

Development of tools for the analysis and visualisation of second generation sequencing data for *Brassica* species

Chris Duran
University of Queensland, Australia
C.Duran@uq.edu.au

Outline

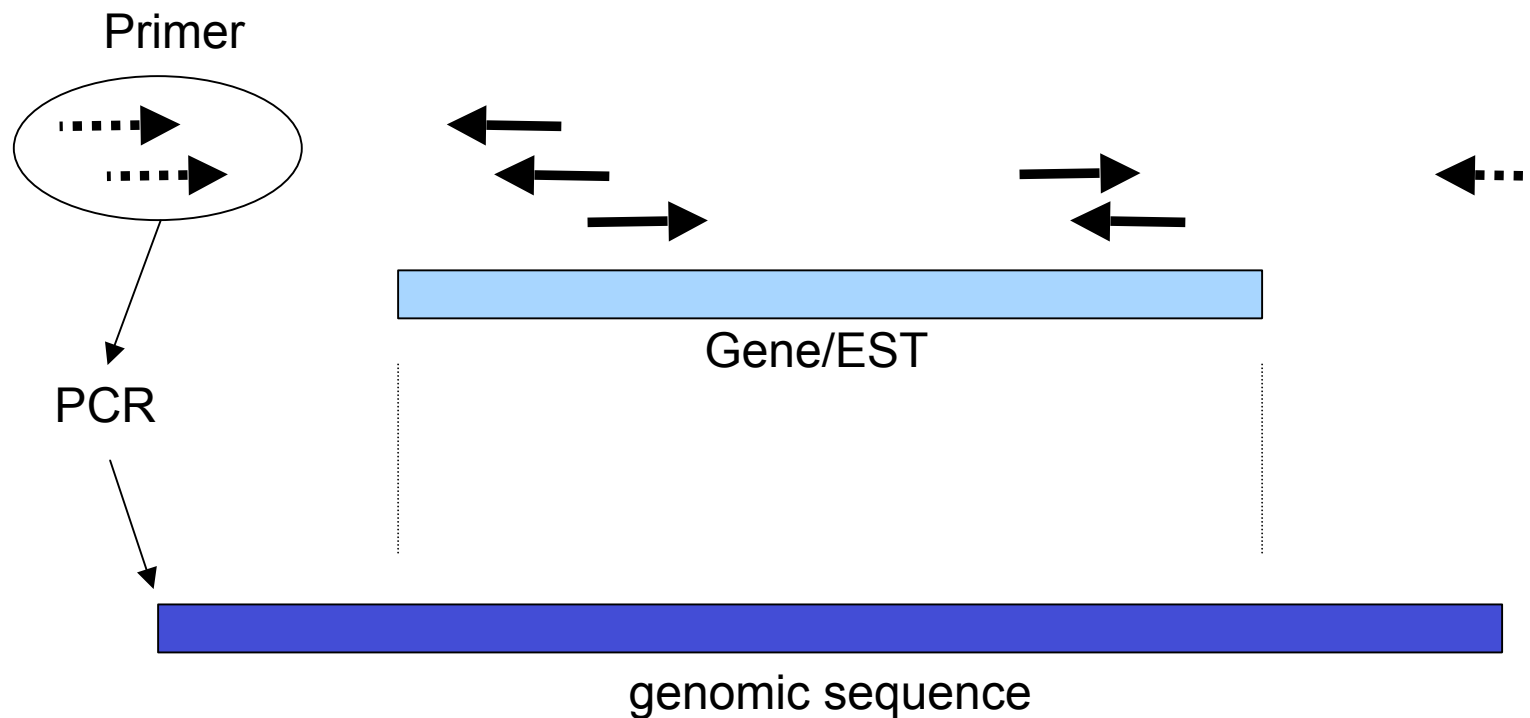
- *Brassica* gene and promoter discovery: TAGdb
- *Brassica* genome sequencing and annotation
- Linking genetic and genomic data using CMap3D

Paired-end short reads



- Illumina GAIIx
- Read length (35bp – 75bp)
- Insert size up to 10Kbp
 - ~ Normal distribution
 - Standard deviation ~ 10% mean

Gene finding and extension



Known (Arabidopsis)

Unknown (Brassica) →

Welcome to ACPFG Bioinformatic's TAGdb.

This service performs BLAST alignment between a single query and short pair reads of selected species.

Please enter a valid email address

Note: Your result will be sent to the specified address.

Sequence data

Either: Select the sequence file to upload:

Otherwise: Enter a sequence in FASTA format:

Note: Query sequence must be less than 5000 nucleotides.

Species selection

Please choose a query species

- Barley
- Brassica
- Pongamia
- Wheat

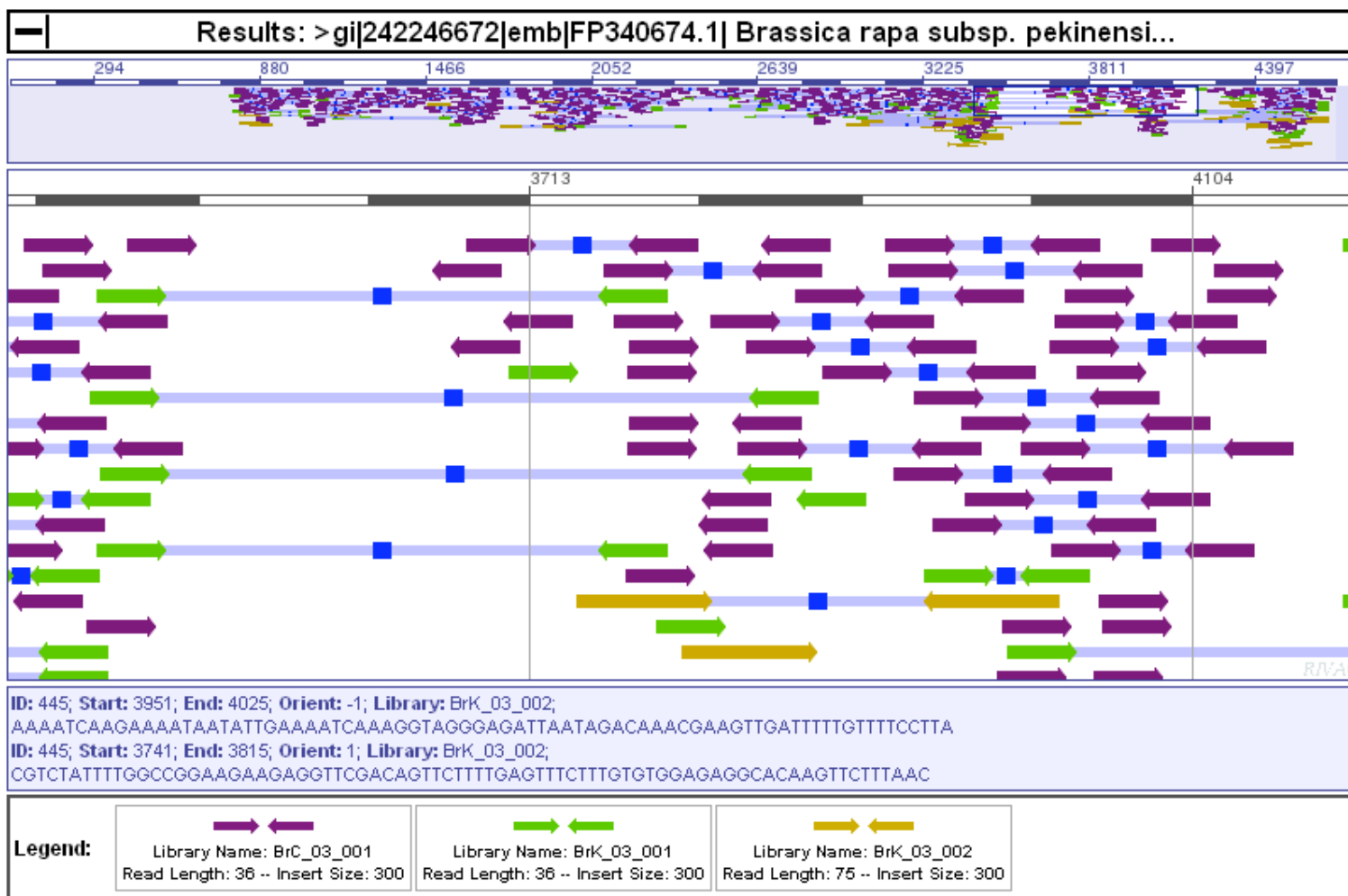


Short paired-read library selection

Please select one or more paired-read libraries to search

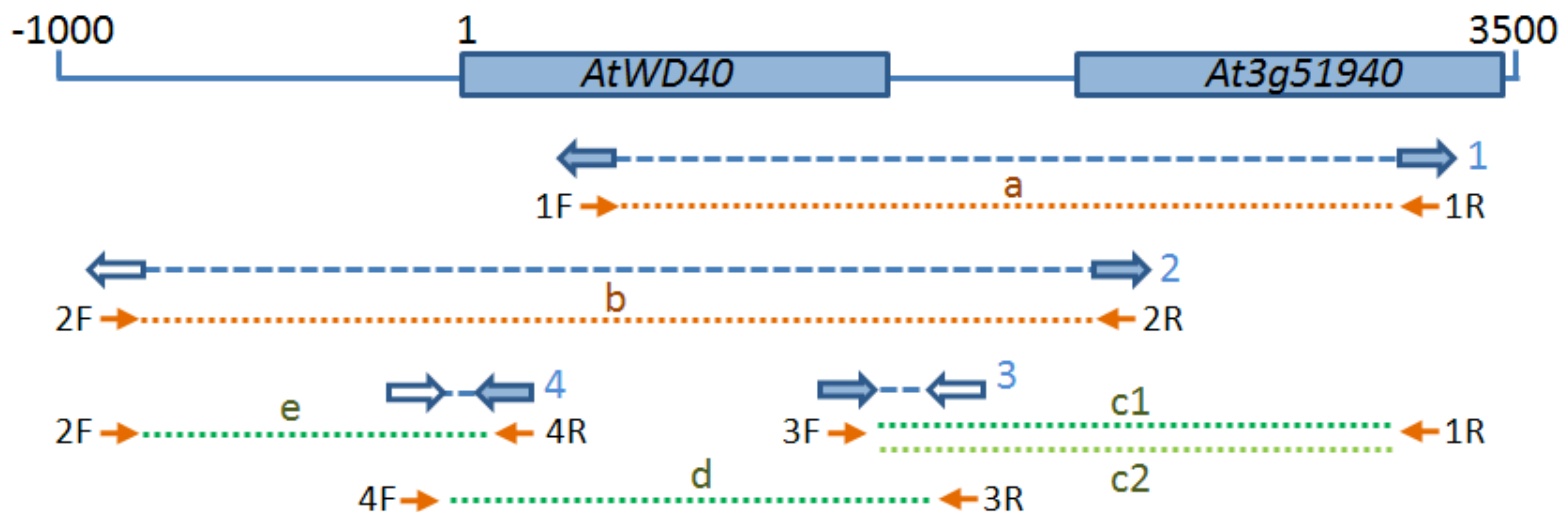
Format: SourceName - ReadLength - InsertSize - LibraryName

TAGdb



<http://flora.acpfg.com.au/tagdb/cgi-bin/results?jobID=bK85Lk10fVzMlw5e33FSuYBYr>

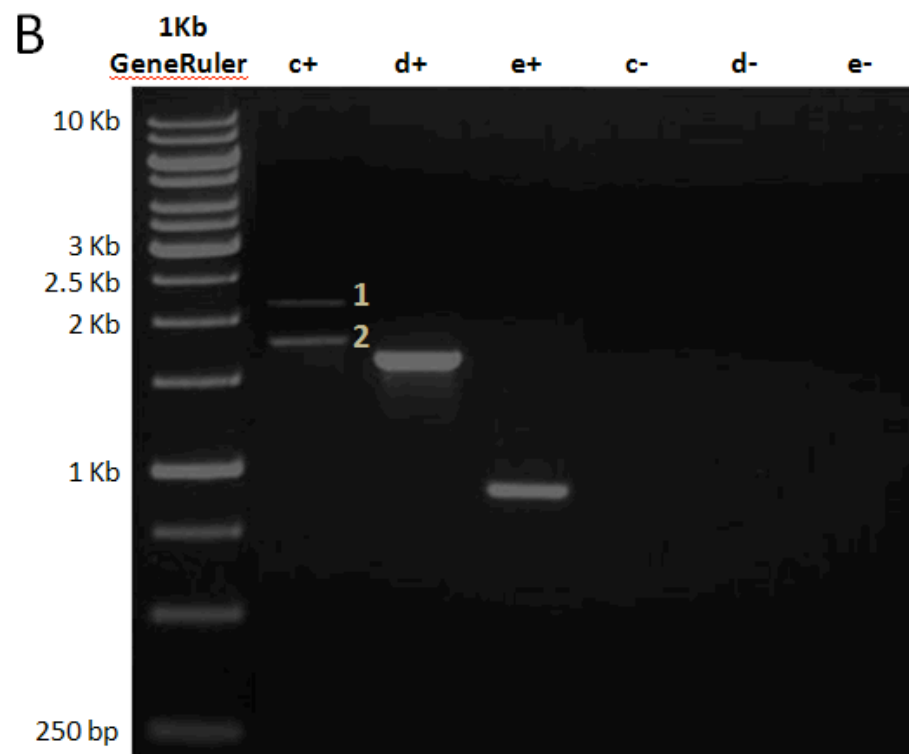
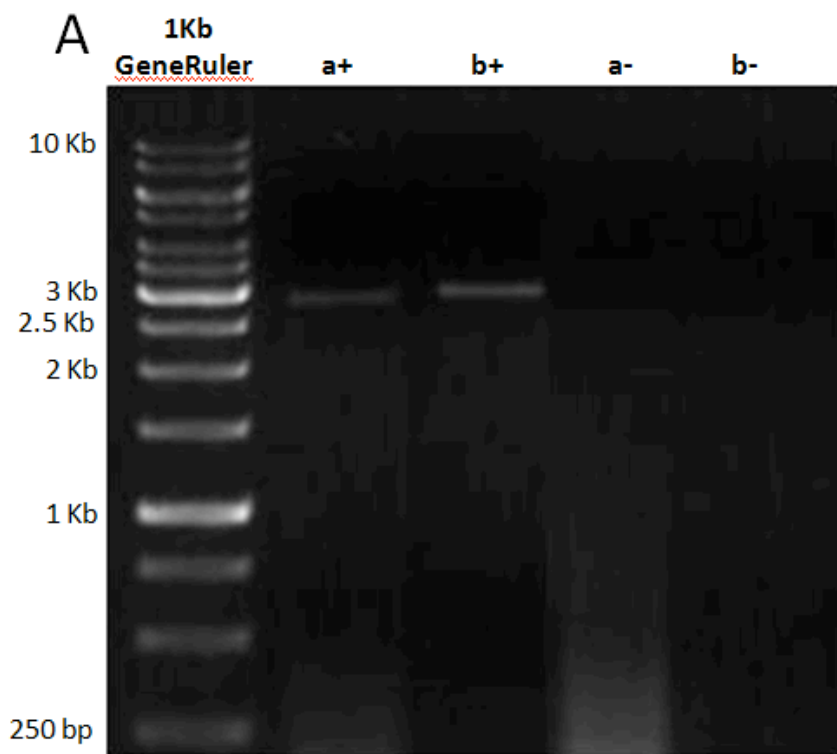
Example: AtWD40



→ ← : Primers
 → ← : Large intra-tag-pair PCR products
 → ← : Small inter-tag-pair PCR products

↔ : Brassica large-insert tag pairs
 ↔ : Brassica small-insert tag pairs
 ← : Tags lacking significant similarity to the reference in TAGdb

Example: AtWD40



1% TAE-agarose gels showing amplification products (+) and negative controls (-) for products (a) and (b) (Panel A) and (c) to (e) (Panel B) in *B. rapa*.

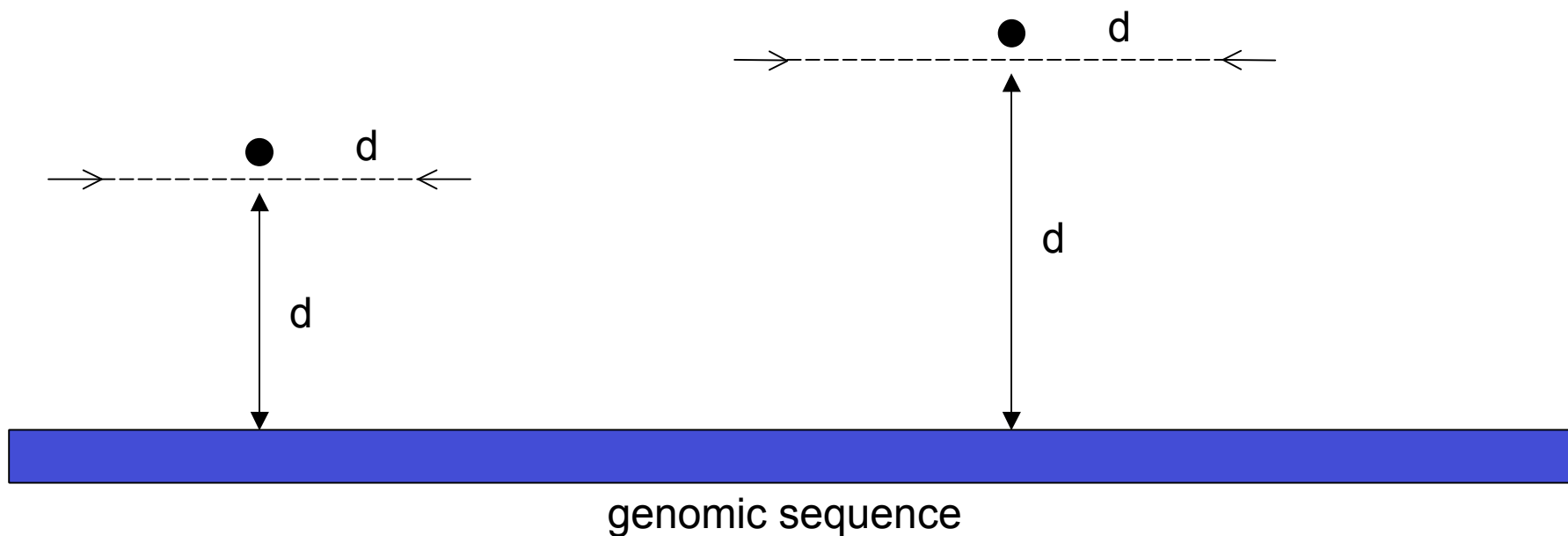
Data

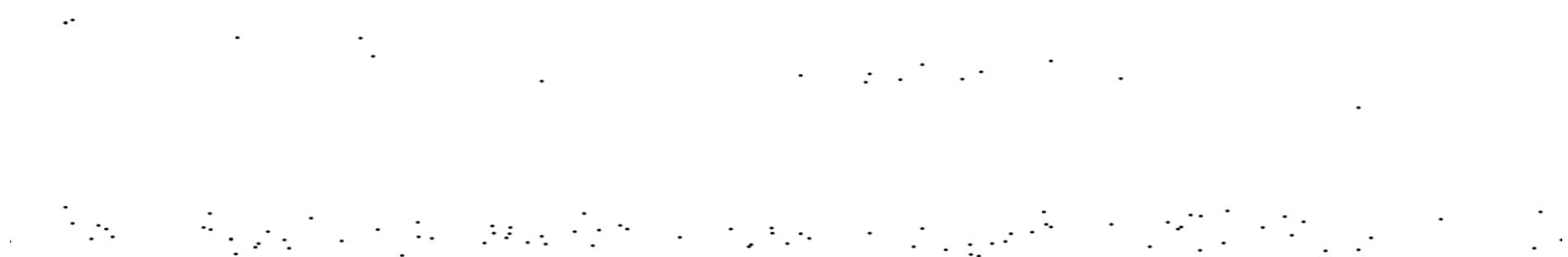
<i>Brassica rapa</i>	5 Gbp
<i>Brassica oleracea</i>	1 Gbp
<i>Brassica nigra</i>	1 Gbp
Wheat	2.3 Gbp
Wheat 7DS	4.2 Gbp
Barley	2.9 Gbp
<i>Pongamia</i>	0.45 Gbp
<i>Nicotiana</i>	10.2 Gbp

TagDB

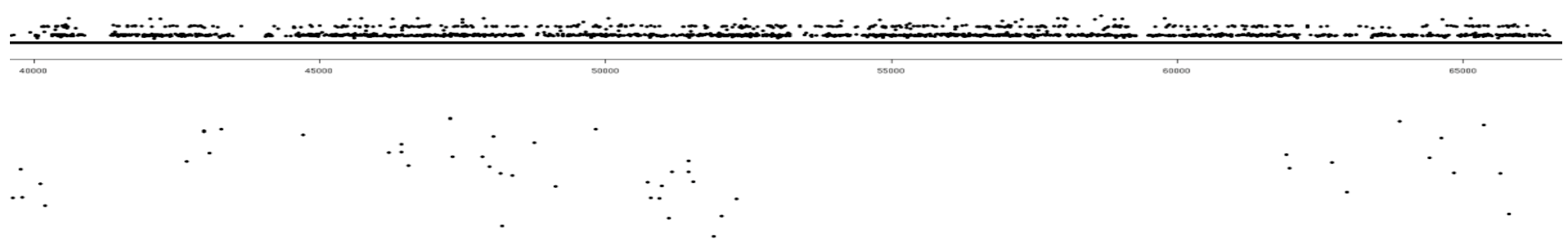
- Web-based tool for short read comparison
 - Short reads stored on server
 - User uploads query sequence
- <http://flora.acpfg.com.au/tagdb>

Visualising read pairs for comparative genomics





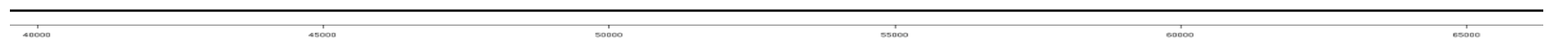
B. rapa Chiifu



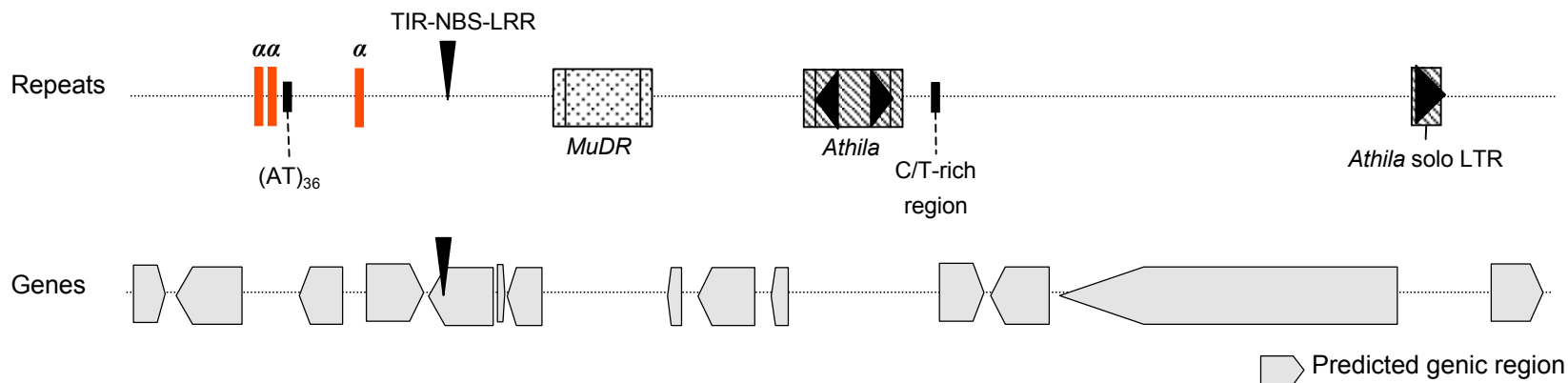
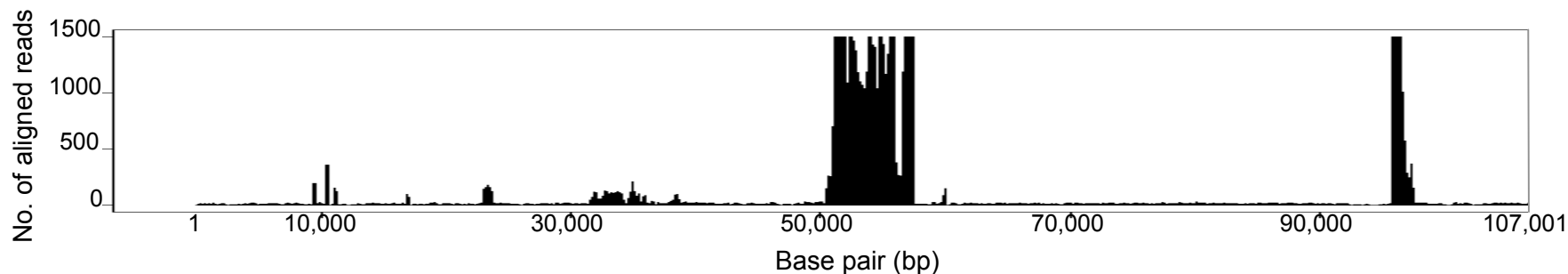
B. oleracea



B. nigra



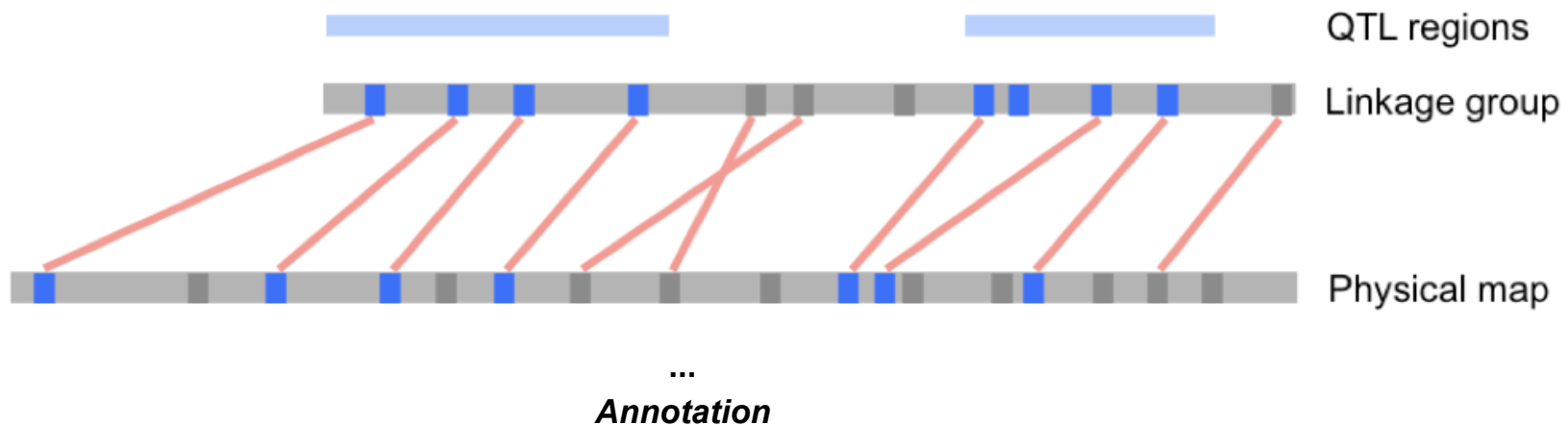
Genome annotation



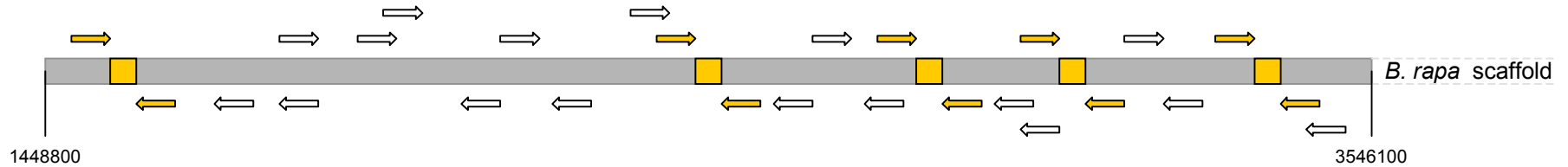
High-covered regions of short reads and their corresponding annotation in a *B. rapa* BAC.

CMap3D

- Finding the genes for the traits
- Integration of genetic data with genomic data
 - Mapping of QTL regions to genomic data



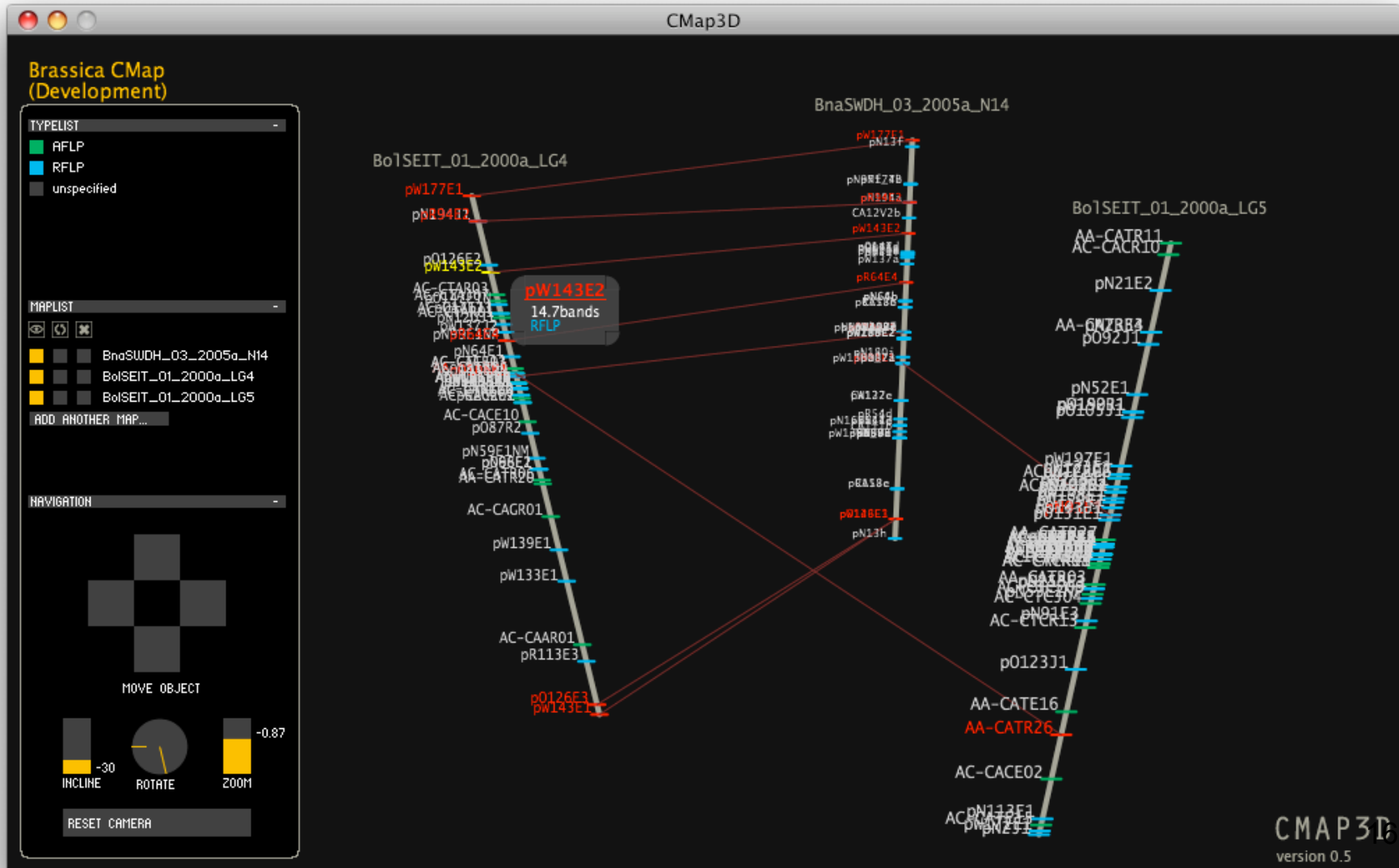
From genetic to physical maps



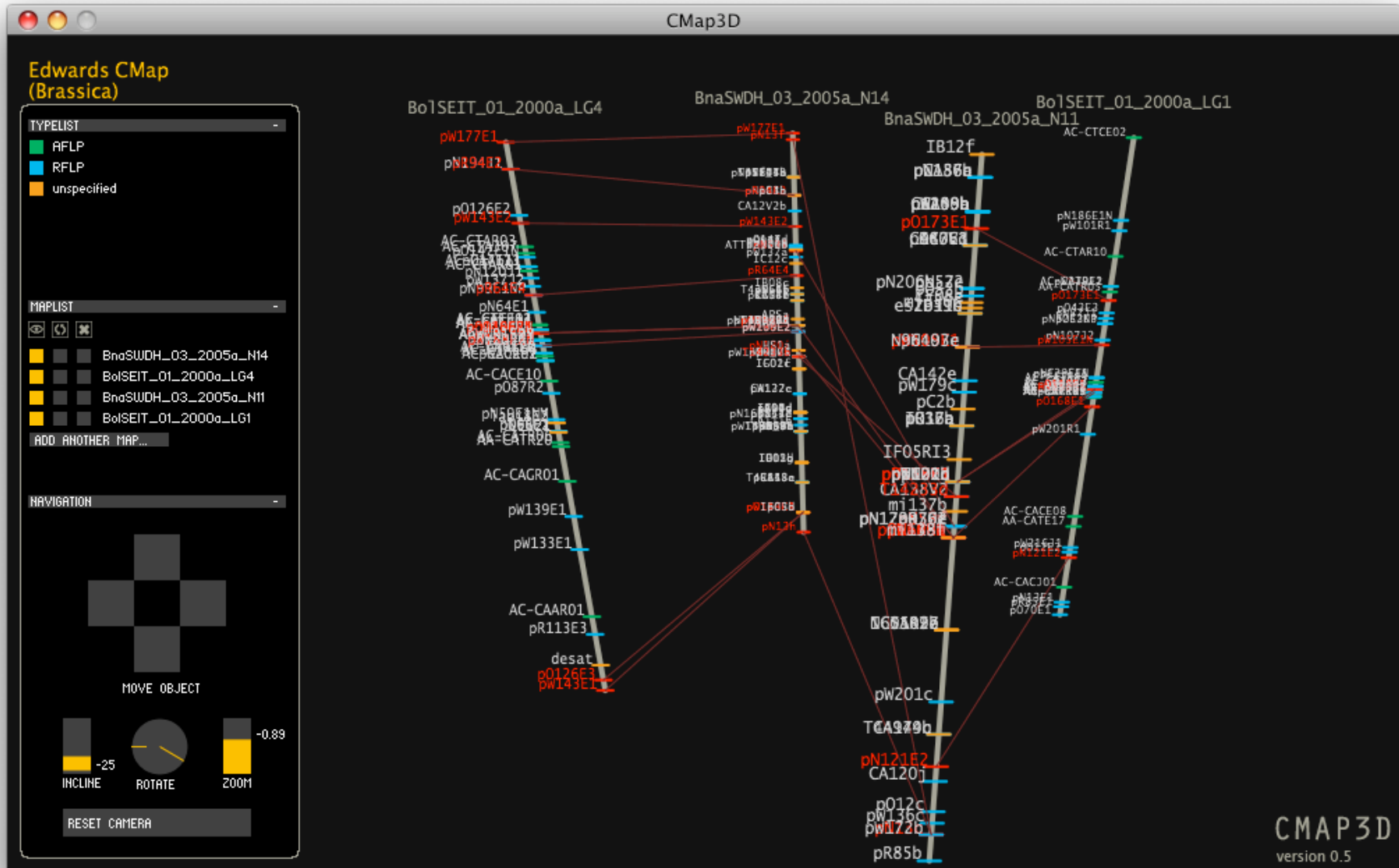
id	seq	qual	hits	pair	length	align	ref	loc	types				
pr3_FITO444	TTGCGGAGCATAGCCTTTAG	hhhhhhhhhhhhhhhhhhhh	1	a	20	-	Scaffold000001	1448812	0 20M		20		
pr5_CB10200	GGGCACTATGGGAAAATTA	hhhhhhhhhhhhhhhhhhhh	1	a	19	+	Scaffold000001	1500093	0 19M		19		
pr3_CB10200	TTGAACGGATTTTCCAA	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	Scaffold000001	1500275	0 18M		18		
pr5_FITO055	TGCTTCATCTTCTTCTTCTTG	hhhhhhhhhhhhhhhhhhhh	1	a	21	-	Scaffold000001	1690566	0 21M		21		
pr5_FITO021	TTTACAAGTCTCCTCCTCCTC	hhhhhhhhhhhhhhhhhhhh	2	a	21	+	Scaffold000001	1745559	2 A->1T40	C->7G40	21M		1A5C13
pr3_FITO001	TACCACTCTCCTCCACCACC	hhhhhhhhhhhhhhhhhhhh	1	a	20	-	Scaffold000001	1745609	0 20M		20		
pr5_CB10636	GTCGCAACACTCCTTGAG	hhhhhhhhhhhhhhhhhhhh	1	a	18	+	Scaffold000001	1864309	2 C->1T40	T->2C40	18M		1CT15
pr5_FITO349	GACGATGATAAGGATGAGGA	hhhhhhhhhhhhhhhhhhhh	1	a	20	+	Scaffold000001	1889134	2 G->9A40	T->2C40	20M		2T17
pr5_Ra3F07	GCACACACAAGATGCCACC	hhhhhhhhhhhhhhhhhhhh	1	a	19	-	Scaffold000001	2066861	2 C->0G40	A->3C40	19M		0C2A15
pr5_Ni4F06	AGATTCGATTCAAAGTGTGG	hhhhhhhhhhhhhhhhhhhh	1	a	21	+	Scaffold000001	2084727	0 21M		21		
pr3_Na12E01	CGACAGAGACAAGGGATT	hhhhhhhhhhhhhhhhhhhh	1	a	19	-	Scaffold000001	2176076	2 T->1G40	A->5G40	19M		1T3A13
pr5_CB10435	GGTCTGTCAAGCGTCTA	hhhhhhhhhhhhhhhhhhhh	1	a	18	+	Scaffold000001	2187790	2 A->1G40	A->4C40	18M		1A2A13
pr5_OI10F05	TCTATCAAATTTGTTTCTCAGGC	hhhhhhhhhhhhhhhhhhhh	1	a	23	+	Scaffold000001	2316408	0 23M		23		
pr3_OI10F05	GCCATTGAAGAACGCAATG	hhhhhhhhhhhhhhhhhhhh	1	a	19	-	Scaffold000001	2317098	0 19M		19		
pr5_CB10372	AAGCTGTTCAAGCAAGCTG	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	Scaffold000001	2509833	0 18M		18		
pr3_Ni3F02	ACCATTGAAACGTTGAACCC	hhhhhhhhhhhhhhhhhhhh	1	a	20	+	Scaffold000001	2930124	2 G->1C40	G->5T40	20M		1G3G14
pr3_Na10B01	GGTCTGGTTTCTTCGATCGA	hhhhhhhhhhhhhhhhhhhh	1	a	20	-	Scaffold000001	2996263	0 20M		20		
pr3_CB10057	TAGTTTCTTCTCCTGCTATC	hhhhhhhhhhhhhhhhhhhh	1	a	21	+	Scaffold000001	3062536	0 21M		21		
pr5_CB10057	TGACAATCTTCCTTAGCCTAG	hhhhhhhhhhhhhhhhhhhh	1	a	21	-	Scaffold000001	3062748	0 21M		21		
pr5_FITO444	TTGTTGTTGTTGTTGTTGTTG	hhhhhhhhhhhhhhhhhhhh	29	a	21	-	Scaffold000001	3245771	0 21M		21		
pr5_FITO380	GTTCCTTGCCACCAATG	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	Scaffold000001	3264971	2 C->1T40	T->3C40	18M		1C1T14
pr5_CB10423	CAATCATGCATTATGCC	hhhhhhhhhhhhhhhhhhhh	1	a	18	+	Scaffold000001	3364589	0 18M		18		
pr3_CB10423	GTGGTGATCTTTCCTTG	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	Scaffold000001	3364778	0 18M		18		
pr5_OI13E03	CTCTGTTTCTCCTGCCCAAC	hhhhhhhhhhhhhhhhhhhh	1	a	20	+	Scaffold000001	3371580	0 20M		20		
pr5_BRAS038	AGAAAGTGGTAACAGCTCA	hhhhhhhhhhhhhhhhhhhh	1	a	19	-	Scaffold000001	3516013	0 19M		19		
pr5_CB10329	AATCATCGAGTGGACGAA	hhhhhhhhhhhhhhhhhhhh	1	a	18	+	Scaffold000001	3539302	0 18M		18		
pr3_CB10329	GACCCGGAGTGATCTCA	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	Scaffold000001	3539409	0 18M		18		
pr5_CB10008	AATCAACAAAGAAGAAGAAG	hhhhhhhhhhhhhhhhhhhh	2	a	20	-	Scaffold000001	3546084	2 A->2T40	A->6C40	20M		2A3A13

Ordered subset of SOAP2 output, with matching primer pairs highlighted

Brassica CMap3D



Brassica CMap3D



Brassica CMap

- 23 map sets
- 318 linkage groups
- 4899 markers

Summary

- There are a lot of useful things you can do with short paired read sequence data
- Use CMap3D to link Brassica genetics and genomics
- Tools available at: <http://flora.acpfg.com.au/>
(or type ACPFG bioinformatics into Google)



Acknowledgements



Australian Government
Australian Research Council



Paul Berkman
Lauren Bragg
Terry Clark
Dominic Eales
Chang Pyo Hong
Michael Imelfort
Edmund Ling
Megan McKenzie
Jiri Stiller
David Edwards

Jacqueline Batley
Xiaowu Wang
Harsh Raman
Kaye Basford



Daniel Marshall
Nikki Appleby
Ping Zhang
Zoran Boskovic

