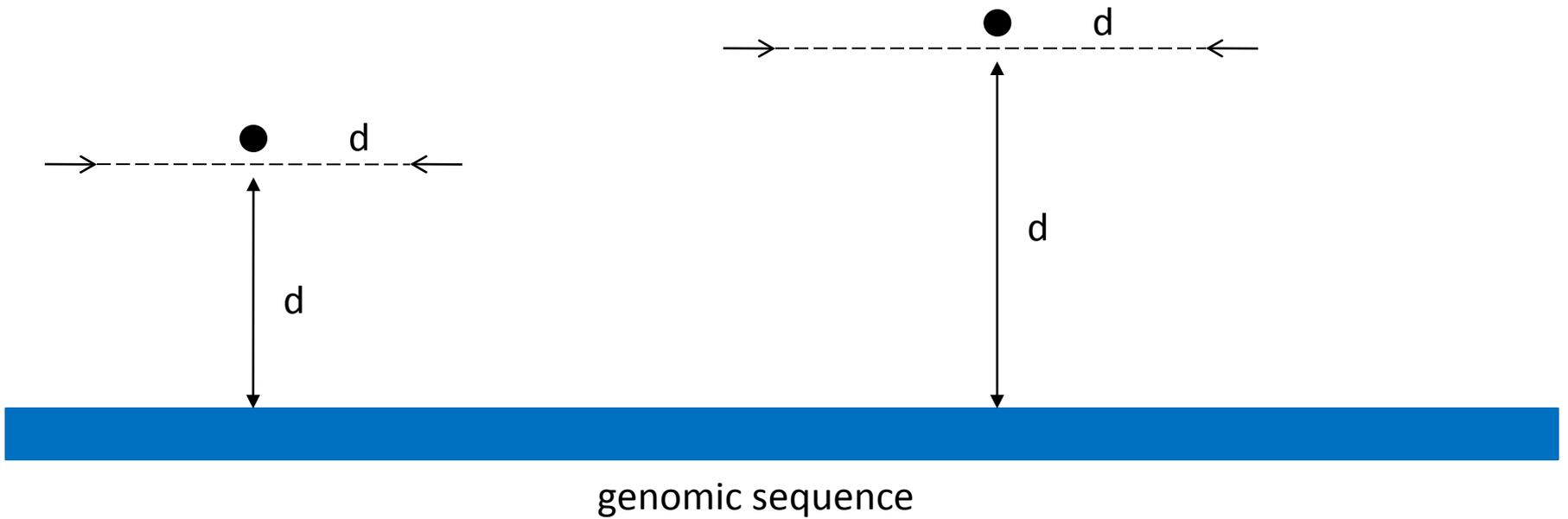
Paired-coverage Plot of Brassica rapa BAC AC189312 vs Brassica napus paired-end short-read libraries



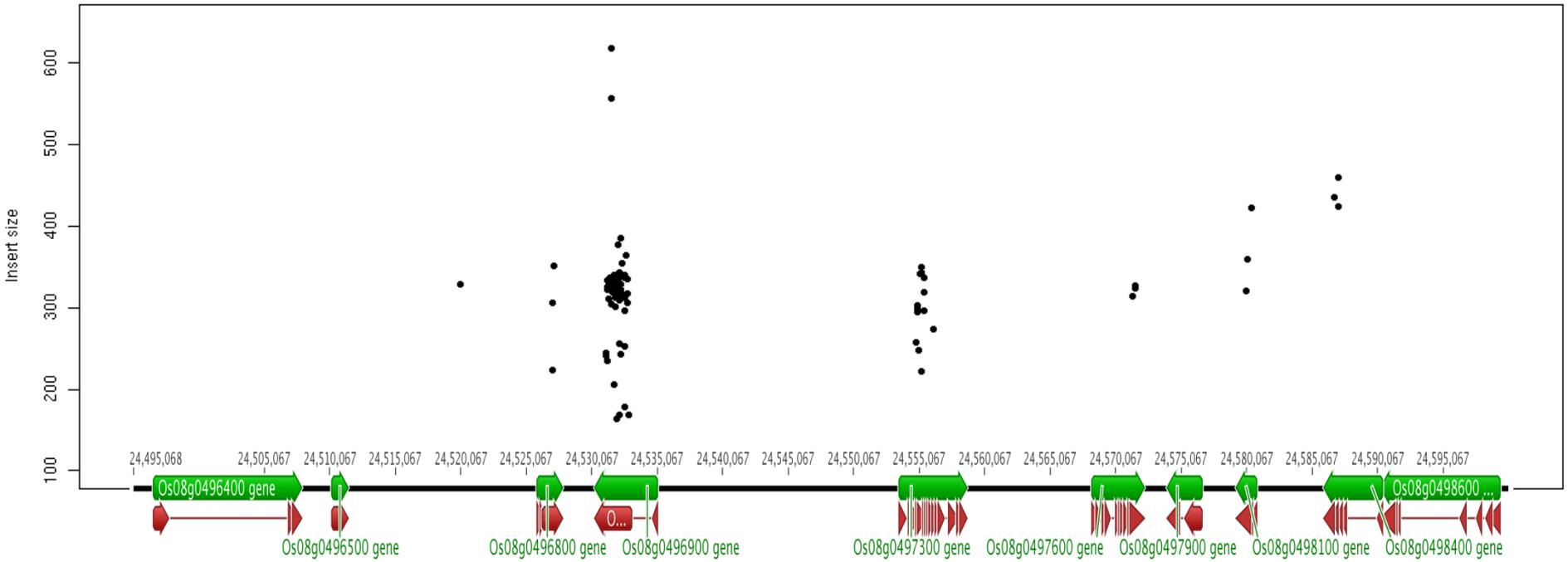
Insert size of Brassica napus paired-end short-read libraries vs Read Pair Position on Brassica rapa BAC AC189312



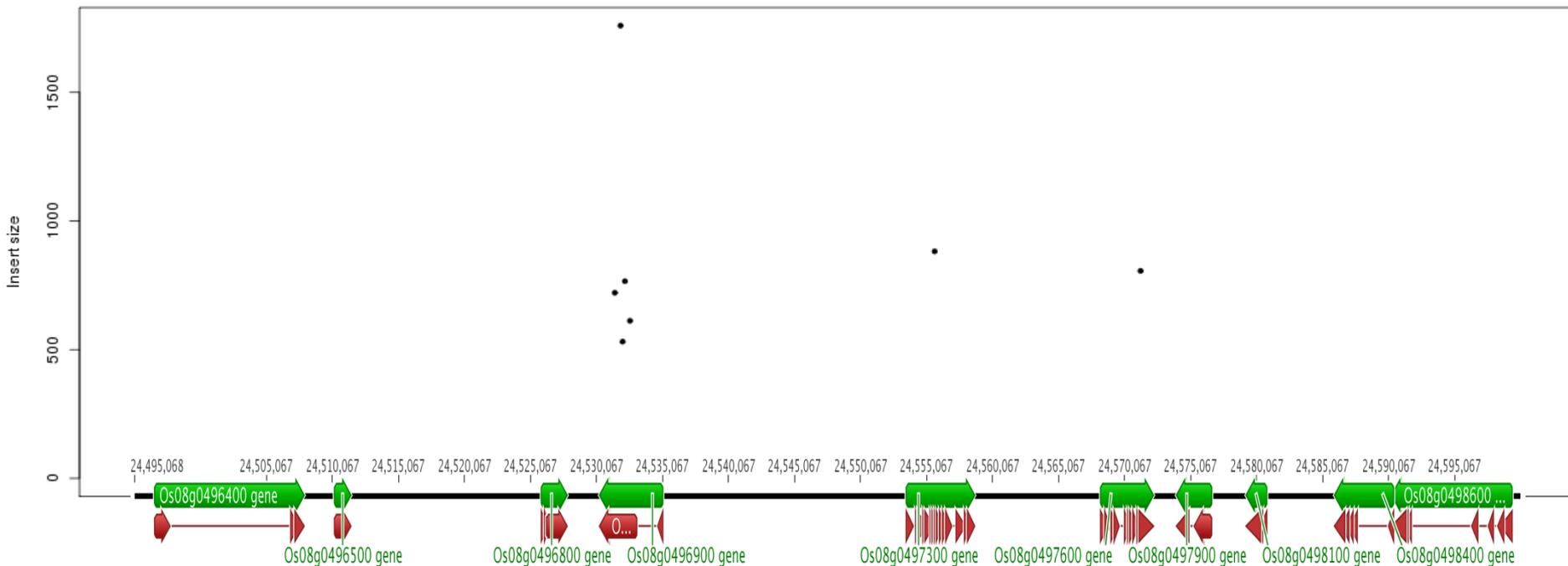
Comparison with rice



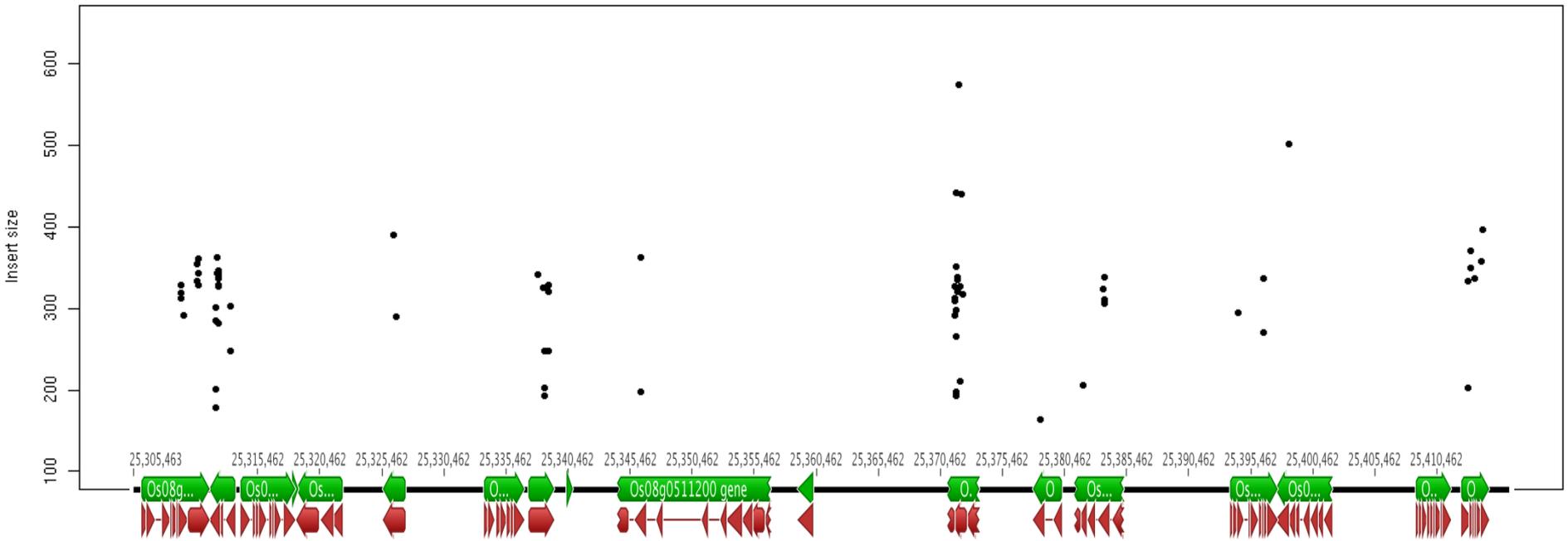
Insert size vs Nucleotide position of 7DS short-read libraries versus OsCh8 24495kb-24600kb genomic fragment



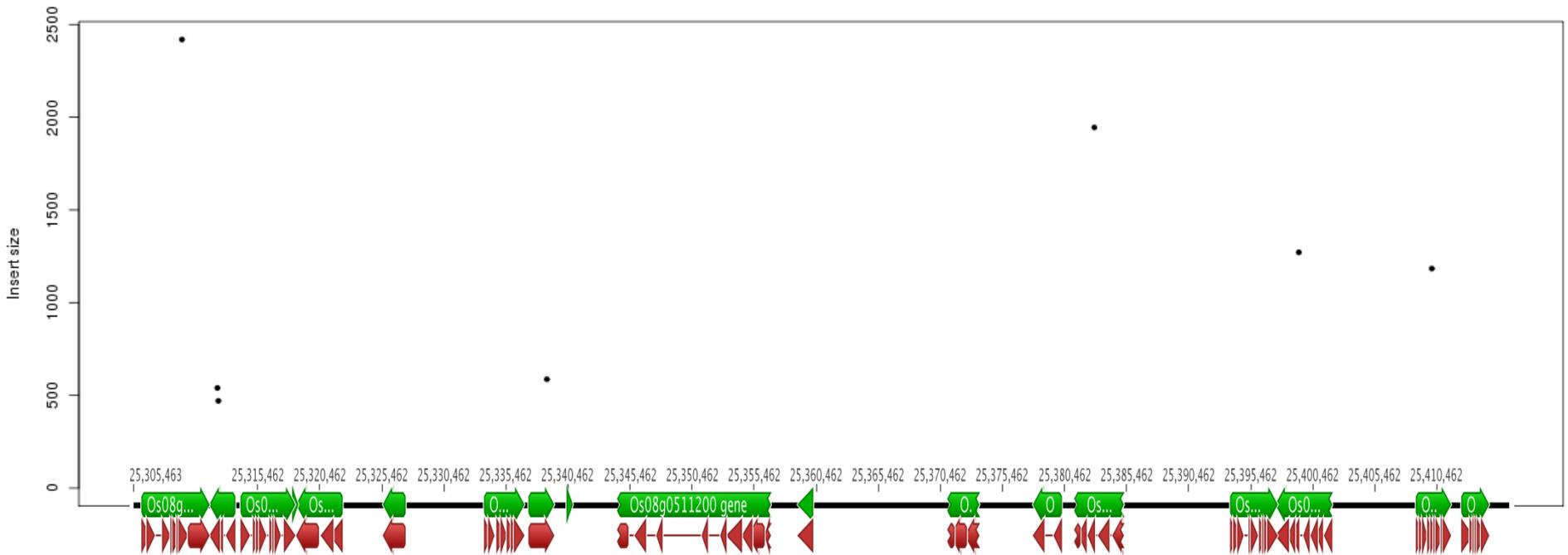
Insert size of 7DS_37_003 read-pairs vs nucleotide position of read-pair hits on OsCh8_24495kb-24600kb



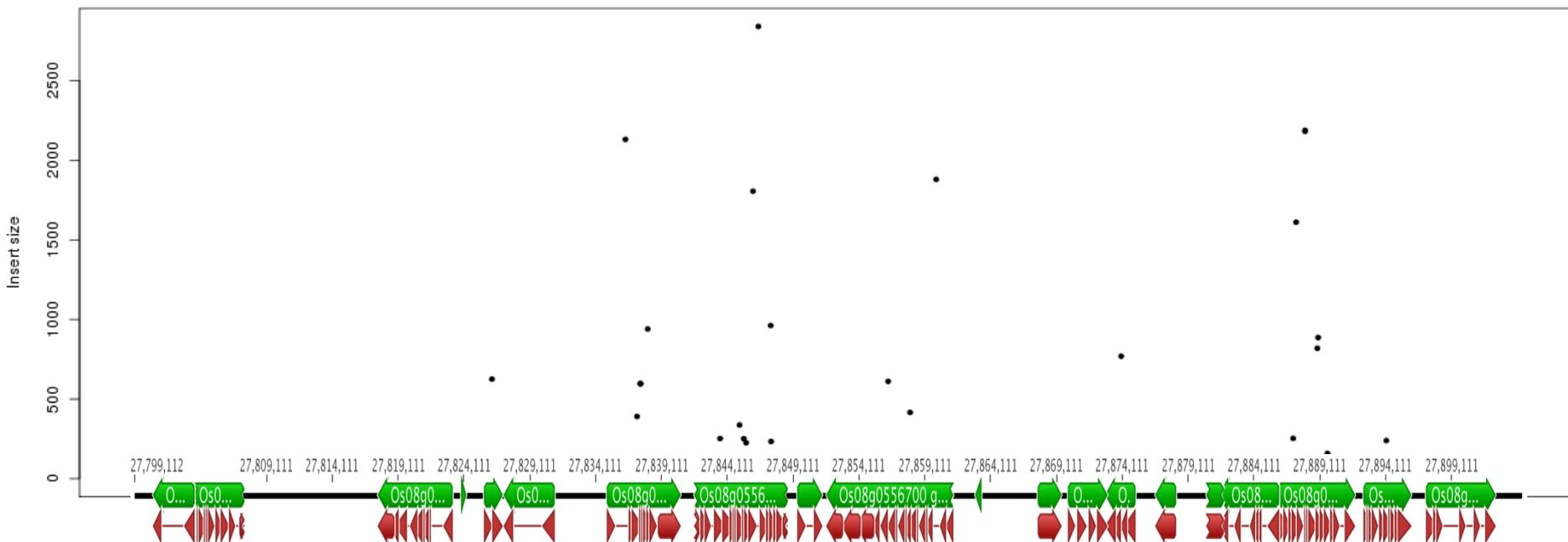
Insert size vs Nucleotide position of 7DS short-read libraries versus OsCh8 25305kb-25415kb genomic fragment



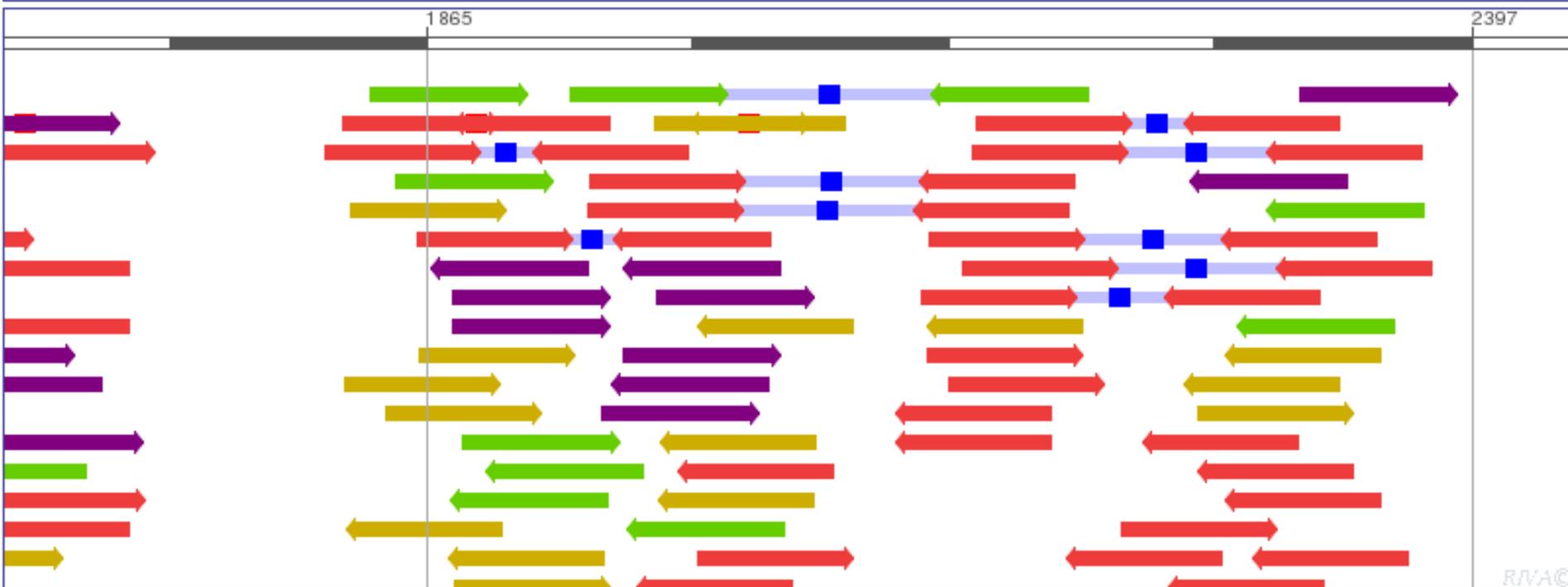
Insert size of 7DS_37_003 read-pairs vs nucleotide position of read-pair hits on OsCh8_25305kb-25415kb



Insert size of 7DS_37_003 read-pairs vs nucleotide position of read-pair hits on OsCh8_27800kb-27905kb



Results: >gi|115469169|ref|NM_001064719.1| *Oryza sativa* (japonica cul...



TAGdb analysis complete for job

ID: yehd4b1q8Z5iXVTgz2niwLzh9

Created: 06-01-2010

User: p.berkman@uq.edu.au

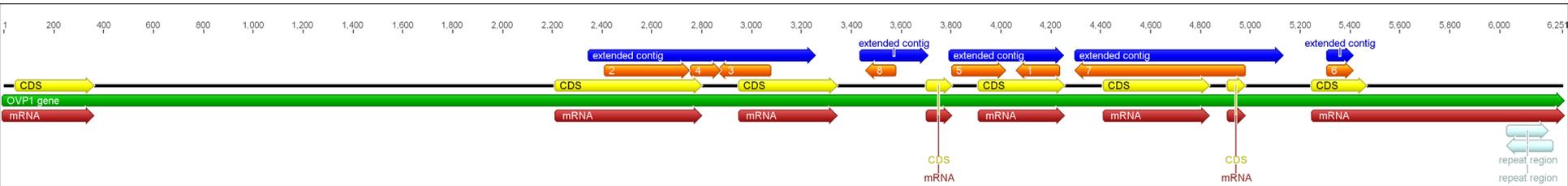
Legend:


Library Name: 7DS_03_002
Read Length: 76 – Insert Size: 320

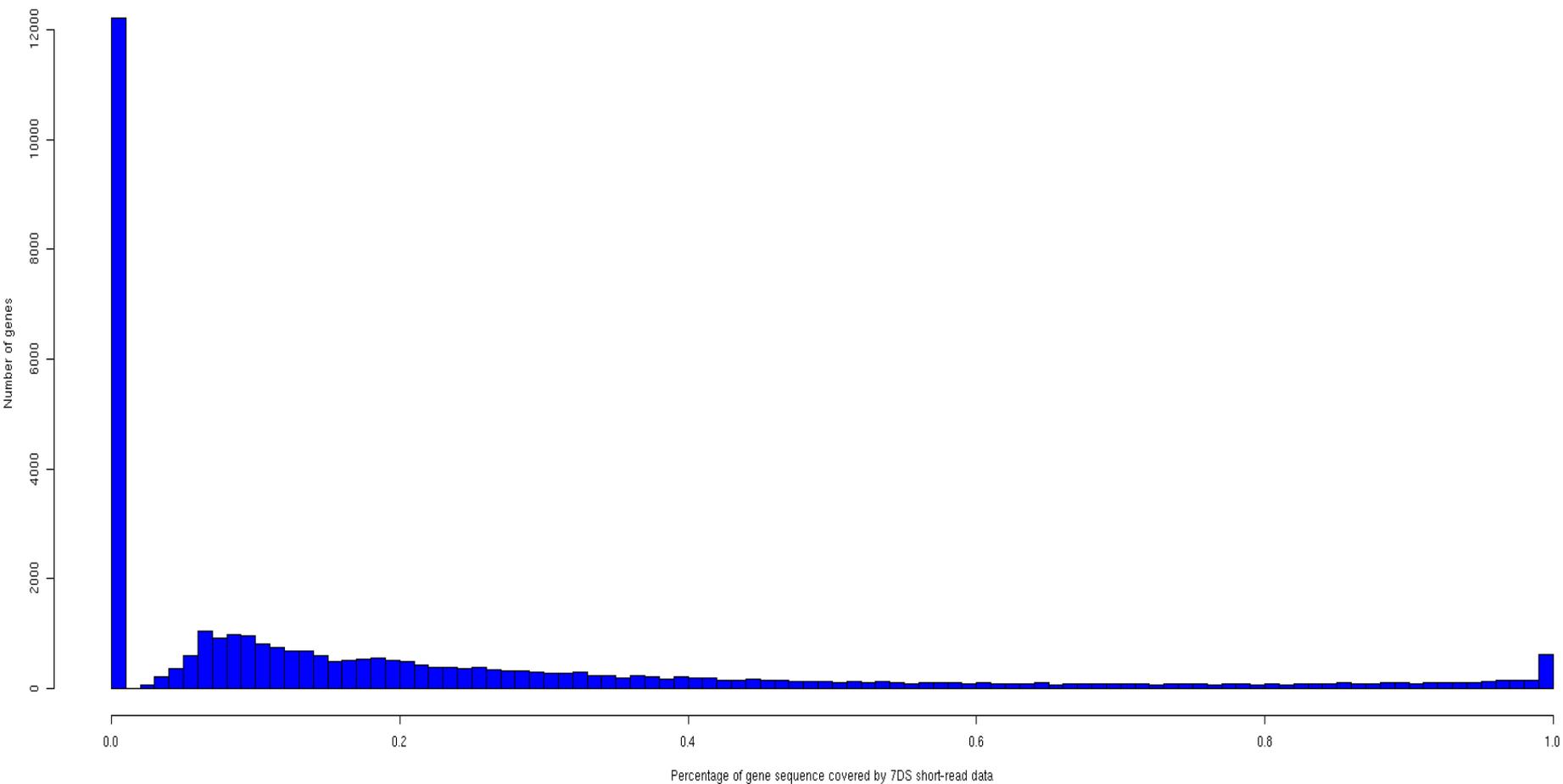

Library Name: 7DS_03_001
Read Length: 76 – Insert Size: 320


Library Name: 7DS_03_003
Read Length: 75 – Insert Size: 320


Library Name: 7DS_03_004
Read Length: 75 – Insert Size: 320



Histogram of percentage coverage of wheat gene sequences by chromosome arm 7DS short-read data



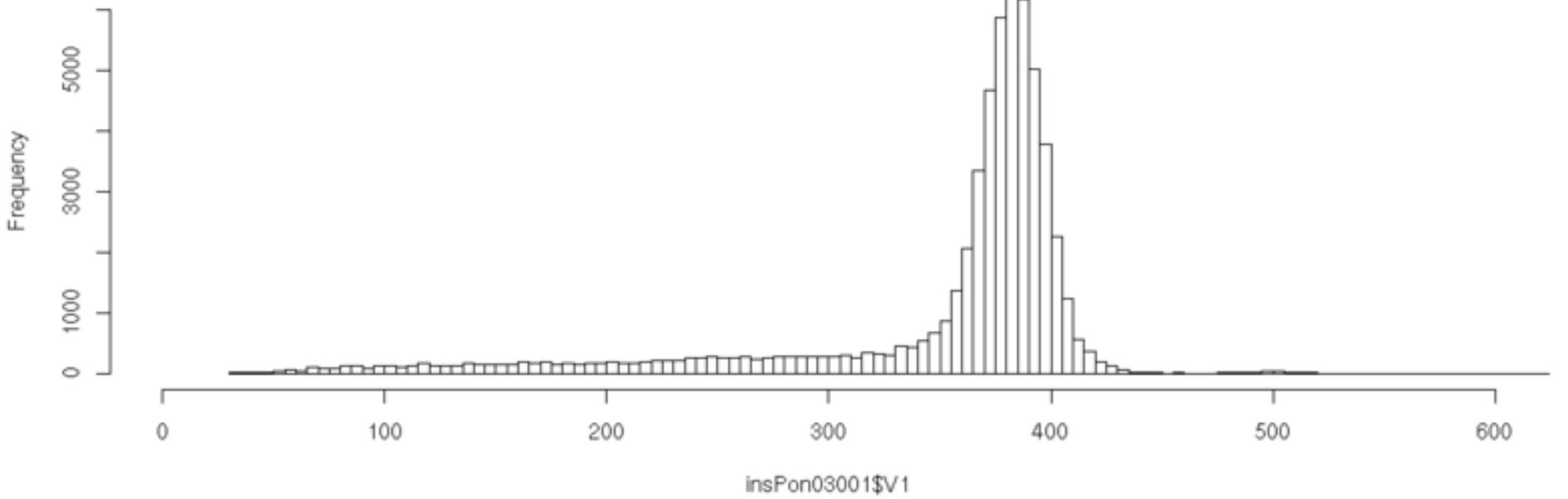
MSk	Skippy, Kangaroo rumen metagenomic sample
MDa	Daisy, Cow rumen metagenomic sample

Illumina Data

[edit]

Approx. Date of Run	Lane Number/s	Species	Biosource/Cultivar	Library Name	Read Length	Predicted Insert(*)	Actual
Prior to Dec 2008	PBII	<i>Brassica rapa</i>	Chiifu	BrC_37_001	35 bp	3-4 kbp	2800 bp e
Prior to Dec 2008	PBII	<i>Brassica rapa</i>	Chiifu	BrC_37_002	35 bp	3-4 kbp	2800 bp e
Prior to Dec 2008	PBII	<i>Brassica rapa</i>	Chiifu	BrC_27_001	35 bp	3-4 kbp	2700 bp e
17 Dec 2008****	1-5	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
.-.	1	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
.-.	2	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
.-.	3	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
.-.	4	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
.-.	5	<i>Brassica rapa</i>	Chiifu	BrC_03_001	36 bp	300 bp	140 bp e
17 Dec 2008****	6-7	Barley	Morex	HvM_03_001	36 bp	300 bp	210 bp e
.-.	6	Barley	Morex	HvM_03_001	36 bp	300 bp	210 bp e
.-.	7	Barley	Morex	HvM_03_001	36 bp	300 bp	210 bp e
20 Feb 2009	1	<i>Pongamia pinnata</i>	Fp4PG	Pon_03_001	36 bp	300 bp	390 bp e
20 Feb 2009	2	Wheat	Chinese Spring	WCs_03_001	36 bp	300 bp	300 bp e
20 Feb 2009	3	Barley	Morex	HvM_03_002	36 bp	300 bp	210 bp e
20 Feb 2009	4	<i>Brassica rapa</i>	Chiifu	BrC_03_002	36 bp	300 bp	300 bp e

Histogram of insPon03001\$V1



1. Coordination of efforts

Gather information on who is sequencing which species/varieties
Possibly collection of all gene sequences in a combined database

2. Methods for library production

Determine if efforts should be made to remove mt and chloroplast DNA

3. Standard QC for data evaluation and comparison between datasets

It would be important to have a discussion what people use as criteria.
Sequence quality for assemblies,
Cutoff for BLAST searches,
etc.

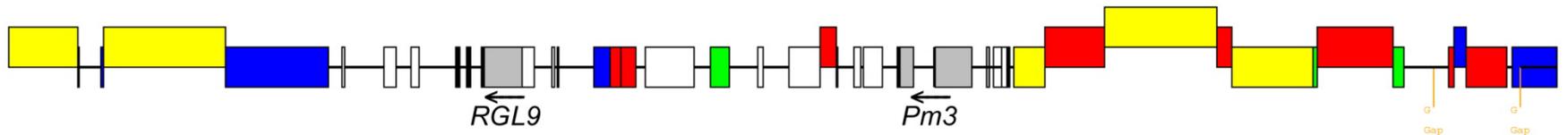
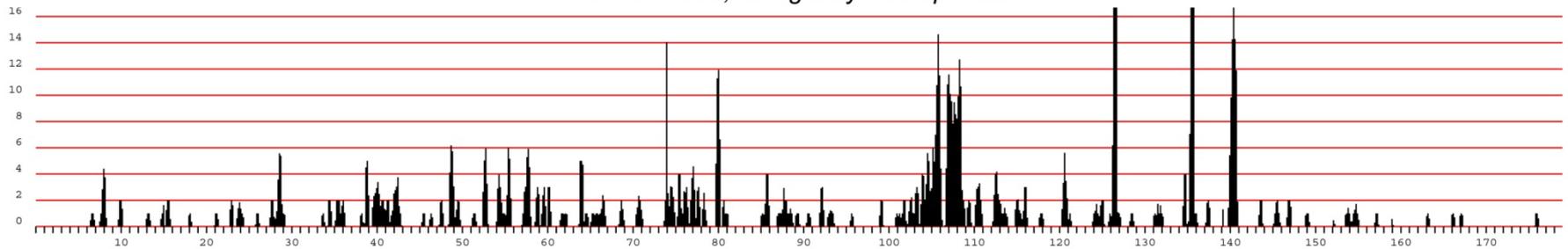
4. Approaches to SNP discovery

Consider whether (and which) different varieties or species should be used
Identification of repeats and genes
Identification and comparison of true orthologs from different varieties/species

Coverage of 1AS genes

Pm3 locus from cv. *Chinese spring*

Linear scale, stringency 100 bp / 98%



Logarithmic scale, stringency 80 bp / 80 %

