Physical Mapping and Shotgun Sequencing of Wheat Chromosome 2A

Department of Biotechnology
Ministry of Science and Technology

Kuldeep Singh, Punjab Agril. Univ. Ludhiana India

IWGSC, Sept 7 2013 Yokohama Japan

Team India

- Punjab Agricultural University, Ludhiana (Kuldeep Singh)
- National Research Centre for Plant Biotechnology (NRCPB), New Delhi (NK Singh)
- University of Delhi, South Campus (UDSC), New Delhi (JP Khurana)

Team India

PAU Team

Principal Investigator: *Dr. Kuldeep Singh,*

Co-investigators: *Dr. P Chhuneja,*
Dr. Yogesh Vikal,
Dr. N.S. Bains,
Dr. O.P. Gupta,
Dr. Priti Sharma
Dr. Neelam Singh
Inderjit Singh Yadav

Research Fellows *Bharat Yadav*
Palvi Malik
Abhishek Bhatnagar

Ph.D. Students *Suruchi Jindal*
Parampreet Kaur

NRCPB Team

Principal Investigator: *Dr. N.K. Singh*
Co-investigators: *Dr. T. R. Sharma*
Dr. A. Mithra
Dr. Kishore Gaikwad
Research Fellows/Ph D students:
Ajay Mohato
Vivek Dogra

UDSC Team

Principal Investigator: *Dr. J P Khurana*
Co-investigators: *Dr. P Khurana*
Dr. Saloni Mathur
Research Fellows/Ph D students:
Navin Sharma
C Chowdhary

IWGSC Time Lines

1. Complete Fingerprinting and Generation of the Physical map to meet IWGSC standards - 2013
2. Anchoring Physical Map to the Genetic map – 2014
3. Enter the 2nd phase
4. Sequencing of the MTP BACs – 2015/16

OBJECTIVES

- To generate chromosome 2AL- and 2AS-specific BAC library with 15X coverage (in collaboration with J. Dolezel, IEB, Czech Republic)
- Fingerprinting of 2AL specific BACs using SNaPShot Technology
- End-sequencing of Chr 2A specific BAC clones using Sanger's method
- Generating high density linkage map for chromosome 2A using SSR, EST, ISBP and other types of markers
- Anchoring contigs to genetic maps using SSRs, ESTs and other markers (generated by network partners) for generating a 2A physical map

OBJECTIVES

- Whole genome shotgun sequencing of flow-sorted chromosome 2AS and 2AL using NextGen Sequencing Technologies
- Generating genome assembly and annotation in collaboration with other network partners
- Integrated draft sequence, identification of markers and discovery of genes
- Human-resource development and institutional capacity building in genomics

Progress made against the Approved Objectives, Targets & Timelines during the Reporting Period

- Shotgun sequencing and Assembly
- HICF of BACs and generation of MTP
- BAC-end Sequencing
- Marker Development and Mapping

2AS NGS DATA

>300 X Illumina Data

- 2 paired end runs from GAII
- 4 paired end run from HiSeq 2000

5.5 X Roche 454 Data

- 5.5X single end run from 454 Flx sequencer

2AL NGS DATA

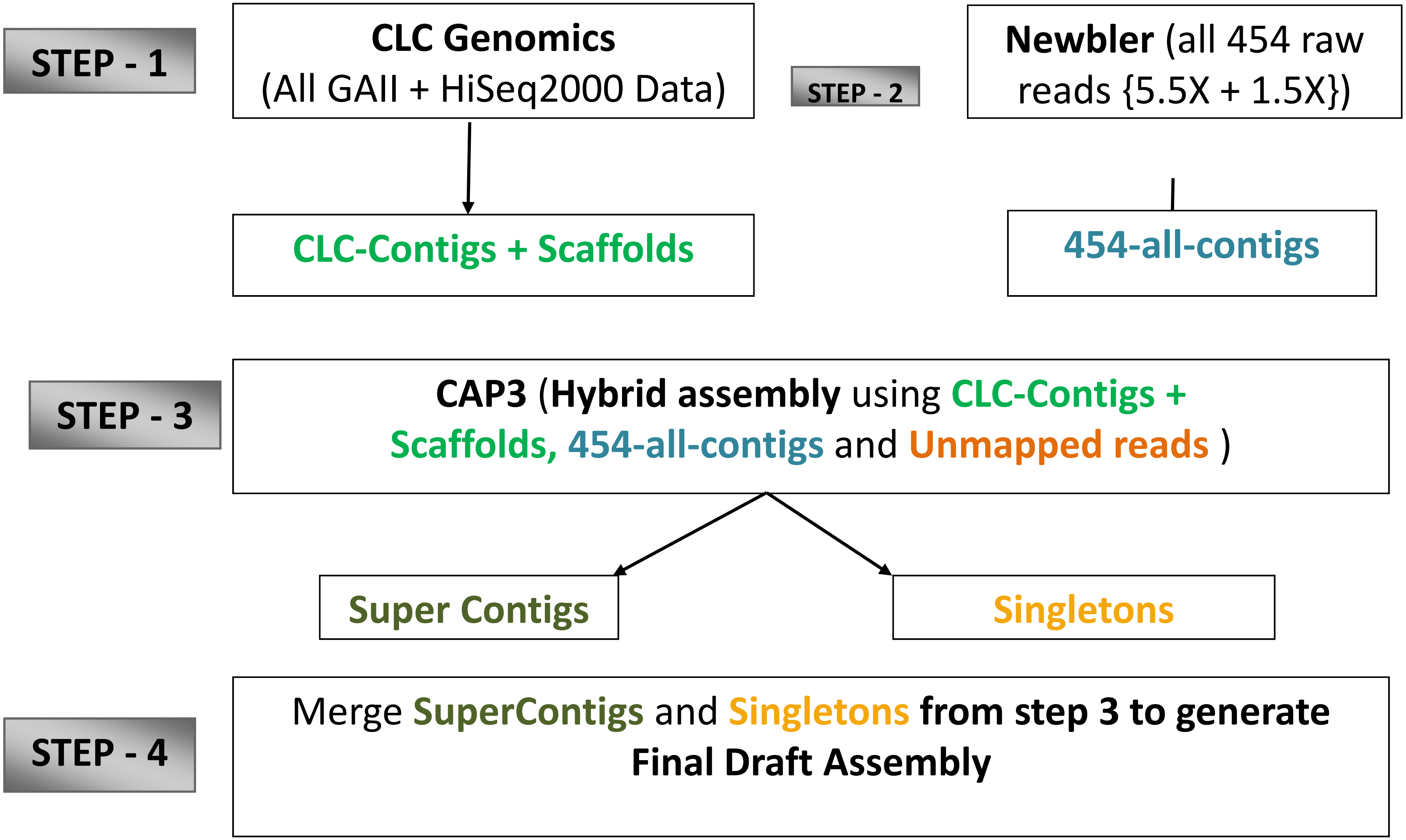
96X Illumina Data

- 5 paired end runs from GAII
- 1 paired end run from HiSeq 2000

7.0 X Roche 454 Data

- 5.5X single end run from 454 Flx sequencer
- 1.5X single end run from 454 Flx sequencer

2A ASSEMBLY STRATEGY



STEP - 1

CLC Genomics
(All GAII + HiSeq2000 Data)

CLC-Contigs + Scaffolds

STEP - 2

Newbler (all 454 raw reads {5.5X + 1.5X})

454-all-contigs

STEP - 3

CAP3 (**Hybrid assembly** using **CLC-Contigs + Scaffolds, 454-all-contigs** and **Unmapped reads**)

Super Contigs

Singletons

STEP - 4

Merge **SuperContigs** and **Singletons** **from step 3 to generate Final Draft Assembly**

2A NGS DATA ASSEMBLY STATISTICS

Statistics	2AS		2AL	
	Illumina	Roche 454	Illumina	Roche 454
Total Number of Raw Reads	1,116,247,283 (113,113,316,981 bp)	7,296,111 (1,992,327,716 bp)	448,978,394 (48,448,861,006 bp)	11,412,646 (3,505,104,650 bp)
Total Number of Trimmed Reads	1,066,201,401 (102,759,183,422 bp- 90.8 %)	4,568,809 (1,168,530,853 bp - 58.7 %)	447,847,122 (45,359,789,456 bp- 93.6 %)	10,722,504 (3,130,752,253 bp - 89.3 %)
Reads included in assembly	1,029,024,103	4,568,420	428,747,015	10,706,898
Unmapped Reads	*37,177,298*	*389*	*19,100,107*	*15,606*
Total Number of Contigs and scaffolds	573,520	106,241	444,331	45,163
Total Number of bases (Contigs + Scaffolds)	334,820,432 bp	105,751,694 bp	278,199,919 bp	68,048,619 bp
Average Contig length	602 bp	995 bp	626 bp	1,506 bp
Maximum Contig Length	41,627 bp	18,719 bp	73,233 bp	8,091 bp
N50	806	1029	851	1469

Contig Assembly Programme (CAP3) STATISTICS

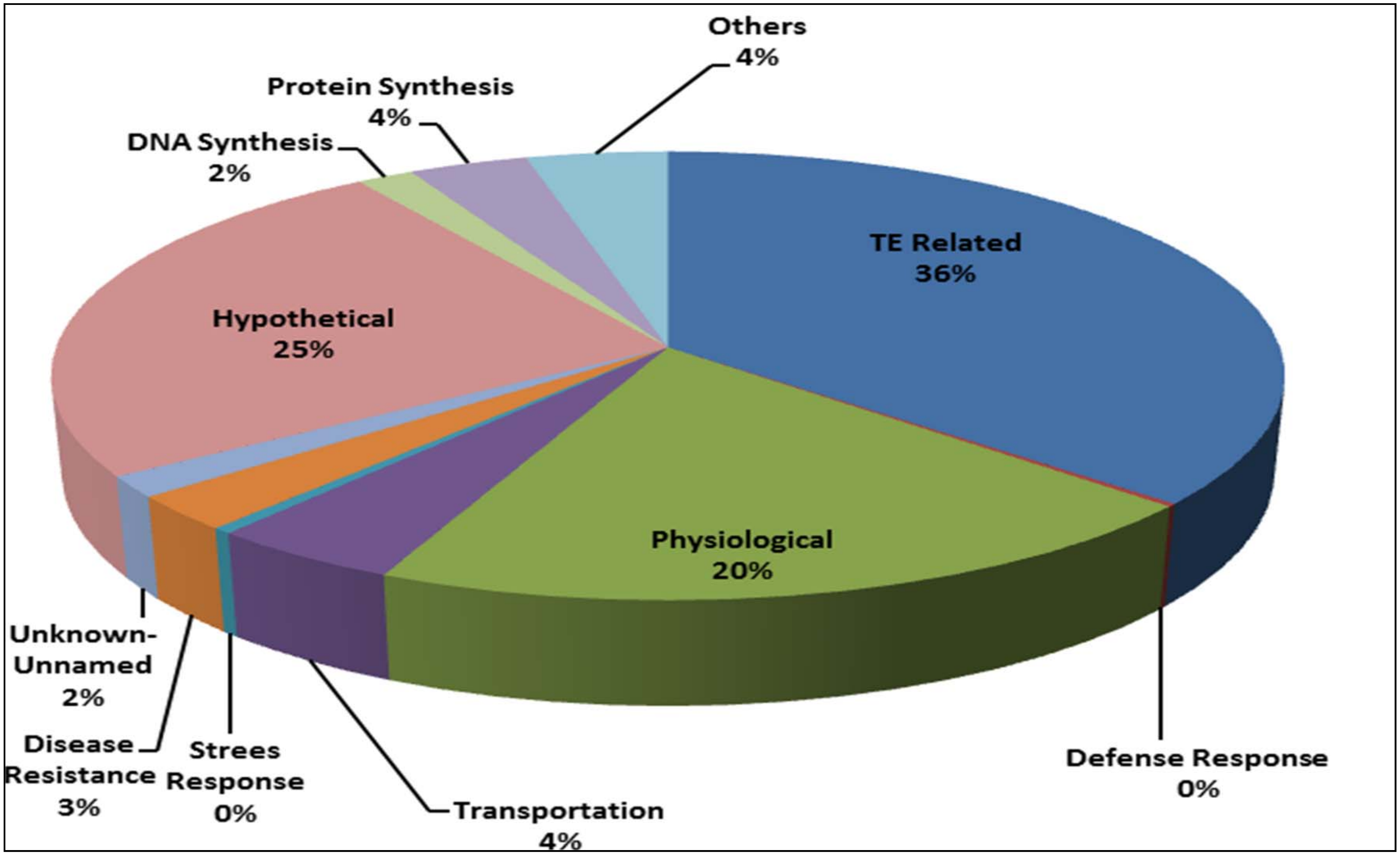
Statistics from CAP3	**2AS**	**2AL**
Total number of Contigs and Scaffolds (Illumina, 454)+ Singletons (Illumina, 454)	679,761 + 37,177,687 = **37,857,448**	489,494 + 19,115,713 = **19,605,207**
Total No of Super-contigs Generated in assembly	**13,013**	**3,314**
Total No of CAP3 Singletons **Unassembled cap3 454/Illumins contigs** **Unassembled 454/Illumina trimmed reads**	**17,215,201** **656,931** 16,558,270	**11,235,698** **422,507** 10,813,191
Draft 2A (CAP3 contigs+ Unassembled CAP3 454/Illumina contigs)	**669,944** **(438,567,520 bp)**	**425,821** **(323,508,104 bp)**
Average Contig Size	670 bp	759 bp
Largest Contig Size	41,627 bp	73,233 bp
N 50	900 bp	1,078 bp

Next Step – Short Oligonucleotide Analysis Package (SOAP) denovo

GENE PREDICTION AND ANNOTATION OF 2A

S. No	Category	**No. of Sequences**
1	**TE Related**	**10,190**
2	**Defense Response**	**57**
3	**Physiological**	**5683**
4	**Transportation**	**1264**
5	**Stress Response**	**123**
6	**Disease Resistance**	**730**
7	**Unknown-Unnamed**	**436**
8	**Hypothetical**	**6947**
9	**DNA Synthesis**	**477**
10	**Protein Synthesis**	**994**
11	**Others**	**1,179**
Total		**28,080**

GENE PREDICTION AND ANNOTATION OF 2A



Comparison with other wheat chromosomes

Chromosome	Platform	Size of Genome	No of Contigs / scaffolds	N50	% Coverage
2AL	**454 Roche + illumina**	**508 Mb**	**425,821**	**1078**	**64.2**
5A	**454 Roche**	**857.8 Mb**	**1,999,592**	**795**	**71.0**
7DS	**Illumina - GAIIx**	**381 Mb**	**571,038**	**1159**	**40.0**

IDENTIFICATION OF REPETITIVE ELEMENTS

APPROACH-1 Homology Based Repeat Masking - Repeat Masker

- TREP – Triticeae Repeat Element Data base – (1717 RE’s)
- RepBase (Grass: 2422 RE’s, Oryza: 578 RE’s)

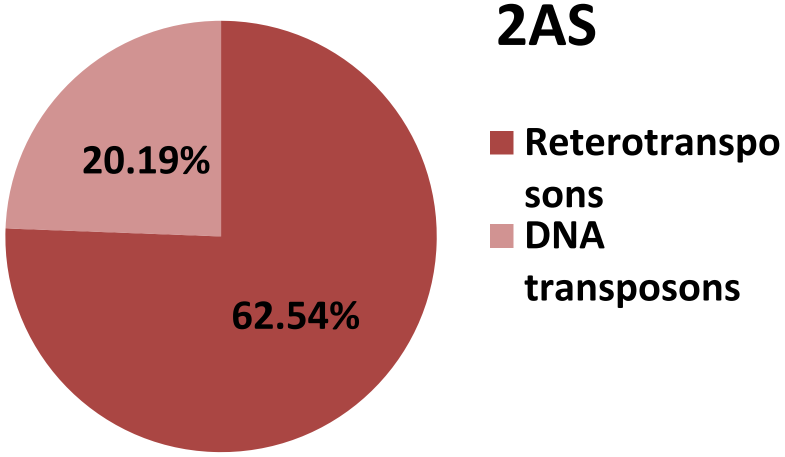
APPROACH-2 De novo Repeat Masking - RepeatModeler

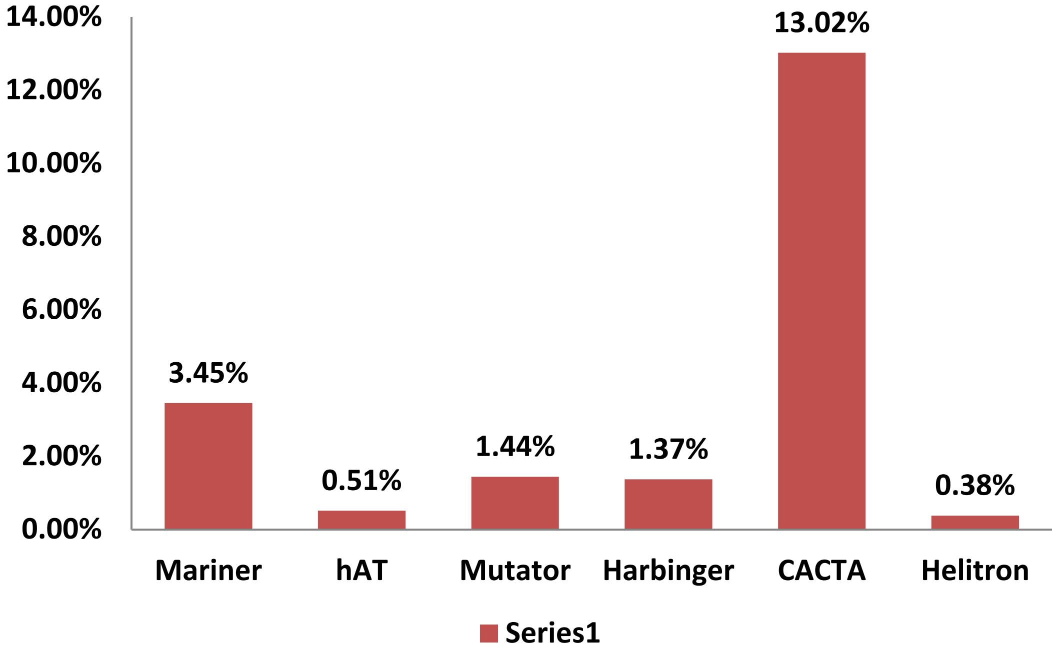
Homology Based Approach Results – 2A

Class I (Retro Transposons)			**2AS**				**2AL**			
			TREP	**Grass-rep**	**Oryza-rep**	**ALL**	**TREP**	**Grass-rep**	**Oryza-rep**	**ALL**
LTR										
	Copia	**RCL**	64361	5608	248	**70217**	54595	4417	312	**59324**
	Gypsy	**RGL**	264235	17065	346	**281646**	196577	11916	227	**208720**
LINE										
	RTE	**RIT**		12		**12**		10		**10**
	L1	**RIL**		10708	1102	**11810**		7337	782	**8119**
SINE					128	**128**			76	**76**
	tRNA	**RST**		619	163	**782**		412	101	**513**
Class II (DNA Transposon) Subclass -1										
TIR										
	Tc1 / Mariner	**DTT**	19904	88	135	**20127**	14389	70	86	**14545**
	hAT	**DTA**	1109	1738	149	**2996**	700	1370	114	**2184**
	Mutator	**DTM**	8397			**8397**	5636			**5636**
	PIF - Harbinger	**DTH**	6986	775	273	**8034**	5464	642	192	**6298**
	CACTA	**DTC**	75940			**75940**	52207			**52207**
Class II (DNA Transposon) Subclass -2										
Hilitron										
	Hilitorn	DHH	1237	885	89	**2211**	734	684	82	**1500**

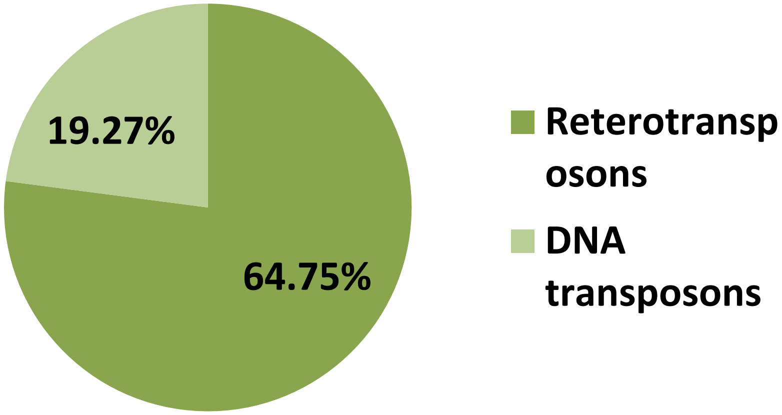
TOTAL **Including Unclassified** **582955** **427402**

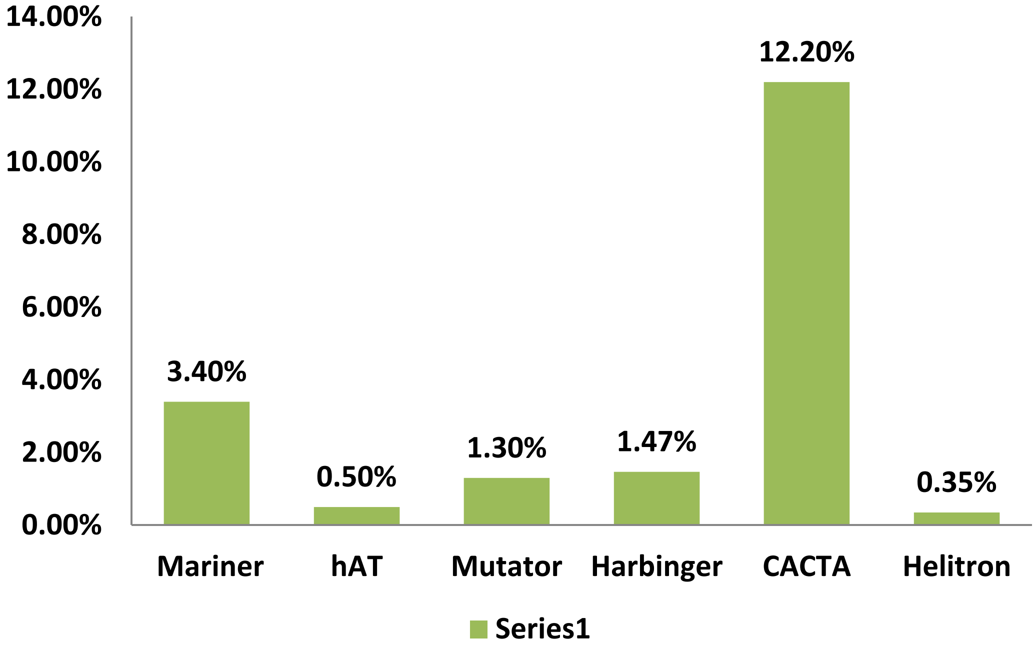
2AS


20.19%
62.54%
Reterotransposons
DNA transposons


14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
3.45%
0.51%
1.44%
1.37%
13.02%
0.38%
Mariner
hAT
Mutator
Harbinger
CACTA
Helitron
Series1

2AL


19.27%
64.75%
🟩 Reterotransposons
🟩 DNA transposons


14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
3.40%
0.50%
1.30%
1.47%
12.20%
0.35%
Mariner
hAT
Mutator
Harbinger
CACTA
Helitron
🟩 Series1

HICF STRATEGY

384 well Master plate (BAC's supplied as STAB)

↓

Maintenance of BAC Clones (Primary Cultures)

↓

Multiplication of BAC Clones and BAC DNA isolation

↓

Restriction Digestion, Labeling and addition of Liz-1200

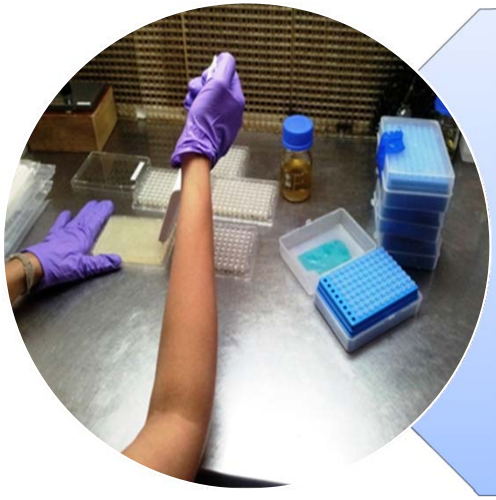
↓

BDX-SAM Purification

↓

High Information Content Fingerprinting

STEP 1: Maintenance Of BAC Clones



Primary Culture

2 YT Medium + Glycerol (7.5%) + Chloramphenicol (12.5 µg/mL) + 2-3ul inoculum

Incubation conditions: 37^{0}C, 160 rpm, 16 hours

STEP 2: Multiplication of BAC Clones and BAC DNA Isolation

Secondary Culture

2 YT Medium + Chloramphenicol (12.5 µg/mL) + 2-3µl inoculum

Incubation conditions: 37^{0} C, 160 rpm, 16 hours

STEP 3: Restriction Digestion and Labelling

Reagent	**1X (µl)**	**220X (µl)**
Buffer	1	220
Water	0.46	102
RNase (10 mg/ml)	0.05	11
100X BSA	0.05	11
EcoRI (1U)	0.01	2.2
BamHI(1U)	0.01	2.2
XhoI(1U)	0.01	2.2
XbaI(1U)	0.01	2.2
HaeIII(1U)	0.1	22
SNaPshot	0.2	44
Beta-Merc (0.01%)	0.1	22
DNA (100 ng)	8	
Total	10	

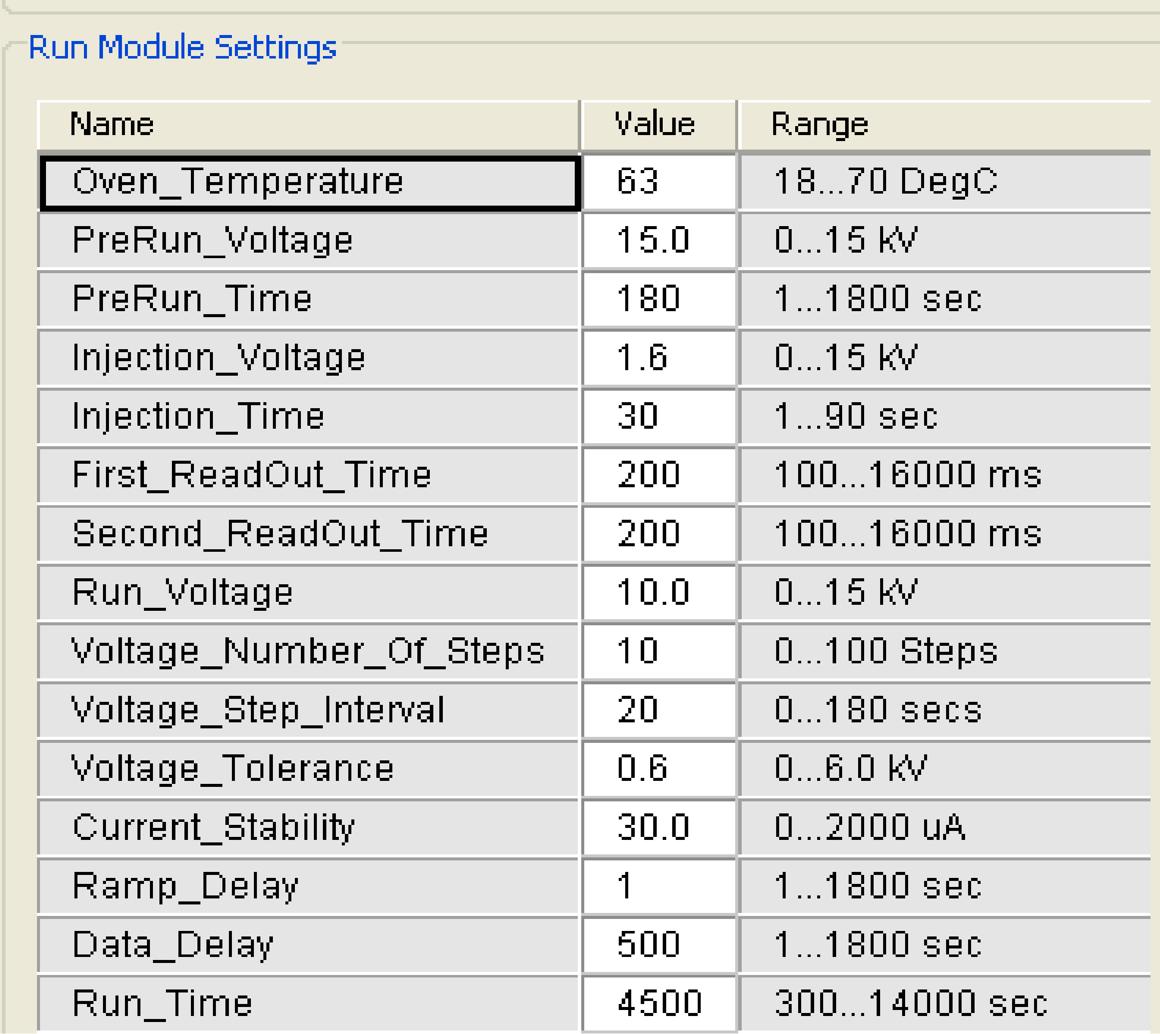
STEP 4: Addition of Size Standard

Reagent	1X (µl)	220X (µl)
Hidi-Formamide	9.7	2134
Liz-1200	0.3	66
Denaturation at 95^0 C for 3 minutes followed by Snapchill		

STEP 5: Purification

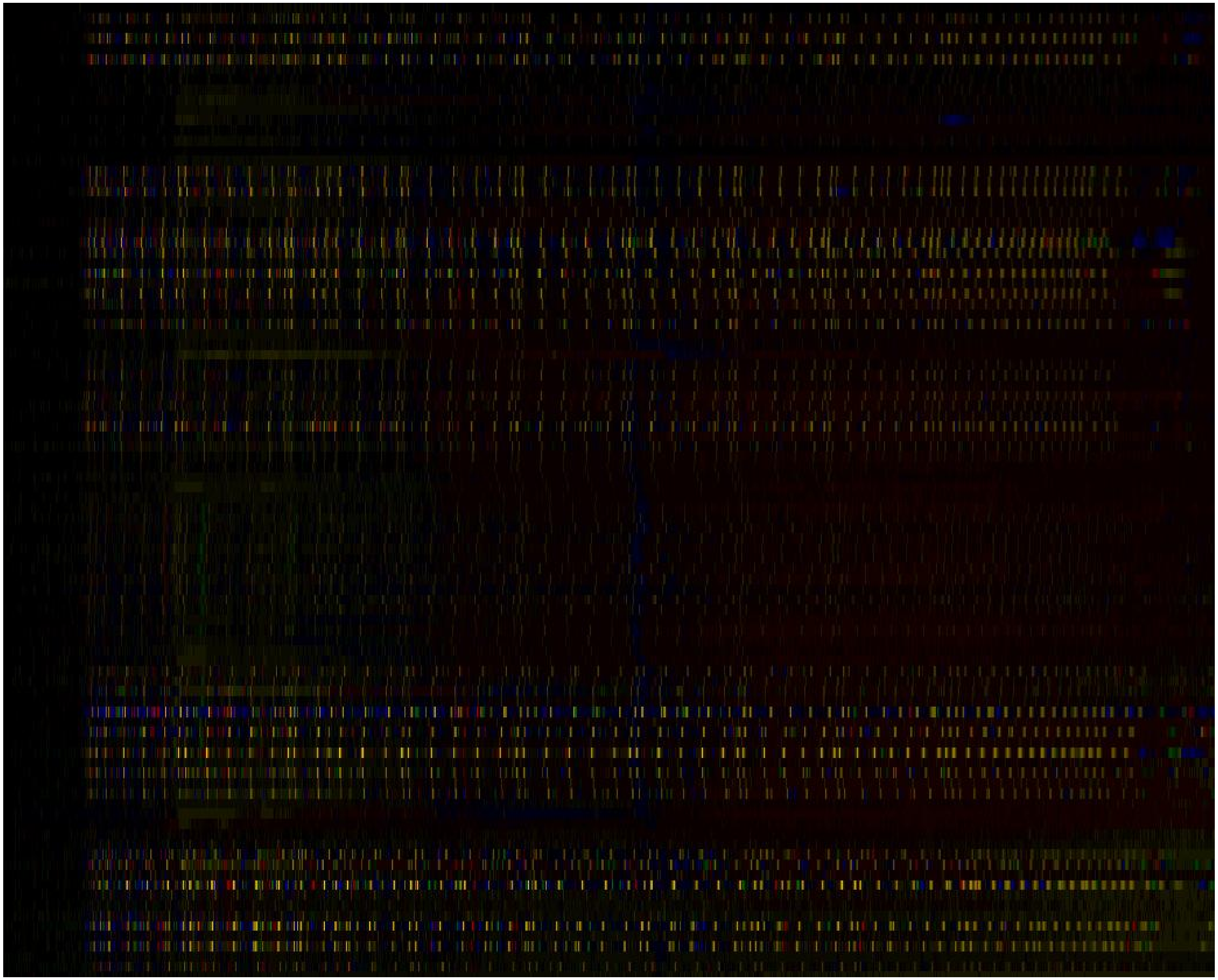
Reagent	1X (µl)	220X (µl)
BDX	4	880
SAM	16	3520
Vortex at 1800 rpm for 30 minutes		

STEP 6: Addition of Plates to ABI 3730xL DNA Analyzer


Run Module Settings

Name	Value	Range
Oven_Temperature	63	18...70 DegC
PreRun_Voltage	15.0	0...15 kV
PreRun_Time	180	1...1800 sec
Injection_Voltage	1.6	0...15 kV
Injection_Time	30	1...90 sec
First_ReadOut_Time	200	100...16000 ms
Second_ReadOut_Time	200	100...16000 ms
Run_Voltage	10.0	0...15 kV
Voltage_Number_Of_Steps	10	0...100 Steps
Voltage_Step_Interval	20	0...180 secs
Voltage_Tolerance	0.6	0...6.0 kV
Current_Stability	30.0	0...2000 uA
Ramp_Delay	1	1...1800 sec
Data_Delay	500	1...1800 sec
Run_Time	4500	300...14000 sec

STEP 7: Array View (Earlier Arrays)



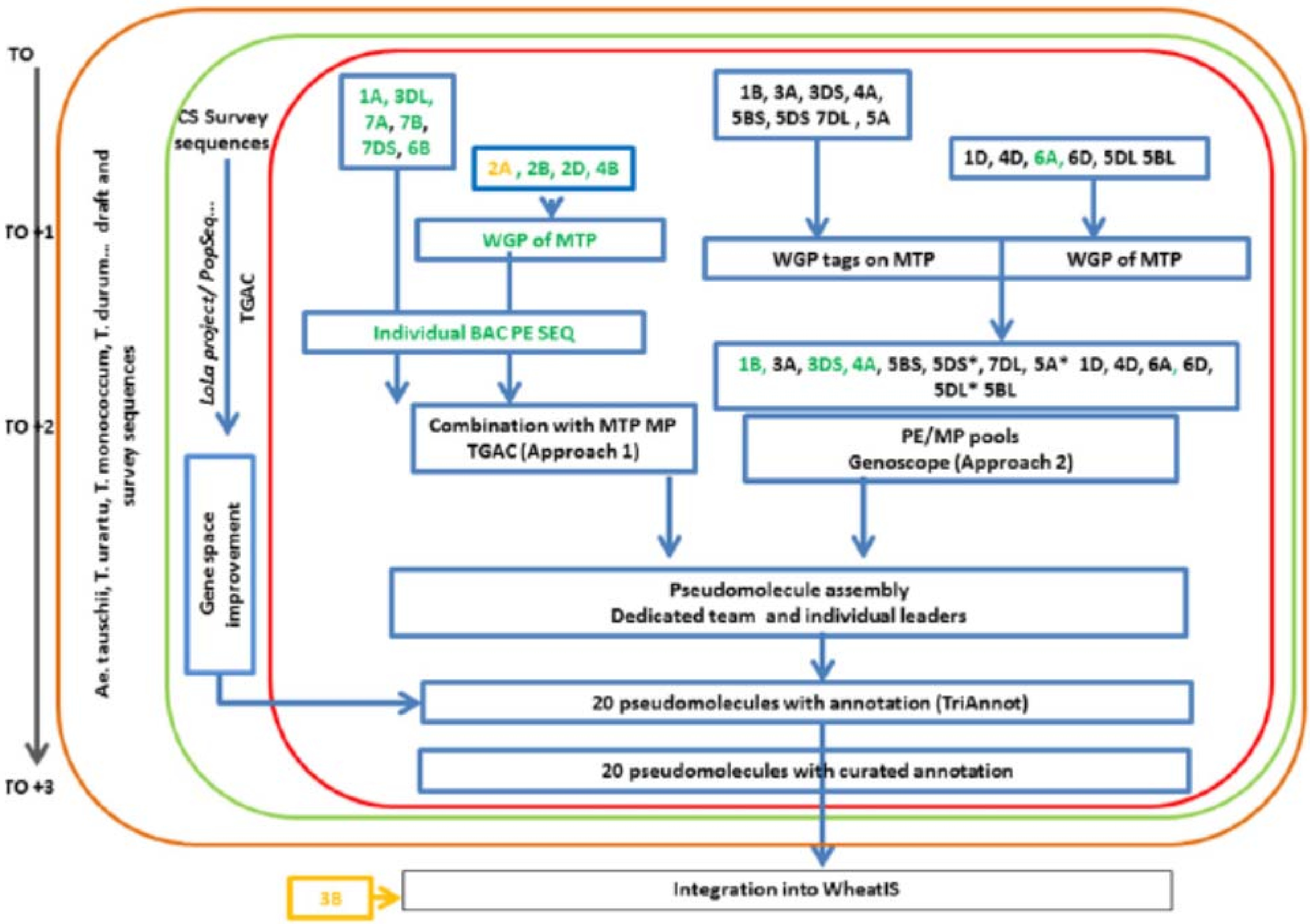
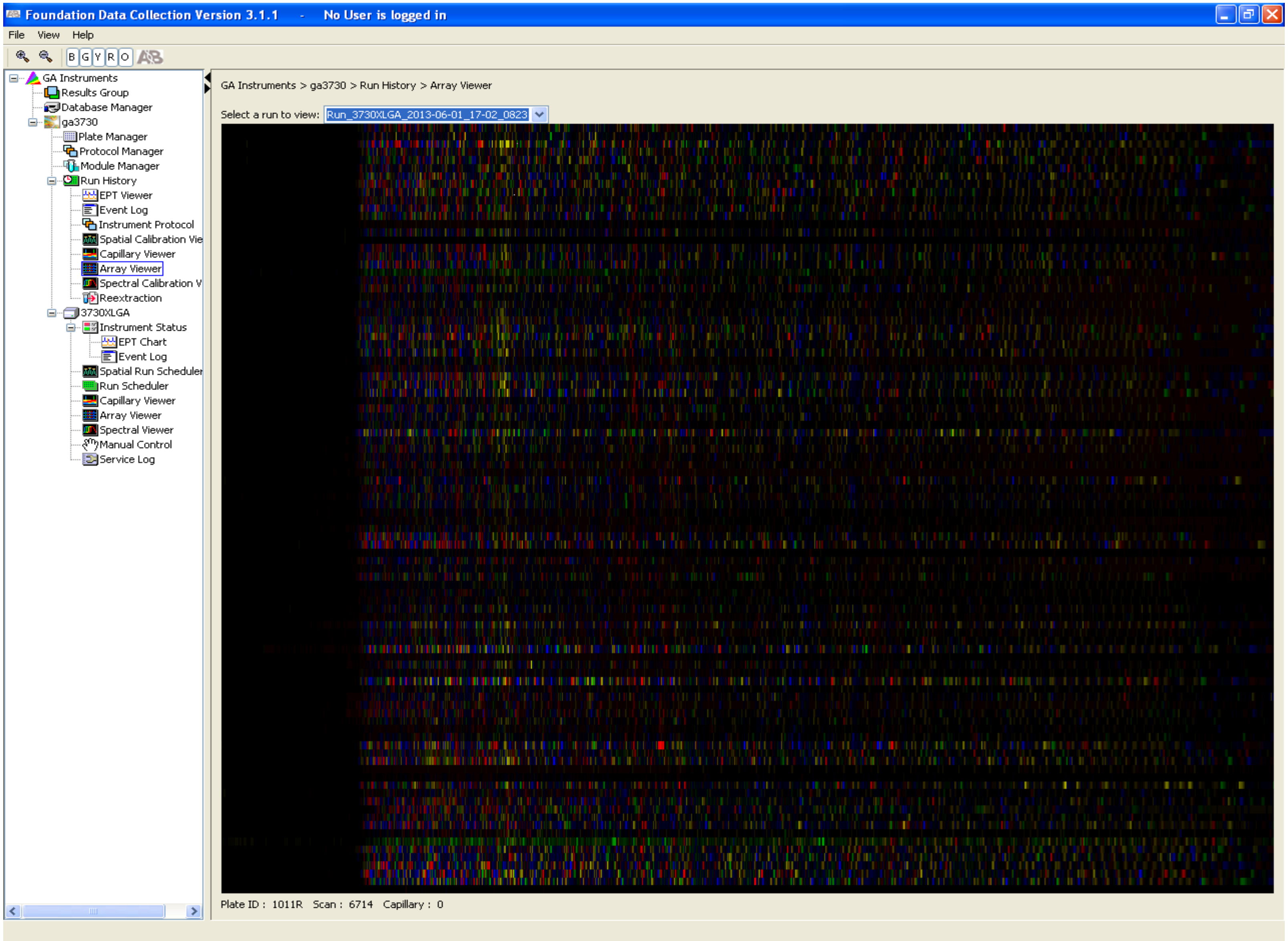


Figure 2. Project flow chart. Activities within (1) the red outline would be included in the WI-IWGSC project, (2) the green outline relates to the parallel IWGSC-TGAC integration of whole genome gene space survey sequences with the IWGSC CSS and gene models, and (3) the brown outline relates to collaborations with survey sequences of the D-genome and A-genome progenitors of bread wheat as well as Durum.

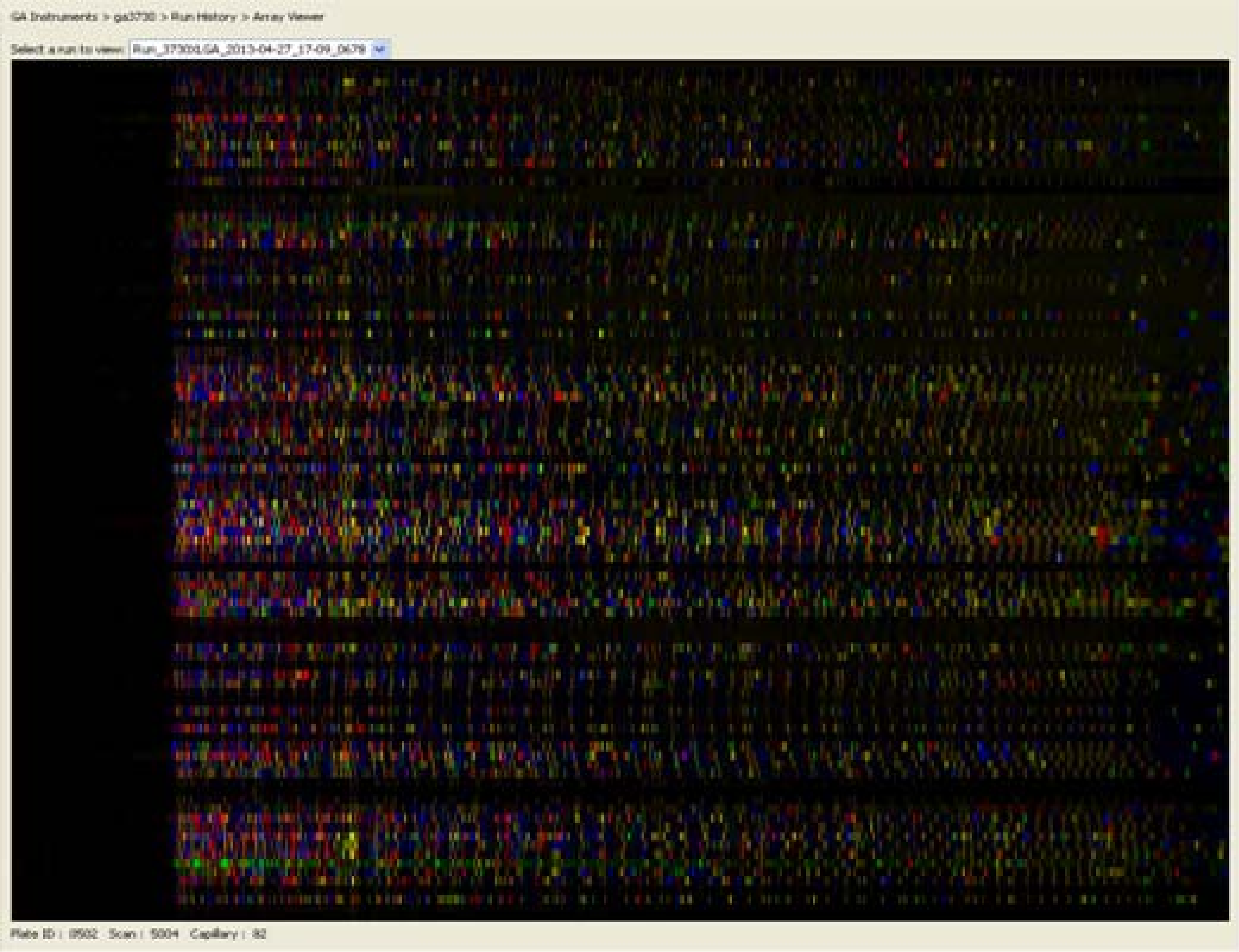
STEP 7: Array View



Select a run to view: Run_3730XLGA_2013-06-01_17-02_0823

Plate ID : 1011R Scan : 6714 Capillary : 0

STEP 7: Array View



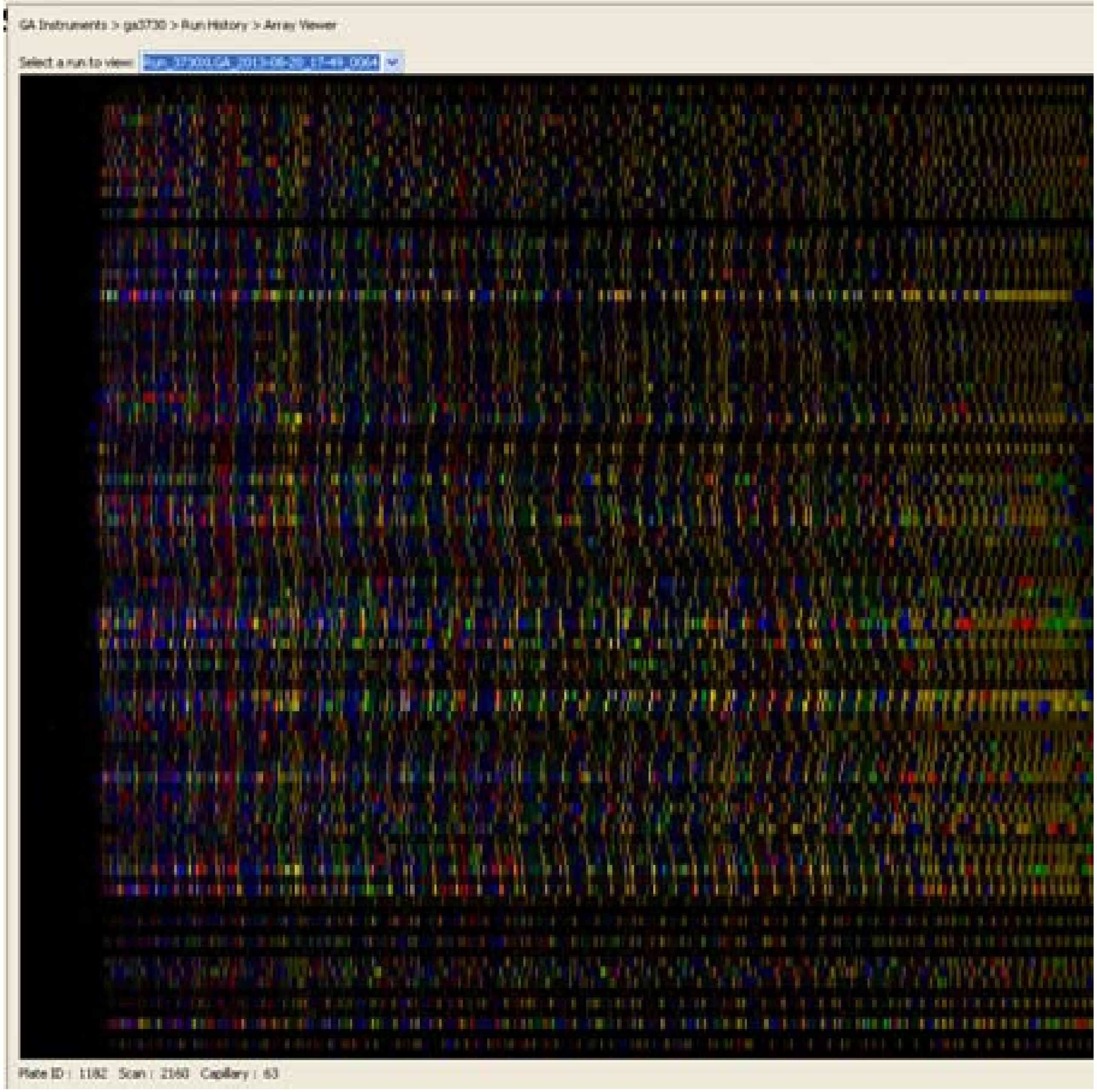
Select a run to view: Run_3730xlGA_2013-04-27_17-09_0679

Plate ID : 0502 Scan : 5004 Capillary : 82

STEP 7: Array View

Plate ID : 0291_20082013 Scan : 3216 Capillary : 1

STEP 7: Array View

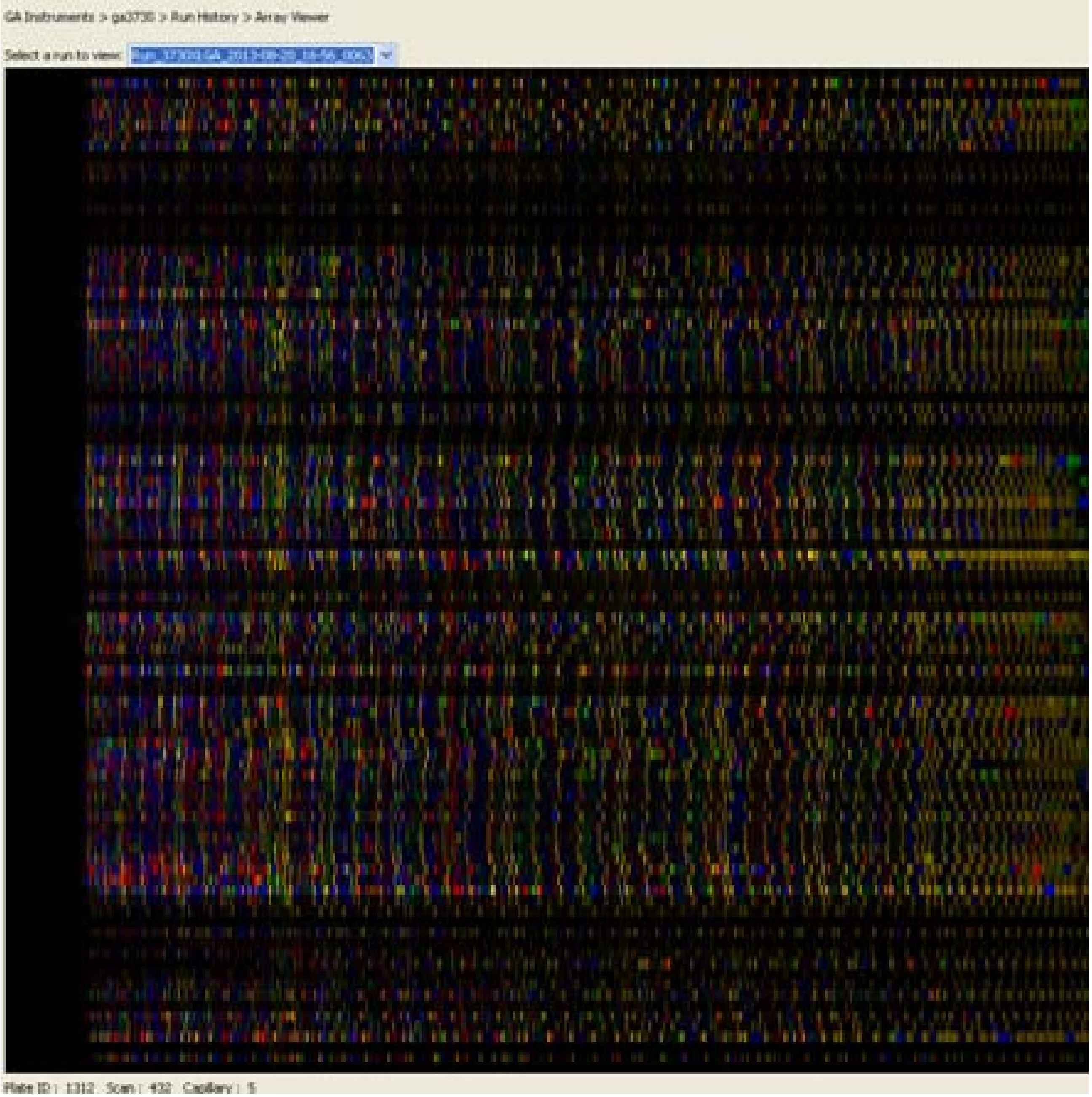


GA Instruments > ga3730 > Run History > Array Viewer

Select a run to view:

Plate ID : 1182 Scan : 2160 Capillary : 63

STEP 7: Array View

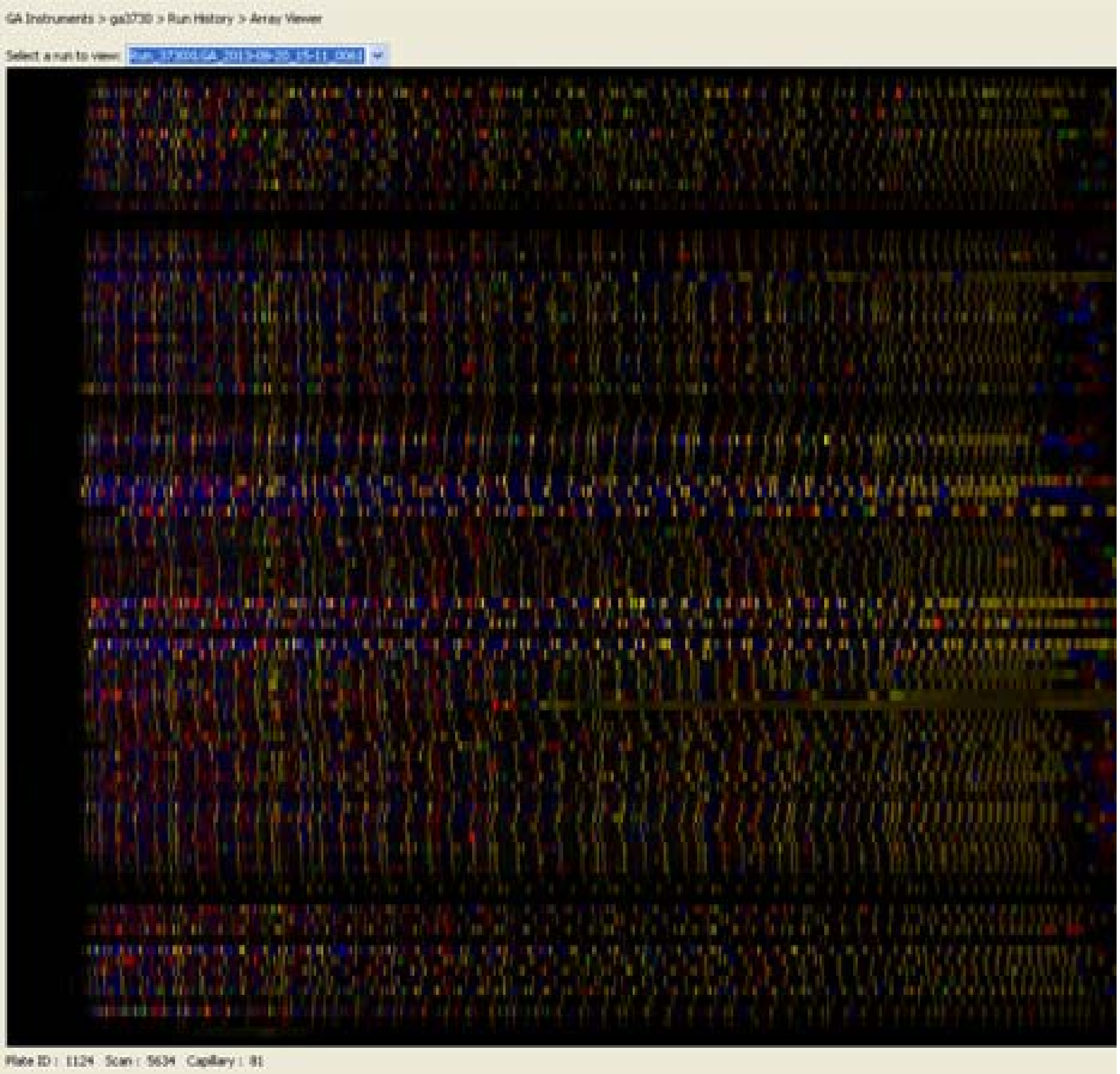


GA Instruments > ga3730 > Run History > Array Viewer

Select a run to view:

Plate ID : 1312 Scan : 432 Capillary : 5

STEP 7: Array View

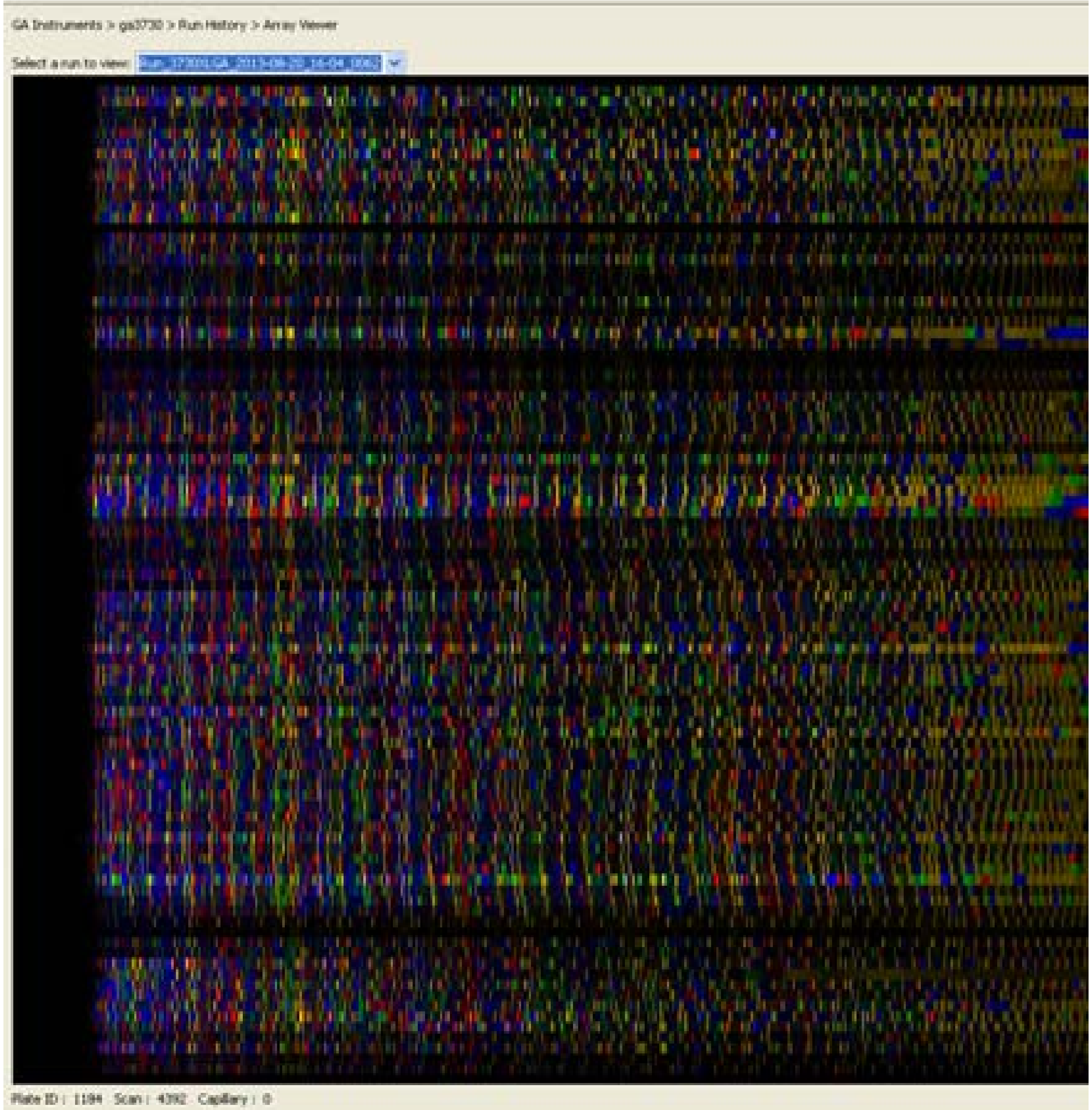


GA Instruments > ga3730 > Run History > Array Viewer

Select a run to view:

Plate ID : 1124 Scan : 5634 Capillary : 81

STEP 7: Array View

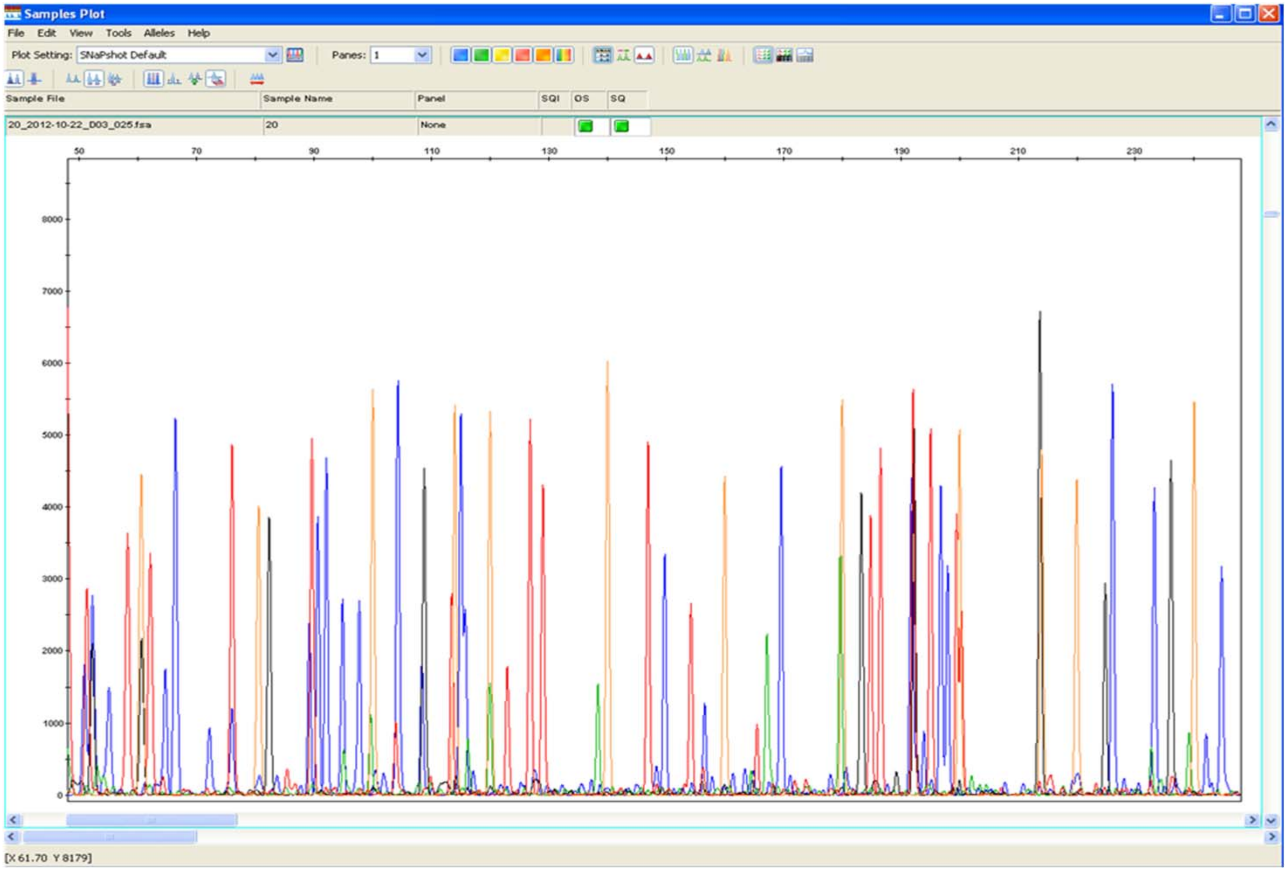


GA Instruments > ga3730 > Run History > Array Viewer

Select a run to view:

Plate ID : 1184 Scan : 4392 Capillary : 0

STEP 8: Gene Mapper Analysis



Plot Setting: SNaPshot Default

Panes: 1

Sample File	Sample Name	Panel	SQI	OS	SQ
20_2012-10-22_D03_025.fsa	20	None		🟩	🟩

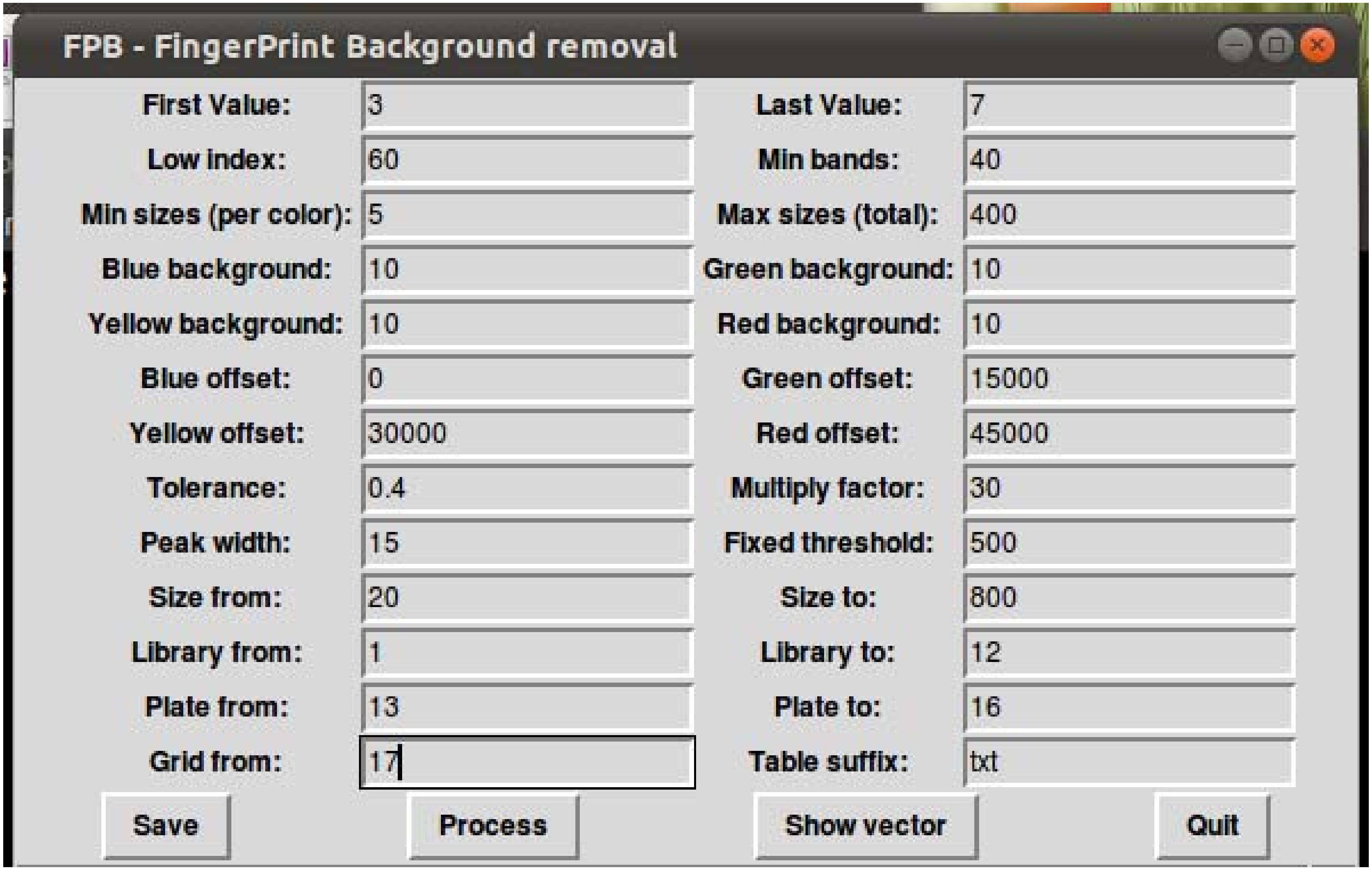
[X 61.70 Y 8179]

54,618 Clones of 2AL have been fingerprinted

Success rate : 60-70%

CONTIG ASSEMBLY USING FPC V9.4

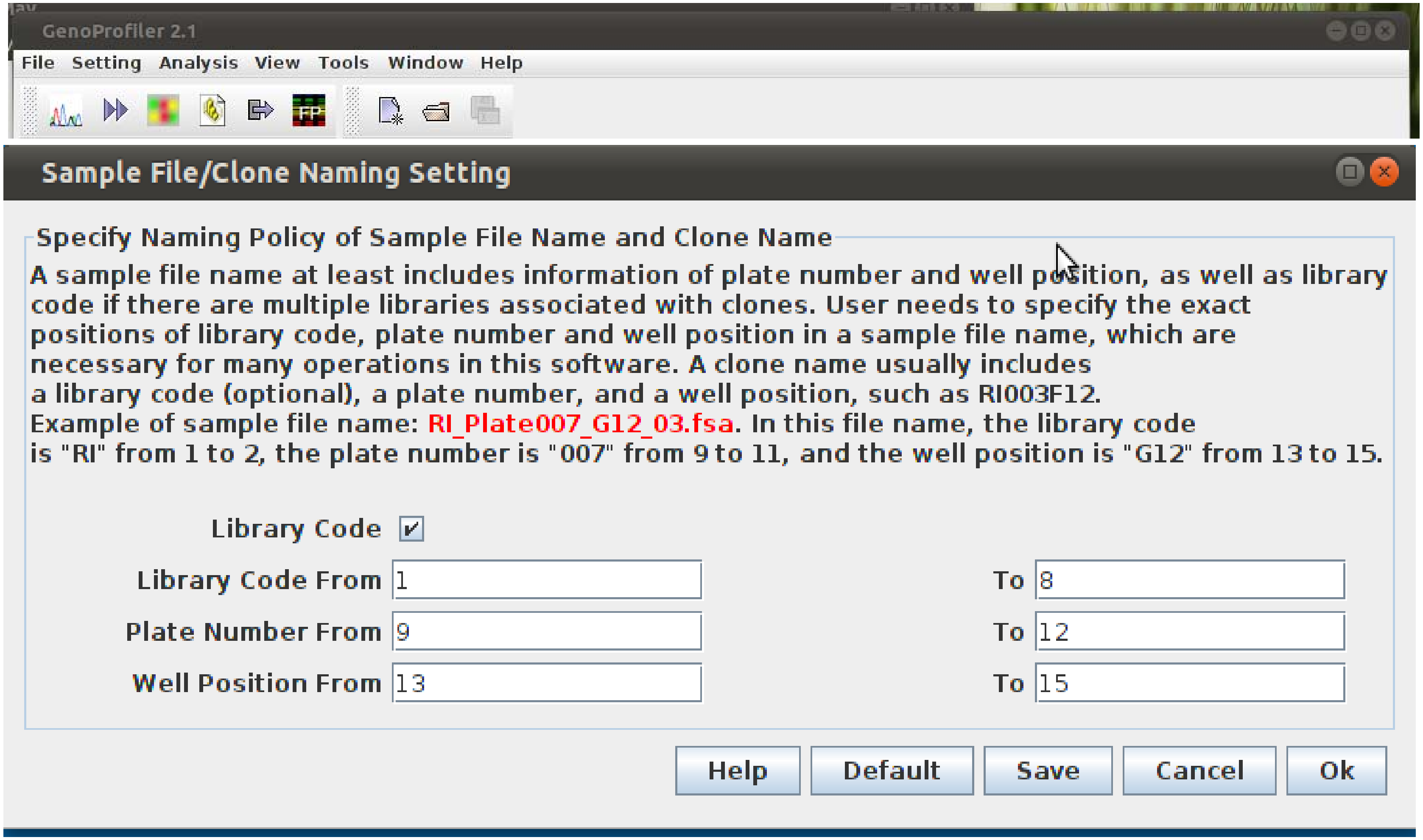
STEP 1: FPB Removal



FPB - FingerPrint Background removal

First Value:	3	Last Value:	7
Low index:	60	Min bands:	40
Min sizes (per color):	5	Max sizes (total):	400
Blue background:	10	Green background:	10
Yellow background:	10	Red background:	10
Blue offset:	0	Green offset:	15000
Yellow offset:	30000	Red offset:	45000
Tolerance:	0.4	Multiply factor:	30
Peak width:	15	Fixed threshold:	500
Size from:	20	Size to:	800
Library from:	1	Library to:	12
Plate from:	13	Plate to:	16
Grid from:	17	Table suffix:	txt

STEP 2: Size Editing Using GenoProfiler



Specify Naming Policy of Sample File Name and Clone Name

A sample file name at least includes information of plate number and well position, as well as library code if there are multiple libraries associated with clones. User needs to specify the exact positions of library code, plate number and well position in a sample file name, which are necessary for many operations in this software. A clone name usually includes a library code (optional), a plate number, and a well position, such as RI003F12.

Example of sample file name: RI_Plate007_G12_03.fsa. In this file name, the library code is "RI" from 1 to 2, the plate number is "007" from 9 to 11, and the well position is "G12" from 13 to 15.

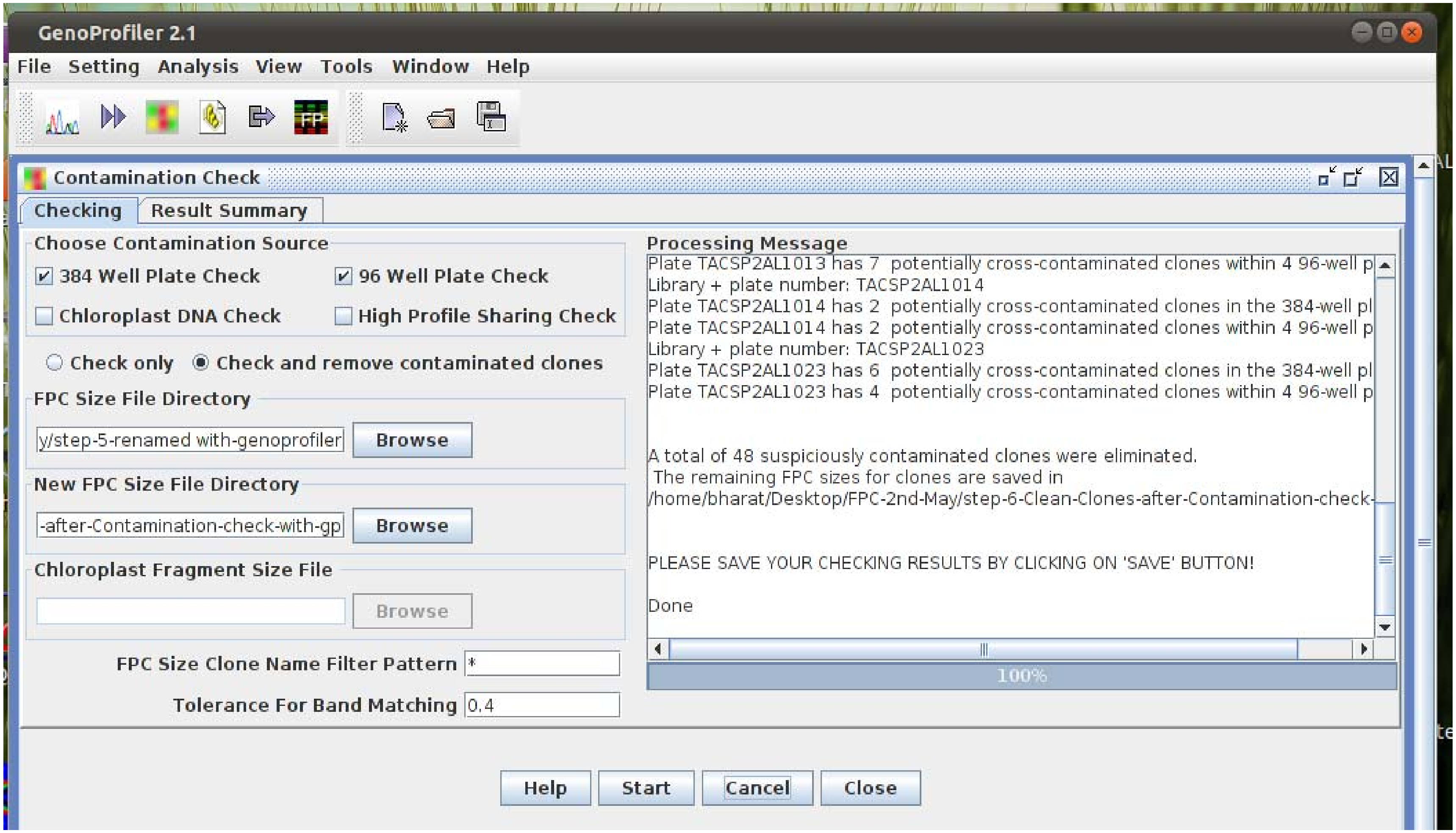
Library Code ☑

Library Code From 1 To 8

Plate Number From 9 To 12

Well Position From 13 To 15

STEP 3: Contamination Check



Choose Contamination Source

☑ 384 Well Plate Check ☑ 96 Well Plate Check

☐ Chloroplast DNA Check ☐ High Profile Sharing Check

○ Check only ● Check and remove contaminated clones

FPC Size File Directory: y/step-5-renamed with-genoprofiler

New FPC Size File Directory: -after-Contamination-check-with-gp

Chloroplast Fragment Size File:

FPC Size Clone Name Filter Pattern: *

Tolerance For Band Matching: 0.4

Processing Message

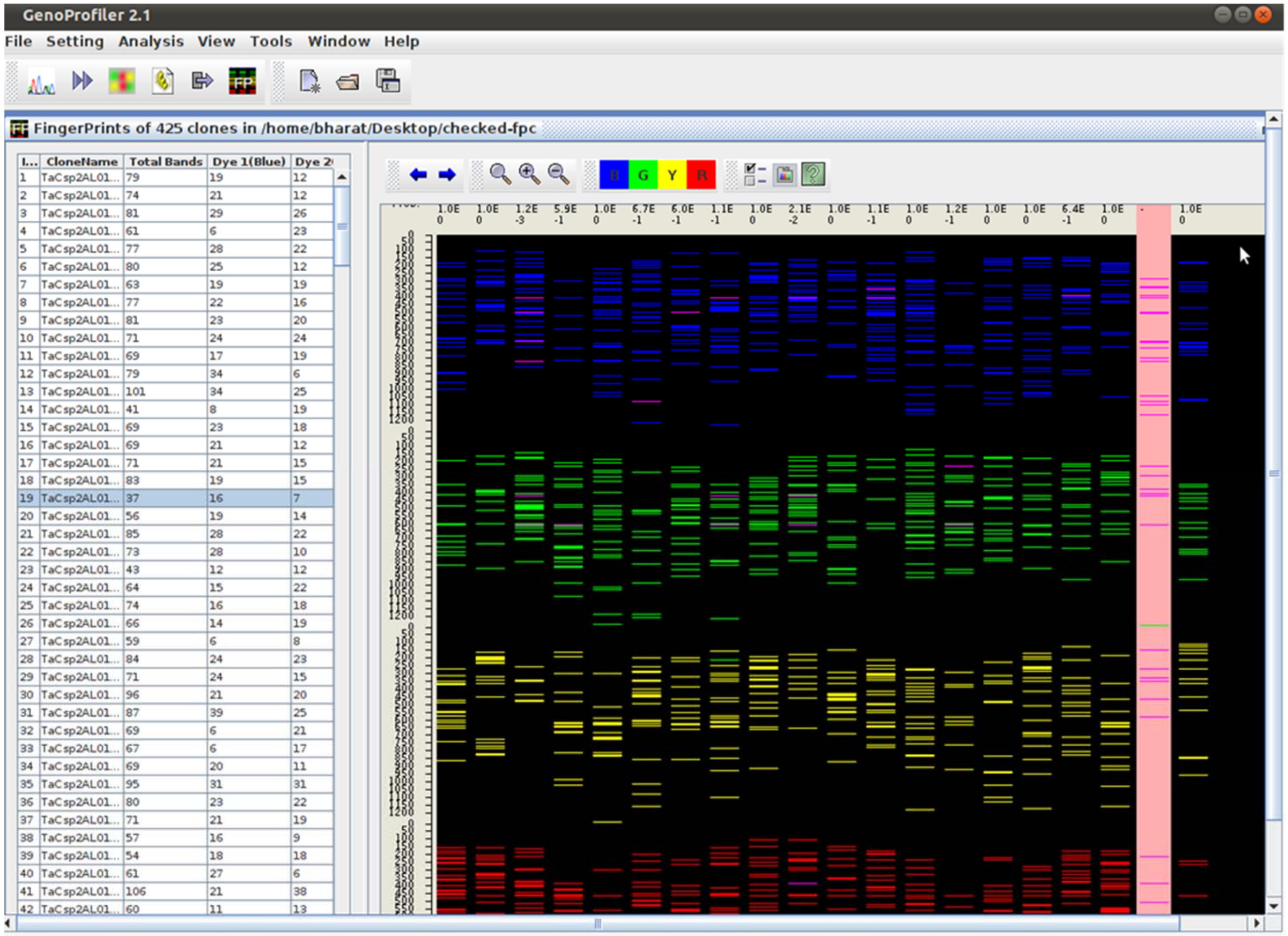
```
Plate TACSP2AL1013 has 7  potentially cross-contaminated clones within 4 96-well p
Library + plate number: TACSP2AL1014
Plate TACSP2AL1014 has 2  potentially cross-contaminated clones in the 384-well pl
Plate TACSP2AL1014 has 2  potentially cross-contaminated clones within 4 96-well p
Library + plate number: TACSP2AL1023
Plate TACSP2AL1023 has 6  potentially cross-contaminated clones in the 384-well pl
Plate TACSP2AL1023 has 4  potentially cross-contaminated clones within 4 96-well p


A total of 48 suspiciously contaminated clones were eliminated.
 The remaining FPC sizes for clones are saved in
/home/bharat/Desktop/FPC-2nd-May/step-6-Clean-Clones-after-Contamination-check-


PLEASE SAVE YOUR CHECKING RESULTS BY CLICKING ON 'SAVE' BUTTON!

Done
```

STEP 4: GenoProfiler Output

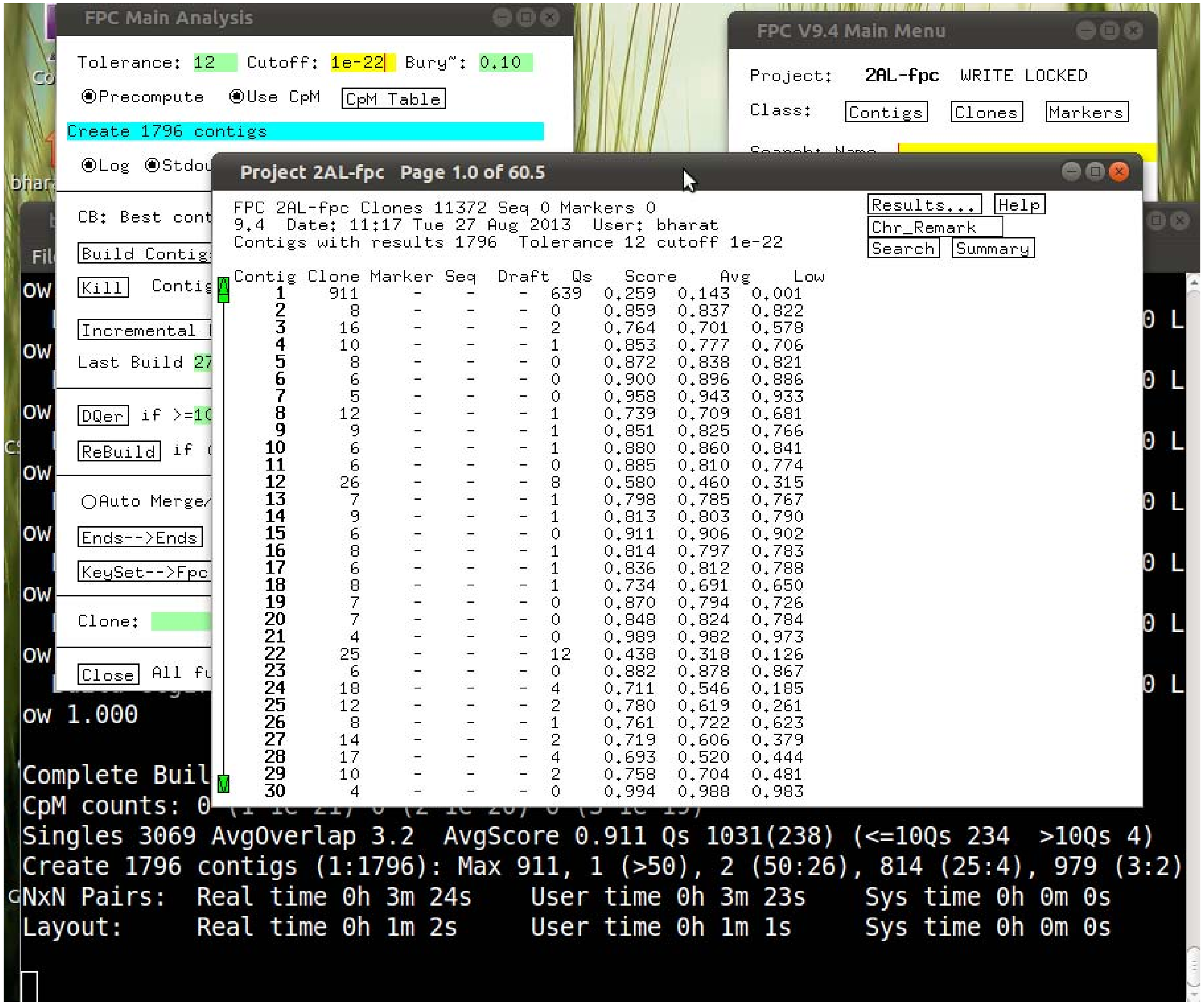


FingerPrints of 425 clones in /home/bharat/Desktop/checked-fpc

I...	CloneName	Total Bands	Dye 1(Blue)	Dye 2
1	TaCsp2AL01...	79	19	12
2	TaCsp2AL01...	74	21	12
3	TaCsp2AL01...	81	29	26
4	TaCsp2AL01...	61	6	23
5	TaCsp2AL01...	77	28	22
6	TaCsp2AL01...	80	25	12
7	TaCsp2AL01...	63	19	19
8	TaCsp2AL01...	77	22	16
9	TaCsp2AL01...	81	23	20
10	TaCsp2AL01...	71	24	24
11	TaCsp2AL01...	69	17	19
12	TaCsp2AL01...	79	34	6
13	TaCsp2AL01...	101	34	25
14	TaCsp2AL01...	41	8	19
15	TaCsp2AL01...	69	23	18
16	TaCsp2AL01...	69	21	12
17	TaCsp2AL01...	71	21	15
18	TaCsp2AL01...	83	19	15
19	TaCsp2AL01...	37	16	7
20	TaCsp2AL01...	56	19	14
21	TaCsp2AL01...	85	28	22
22	TaCsp2AL01...	73	28	10
23	TaCsp2AL01...	43	12	12
24	TaCsp2AL01...	64	15	22
25	TaCsp2AL01...	74	16	18
26	TaCsp2AL01...	66	14	19
27	TaCsp2AL01...	59	6	8
28	TaCsp2AL01...	84	24	23
29	TaCsp2AL01...	71	24	15
30	TaCsp2AL01...	96	21	20
31	TaCsp2AL01...	87	39	25
32	TaCsp2AL01...	69	6	21
33	TaCsp2AL01...	67	6	17
34	TaCsp2AL01...	69	20	11
35	TaCsp2AL01...	95	31	31
36	TaCsp2AL01...	80	23	22
37	TaCsp2AL01...	71	21	19
38	TaCsp2AL01...	57	16	9
39	TaCsp2AL01...	54	18	18
40	TaCsp2AL01...	61	27	6
41	TaCsp2AL01...	106	21	38
42	TaCsp2AL01...	60	11	13

11,012 clones out of 21,120 passed the criteria of FPB and GenoProfiler

FPC Step 1 - Partial view



Tolerance: 12 Cutoff: 1e-22 Bury~: 0.10

Create 1796 contigs

CB: Best cont

Last Build 27

if >=1

Clone:

Project: 2AL-fpc WRITE LOCKED

Class:

Project 2AL-fpc Page 1.0 of 60.5

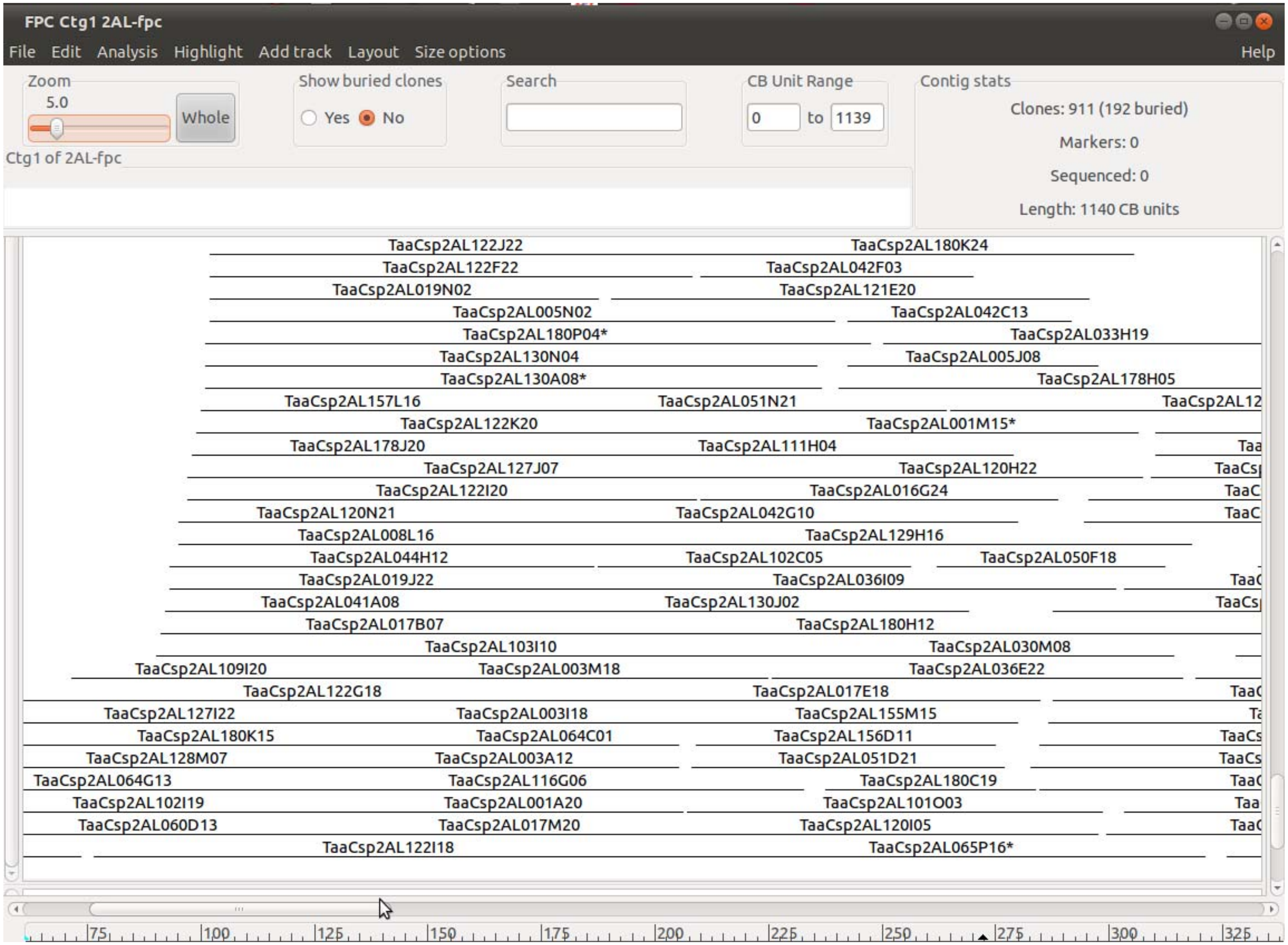
```
FPC 2AL-fpc Clones 11372 Seq 0 Markers 0
9.4  Date: 11:17 Tue 27 Aug 2013  User: bharat
Contigs with results 1796  Tolerance 12 cutoff 1e-22

Contig Clone Marker Seq  Draft  Qs   Score   Avg    Low
    1    911    -     -     -  639  0.259  0.143  0.001
    2      8    -     -     -  0    0.859  0.837  0.822
    3     16    -     -     -  2    0.764  0.701  0.578
    4     10    -     -     -  1    0.853  0.777  0.706
    5      8    -     -     -  0    0.872  0.838  0.821
    6      6    -     -     -  0    0.900  0.896  0.886
    7      5    -     -     -  0    0.958  0.943  0.933
    8     12    -     -     -  1    0.739  0.709  0.681
    9      9    -     -     -  1    0.851  0.825  0.766
   10      6    -     -     -  1    0.880  0.860  0.841
   11      6    -     -     -  0    0.885  0.810  0.774
   12     26    -     -     -  8    0.580  0.460  0.315
   13      7    -     -     -  1    0.798  0.785  0.767
   14      9    -     -     -  1    0.813  0.803  0.790
   15      6    -     -     -  0    0.911  0.906  0.902
   16      8    -     -     -  1    0.814  0.797  0.783
   17      6    -     -     -  1    0.836  0.812  0.788
   18      8    -     -     -  1    0.734  0.691  0.650
   19      7    -     -     -  0    0.870  0.794  0.726
   20      7    -     -     -  0    0.848  0.824  0.784
   21      4    -     -     -  0    0.989  0.982  0.973
   22     25    -     -     -  12   0.438  0.318  0.126
   23      6    -     -     -  0    0.882  0.878  0.867
   24     18    -     -     -  4    0.711  0.546  0.185
   25     12    -     -     -  2    0.780  0.619  0.261
   26      8    -     -     -  1    0.761  0.722  0.623
   27     14    -     -     -  2    0.719  0.606  0.379
   28     17    -     -     -  4    0.693  0.520  0.444
   29     10    -     -     -  2    0.758  0.704  0.481
   30      4    -     -     -  0    0.994  0.988  0.983
```

```
ow 1.000

Complete Buil
CpM counts: 0
Singles 3069 AvgOverlap 3.2  AvgScore 0.911 Qs 1031(238) (<=10Qs 234  >10Qs 4)
Create 1796 contigs (1:1796): Max 911, 1 (>50), 2 (50:26), 814 (25:4), 979 (3:2)
NxN Pairs:  Real time 0h 3m 24s    User time 0h 3m 23s    Sys time 0h 0m 0s
Layout:     Real time 0h 1m 2s     User time 0h 1m 1s     Sys time 0h 0m 0s
```

FPC Step 2 - Partial view



FPC Ctg1 2AL-fpc

Zoom 5.0

Show buried clones: ○ Yes ● No

Search:

CB Unit Range: 0 to 1139

Ctg1 of 2AL-fpc

Contig stats

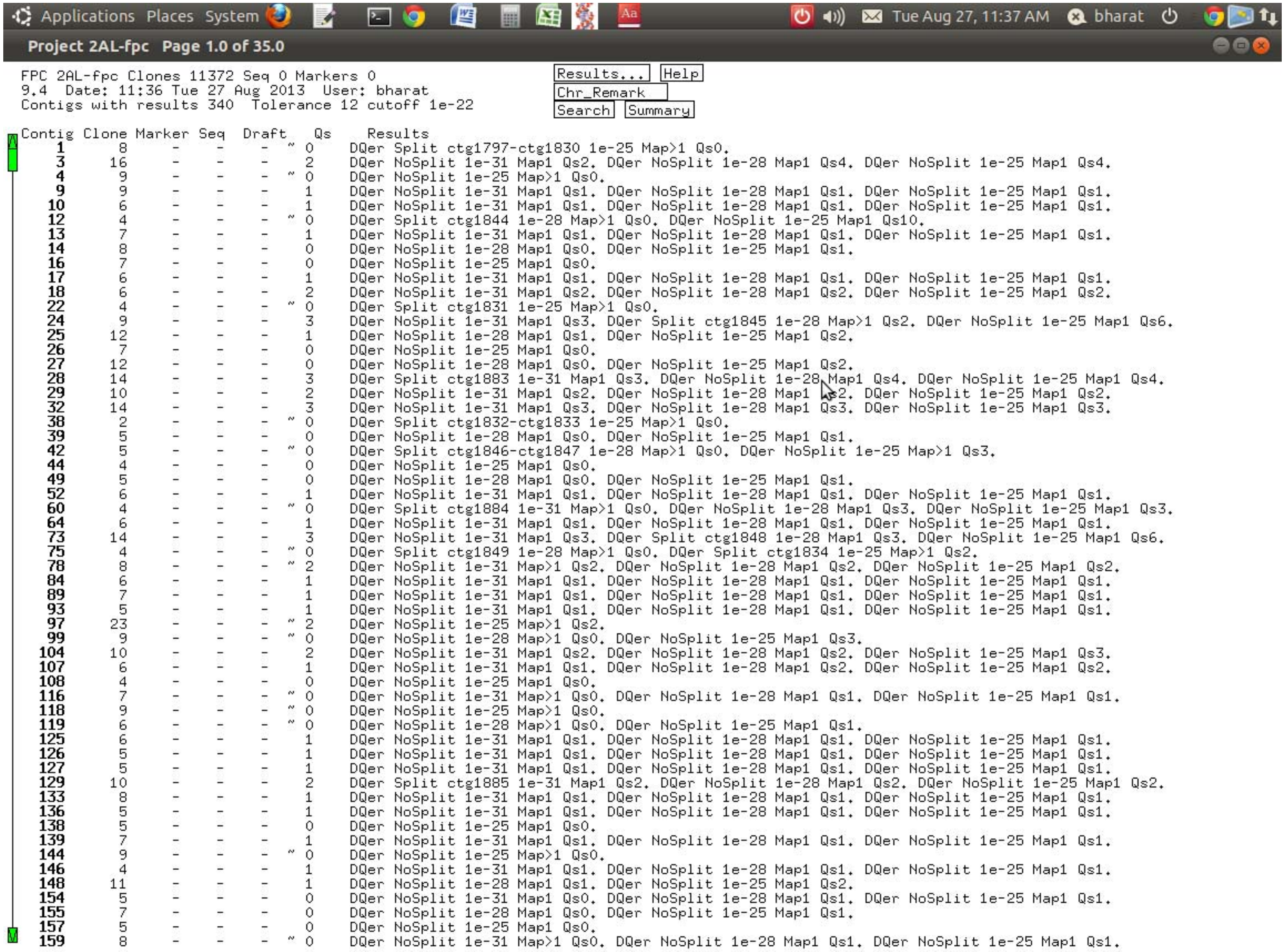
Clones: 911 (192 buried)

Markers: 0

Sequenced: 0

Length: 1140 CB units

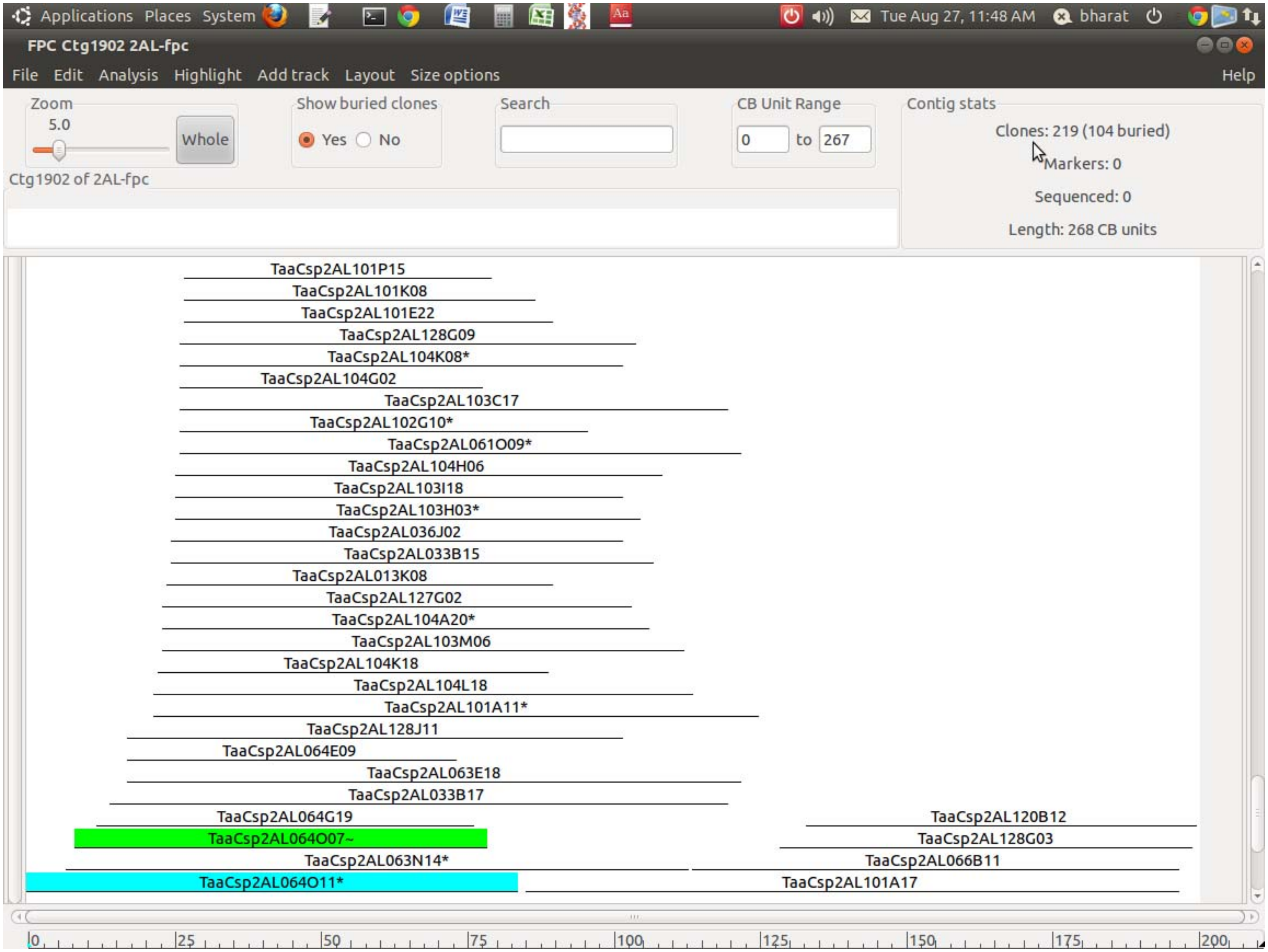
FPC Step 3 - Partial view



```
FPC 2AL-fpc Clones 11372 Seq 0 Markers 0
9.4  Date: 11:36 Tue 27 Aug 2013  User: bharat
Contigs with results 340  Tolerance 12 cutoff 1e-22

Contig Clone Marker Seq  Draft   Qs     Results
     1      8     -    -     -  ~ 0    DQer Split ctg1797-ctg1830 1e-25 Map>1 Qs0.
     3     16     -    -     -    2    DQer NoSplit 1e-31 Map1 Qs2. DQer NoSplit 1e-28 Map1 Qs4. DQer NoSplit 1e-25 Map1 Qs4.
     4      9     -    -     -  ~ 0    DQer NoSplit 1e-25 Map>1 Qs0.
     9      9     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    10      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    12      4     -    -     -  ~ 0    DQer Split ctg1844 1e-28 Map>1 Qs0. DQer NoSplit 1e-25 Map1 Qs10.
    13      7     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    14      8     -    -     -    0    DQer NoSplit 1e-28 Map1 Qs0. DQer NoSplit 1e-25 Map1 Qs1.
    16      7     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
    17      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    18      6     -    -     -    2    DQer NoSplit 1e-31 Map1 Qs2. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs2.
    22      4     -    -     -  ~ 0    DQer Split ctg1831 1e-25 Map>1 Qs0.
    24      9     -    -     -    3    DQer NoSplit 1e-31 Map1 Qs3. DQer Split ctg1845 1e-28 Map>1 Qs2. DQer NoSplit 1e-25 Map1 Qs6.
    25     12     -    -     -    1    DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs2.
    26      7     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
    27     12     -    -     -    0    DQer NoSplit 1e-28 Map1 Qs0. DQer NoSplit 1e-25 Map1 Qs2.
    28     14     -    -     -    3    DQer Split ctg1883 1e-31 Map1 Qs3. DQer NoSplit 1e-28 Map1 Qs4. DQer NoSplit 1e-25 Map1 Qs4.
    29     10     -    -     -    2    DQer NoSplit 1e-31 Map1 Qs2. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs2.
    32     14     -    -     -    3    DQer NoSplit 1e-31 Map1 Qs3. DQer NoSplit 1e-28 Map1 Qs3. DQer NoSplit 1e-25 Map1 Qs3.
    38      2     -    -     -  ~ 0    DQer Split ctg1832-ctg1833 1e-25 Map>1 Qs0.
    39      5     -    -     -    0    DQer NoSplit 1e-28 Map1 Qs0. DQer NoSplit 1e-25 Map1 Qs1.
    42      5     -    -     -  ~ 0    DQer Split ctg1846-ctg1847 1e-28 Map>1 Qs0. DQer NoSplit 1e-25 Map>1 Qs3.
    44      4     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
    49      5     -    -     -    0    DQer NoSplit 1e-28 Map1 Qs0. DQer NoSplit 1e-25 Map1 Qs1.
    52      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    60      4     -    -     -  ~ 0    DQer Split ctg1884 1e-31 Map>1 Qs0. DQer NoSplit 1e-28 Map1 Qs3. DQer NoSplit 1e-25 Map1 Qs3.
    64      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    73     14     -    -     -    3    DQer NoSplit 1e-31 Map1 Qs3. DQer Split ctg1848 1e-28 Map1 Qs3. DQer NoSplit 1e-25 Map1 Qs6.
    75      4     -    -     -  ~ 0    DQer Split ctg1849 1e-28 Map>1 Qs0. DQer Split ctg1834 1e-25 Map>1 Qs2.
    78      8     -    -     -  ~ 2    DQer NoSplit 1e-31 Map>1 Qs2. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs2.
    84      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    89      7     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    93      5     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
    97     23     -    -     -  ~ 2    DQer NoSplit 1e-25 Map>1 Qs2.
    99      9     -    -     -  ~ 0    DQer NoSplit 1e-28 Map>1 Qs0. DQer NoSplit 1e-25 Map1 Qs3.
   104     10     -    -     -    2    DQer NoSplit 1e-31 Map1 Qs2. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs3.
   107      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs2.
   108      4     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
   116      7     -    -     -  ~ 0    DQer NoSplit 1e-31 Map>1 Qs0. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   118      9     -    -     -  ~ 0    DQer NoSplit 1e-25 Map>1 Qs0.
   119      6     -    -     -  ~ 0    DQer NoSplit 1e-28 Map>1 Qs0. DQer NoSplit 1e-25 Map1 Qs1.
   125      6     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   126      5     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   127      5     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   129     10     -    -     -    2    DQer Split ctg1885 1e-31 Map1 Qs2. DQer NoSplit 1e-28 Map1 Qs2. DQer NoSplit 1e-25 Map1 Qs2.
   133      8     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   136      5     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   138      5     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
   139      7     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   144      9     -    -     -  ~ 0    DQer NoSplit 1e-25 Map>1 Qs0.
   146      4     -    -     -    1    DQer NoSplit 1e-31 Map1 Qs1. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   148     11     -    -     -    1    DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs2.
   154      5     -    -     -    0    DQer NoSplit 1e-31 Map1 Qs0. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
   155      7     -    -     -    0    DQer NoSplit 1e-28 Map1 Qs0. DQer NoSplit 1e-25 Map1 Qs1.
   157      5     -    -     -    0    DQer NoSplit 1e-25 Map1 Qs0.
   159      8     -    -     -  ~ 0    DQer NoSplit 1e-31 Map>1 Qs0. DQer NoSplit 1e-28 Map1 Qs1. DQer NoSplit 1e-25 Map1 Qs1.
```

FPC Step 4 - Partial view



FPC Ctg1902 2AL-fpc

Zoom: 5.0

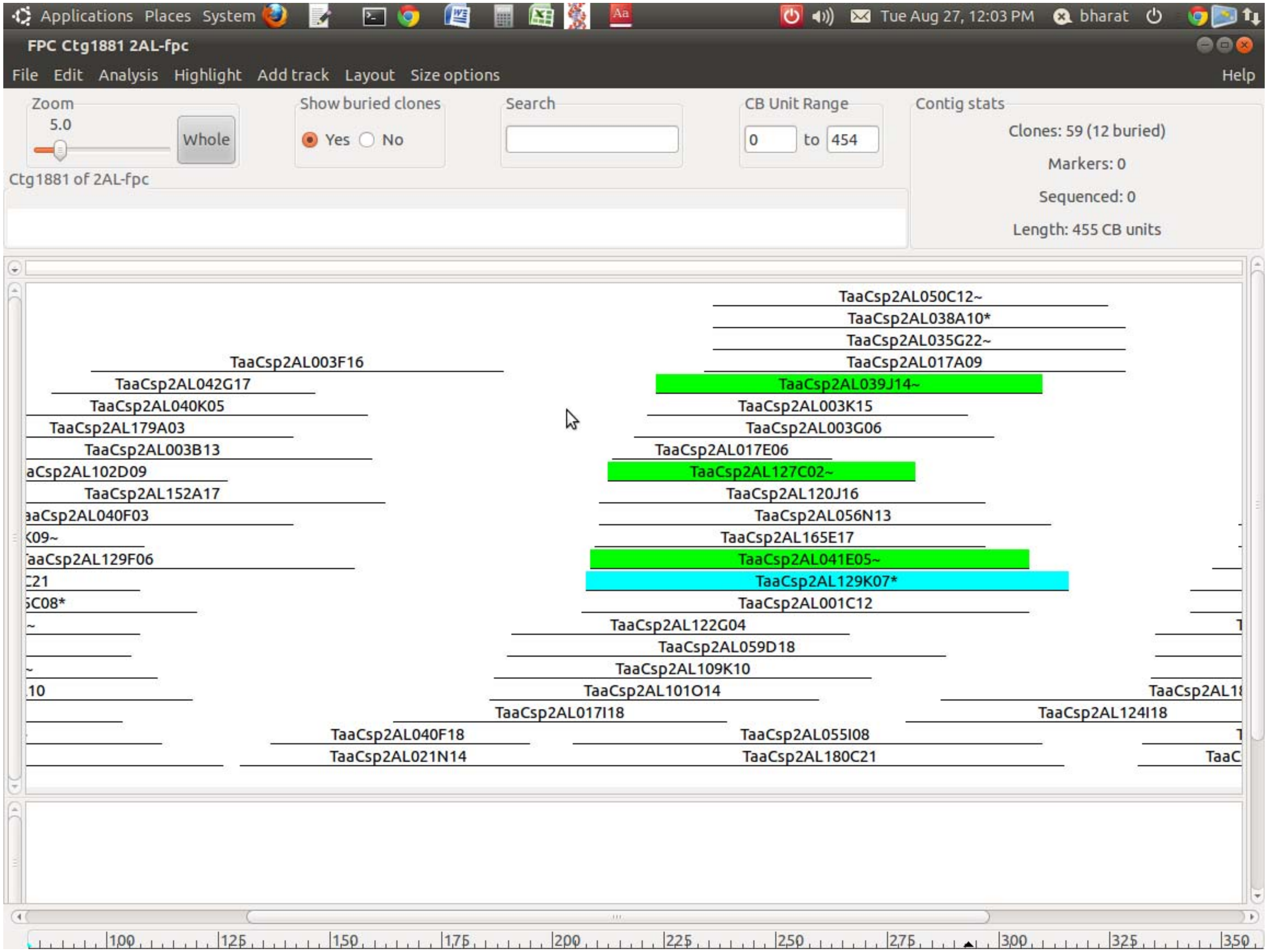
Show buried clones: ● Yes ○ No

CB Unit Range: 0 to 267

Ctg1902 of 2AL-fpc

Contig stats
Clones: 219 (104 buried)
Markers: 0
Sequenced: 0
Length: 268 CB units

FPC Step 5 - Partial view



FPC Ctg1881 2AL-fpc

Zoom 5.0

Show buried clones ● Yes ○ No

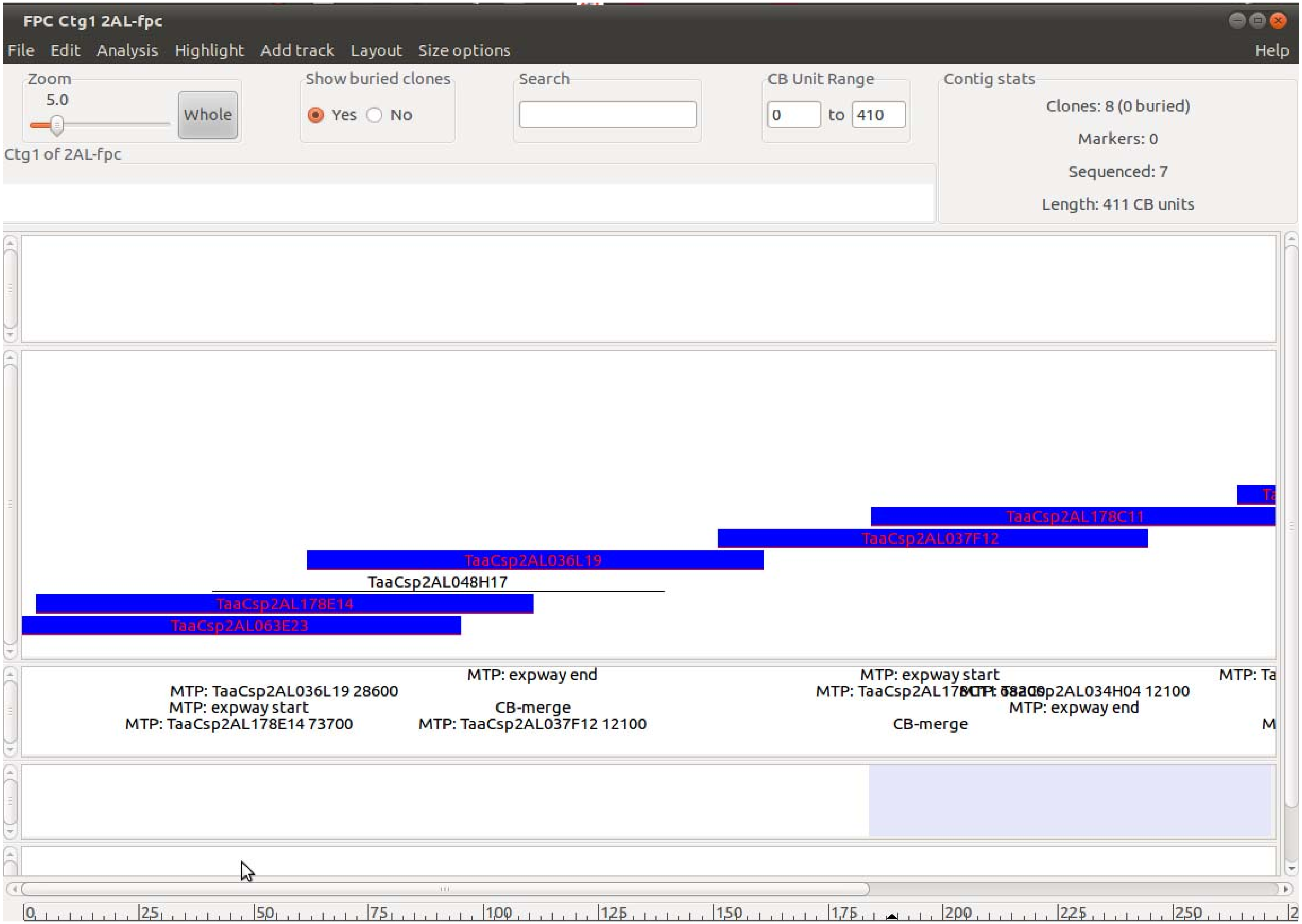
Search

CB Unit Range 0 to 454

Ctg1881 of 2AL-fpc

Contig stats
Clones: 59 (12 buried)
Markers: 0
Sequenced: 0
Length: 455 CB units

BAC Clones showing Minimum Tiling Path (MTP)



Zoom: 5.0

Show buried clones: ● Yes ○ No

CB Unit Range: 0 to 410

Contig stats

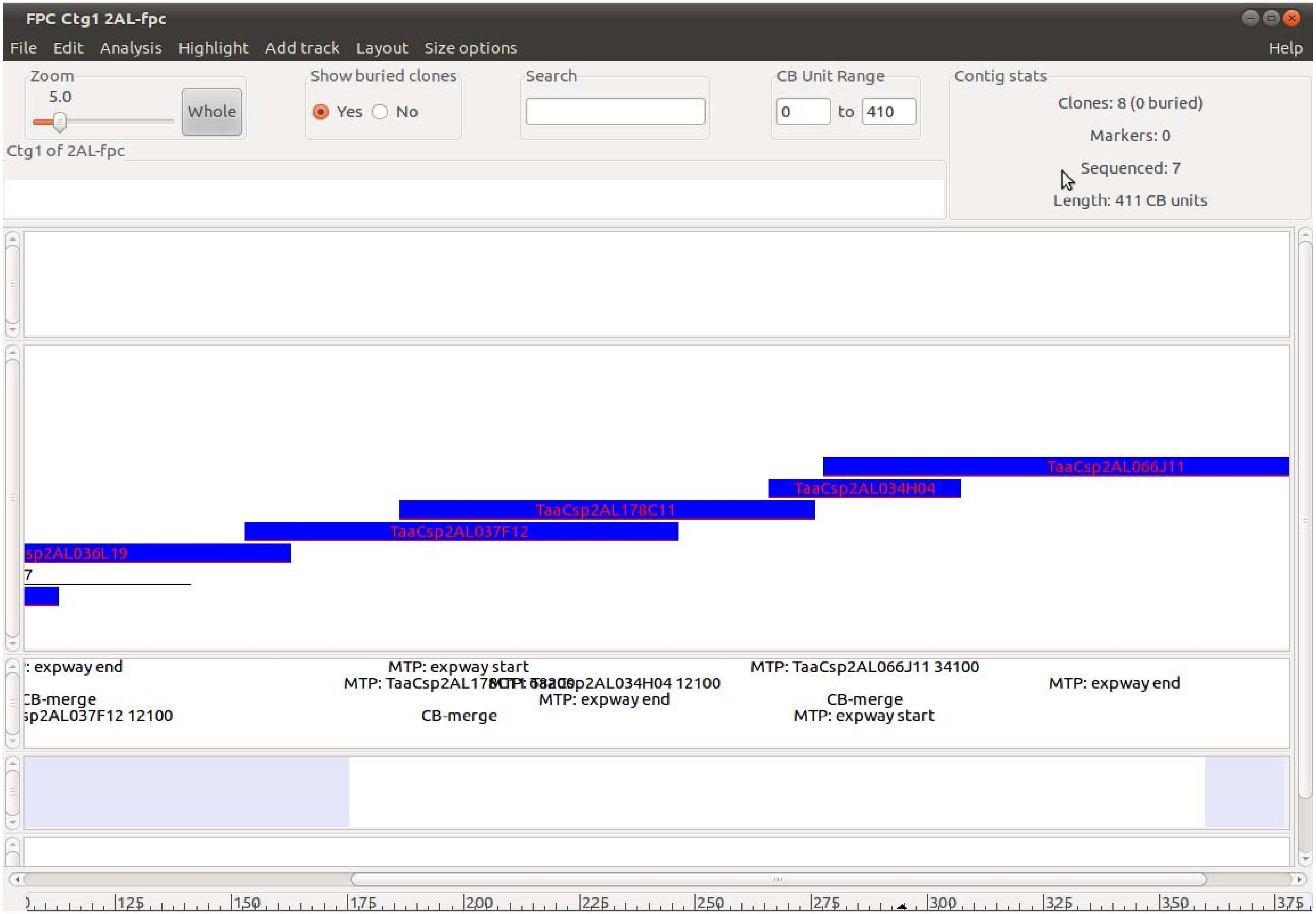
Clones: 8 (0 buried)

Markers: 0

Sequenced: 7

Length: 411 CB units

BAC Clones showing Minimum Tiling Path (MTP)



Zoom: 5.0

Show buried clones: ● Yes ○ No

CB Unit Range: 0 to 410

Ctg1 of 2AL-fpc

Contig stats

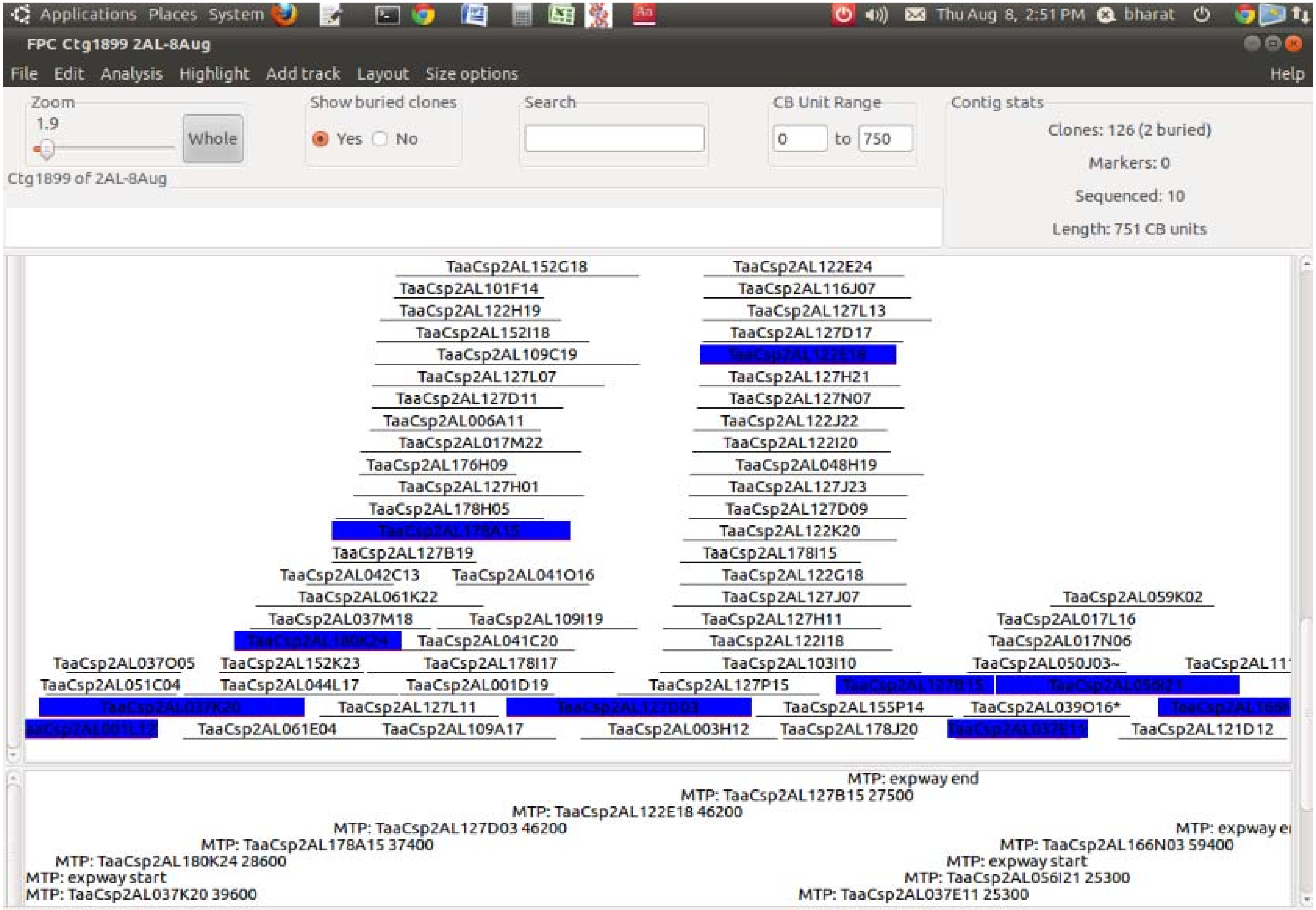
Clones: 8 (0 buried)

Markers: 0

Sequenced: 7

Length: 411 CB units

BAC Clones showing Minimum Tiling Path (MTP)



Zoom 1.9

Show buried clones ● Yes ○ No

CB Unit Range 0 to 750

Contig stats

Clones: 126 (2 buried)

Markers: 0

Sequenced: 10

Length: 751 CB units

```
                                        MTP: expway end
                               MTP: TaaCsp2AL127B15 27500
                     MTP: TaaCsp2AL122E18 46200
           MTP: TaaCsp2AL127D03 46200                                        MTP: expway e
   MTP: TaaCsp2AL178A15 37400                                     MTP: TaaCsp2AL166N03 59400
MTP: TaaCsp2AL180K24 28600                                    MTP: expway start
MTP: expway start                                          MTP: TaaCsp2AL056I21 25300
MTP: TaaCsp2AL037K20 39600                           MTP: TaaCsp2AL037E11 25300
```

Number of contigs assembled at different cut off values using FPC V9.4

Cutoff	Contigs	Singletons	Number of clones in biggest contig
1.0e-75	678	9542	152
1.0e-70	833	9108	164
1.0e-65	1003	8591	176
1.0e-60	1205	7982	180
1.0e-55	1402	7350	186
1.0e-50	1578	6660	195
1.0e-45	1739	5927	195
1.0e-40	1843	5277	201
1.0e-35	1923	4642	202
1.0e-30	1923	3997	346
1.0e-25	1877	3390	692
1.0e-22	1796	3069	911

Current status of SNaPshot Sequencing data -2AS

S. No.	Details	Value
1	Total No of Plates (96 well) fingerprinted	75 x 4 (300 plates)
2	Total No. of plates analyzed from GeneMapper	91 (8554 BAC clones)
3	Total No of BAC clones passed from Genemapper	7678 (87.6 % of analyzed BAC Clones)
4	Total No of analyzed BAC Clones filtered by FPB	5219 (61 % of GeneMapper Passed BAC Clones)

BAC-End Sequencing

BAC libraries in 384-well format for *Triticum aestivum* cv. Chinese Spring chromosome 2A, received in September 2011 from the Czech Republic

Chromosome 2AL:	**200 plates**
Chromosome 2AS:	**148 plates**
Total plates:	**348**
Total BACs:	**1,33,632**
Average insert size:	**120 kb**
Vector:	**pIndigoBAC-5**
Chromosome coverage:	**15X**

Work flow for BAC ends Sequencing

A 384-well BAC plate

↓

Four 96-well plate glycerol stocks

↓

Primary culture

↓

Secondary culture

↓

BAC DNA isolation

↓

Quality and quantity estimation on agarose gel

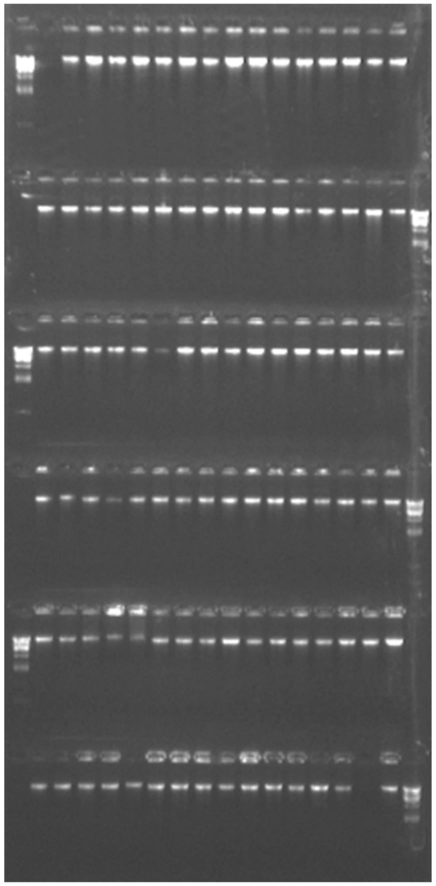
↓

Sequencing using ABI3730xl Sequencers

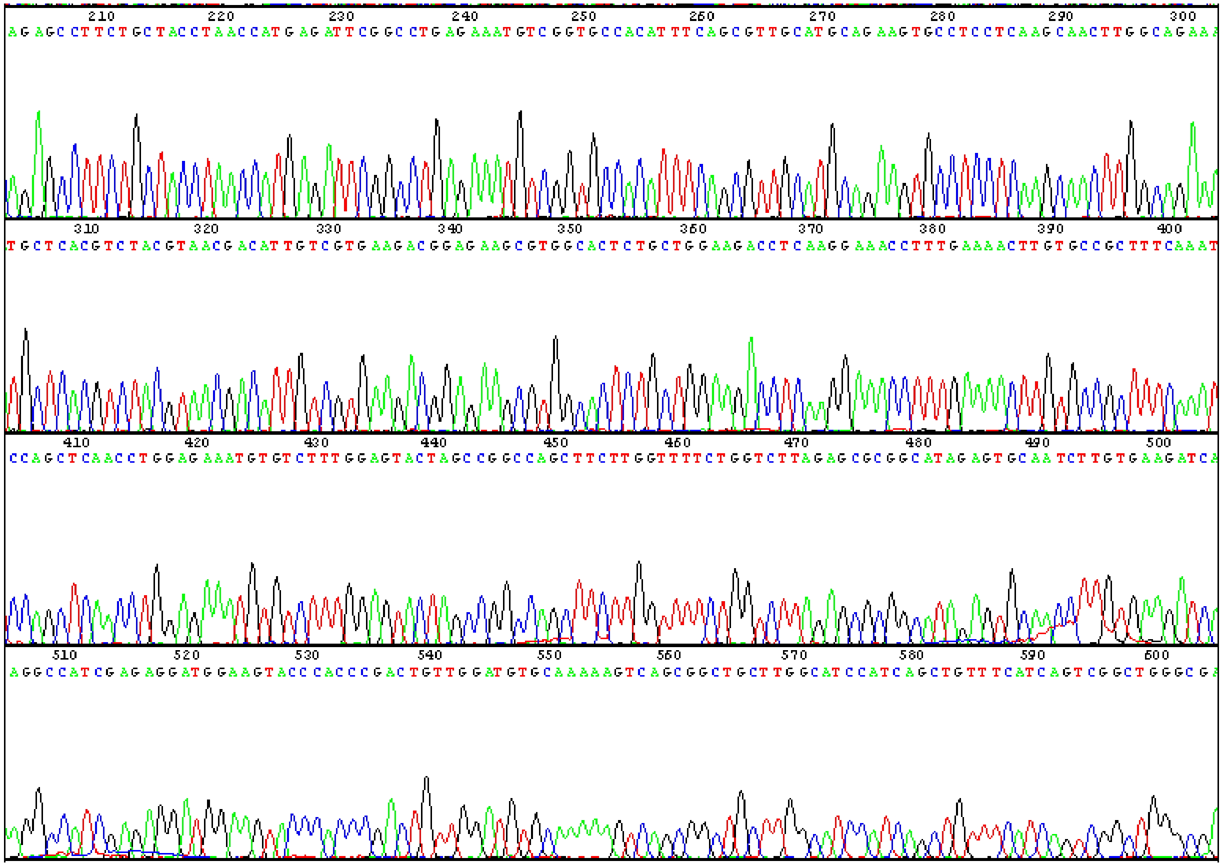
↓

Quality and vector trimming of sequences

BAC Quality Check



BAC DNA from a representative plate



Chromatogram of a representative BAC

Progress *(August, 2013)*

Glycerol Stock	(96 Well Plate) 2AL = 520, 2AS = 196
DNA Isolation	716
Sequenced Plates	716 X 2 (F/R)
No of BAC Clones Sequenced	67,304 2AL = 48880, 2AS = 18424
Total Data Generated	74Mb
Average Sequence length	550bp

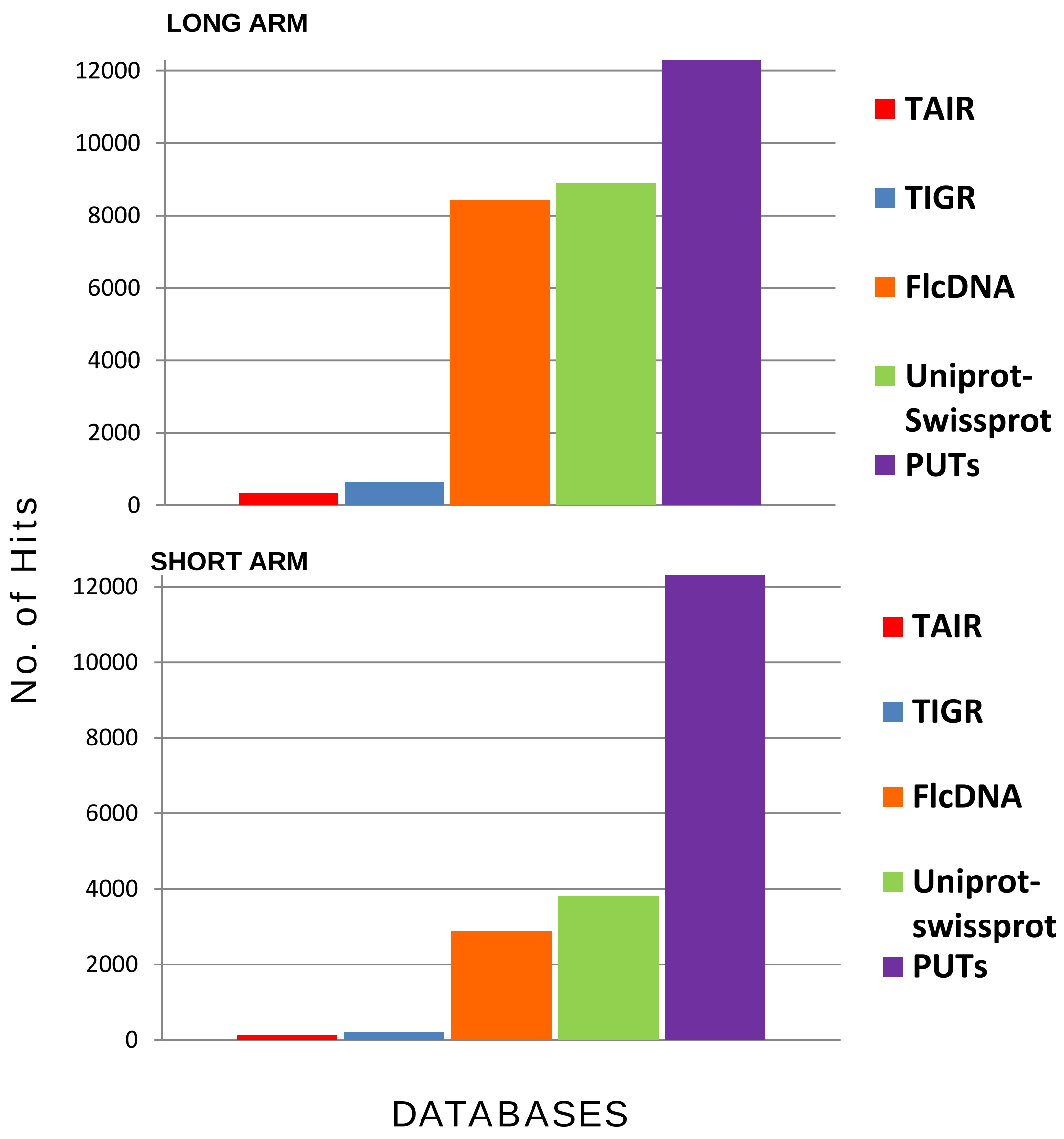
BAC End Sequence Analysis

BAC end sequence obtained after sequencing

BLAST was done with different database

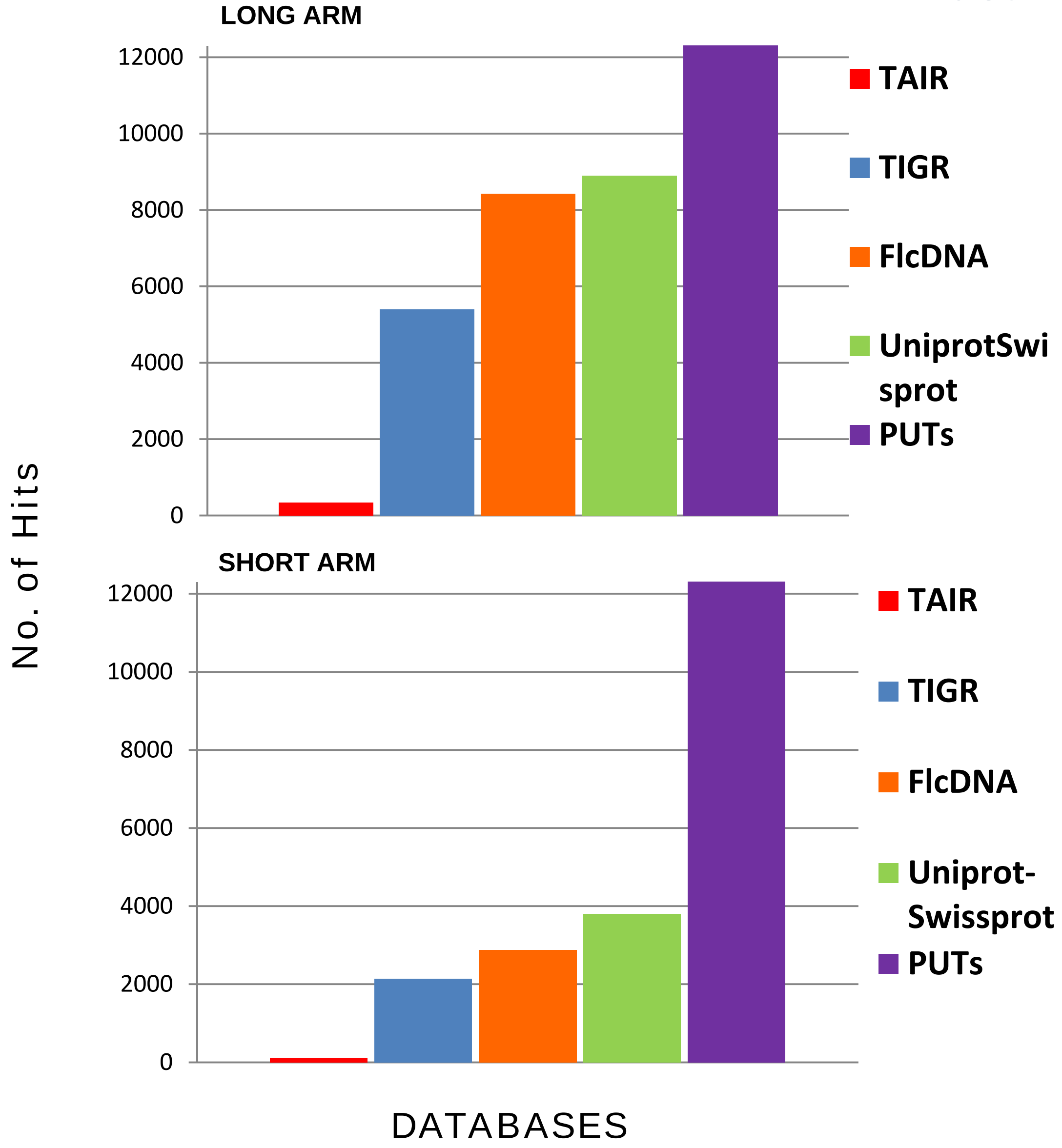
Blastx (e ≤ -10) ,Blastn (e ≤ -10 to e ≤ -50) ,
Bit score ≥ 150

Blast Result – e^{-35}



BAC End Sequence Analysis

Blast Result – e^{-10}



BAC End Sequence Analysis Result – 2AL

S. No.	BAC clone	Name of the gene	Organism
1	TaaCsp2ALhA0126-A04	somatic embryogenesis receptor-like kinase 1	*A. thaliana*
2	TaaCsp2ALhA0080-M02	Flowering-promoting factor 1-like protein	*A. thaliana*
3	TaaCsp2ALhA0068-I22	WHEAT ATP synthase subunit beta	*O. sativa*
4	TaaCsp2ALhA00132-020	Homeobox-leucine zipper protein	*O. sativa*
5	TaaCsp2ALhA0102-F16	Probable WRKY transcription factor	*A. thaliana*
6	TaaCsp2ALhA0075-A08	DEAD-box ATP-dependent RNA helicase	*O. sativa*
7	TaaCsp2ALhA0046-I05	OsFBX150-F-Box domain containing protein	*O. sativa*
8	TaaCsp2ALhA0019-F04	NAD(P)H-quinone oxidoreductase subunit	*T. aestivum*
9	TaaCsp2ALhA0068-L03	Glutathione S-transferase	*A. thaliana*
10	TaaCsp2ALhA0013-C06	Fructokinase-1	*O. sativa*

BAC End Sequence Analysis Result – 2AL

S. No	BAC clone	Name of the gene	Organism
11	TaaCsp2ALhA0087-P20	Scarecrow-like protein 32	*A. thaliana*
12	TaaCsp2ALhA0001-H24	suppressor of stem-loop protein 1	*O. sativa*
13	TaaCsp2ALhA0005-E8	Na+ transporter	*O. sativa*
14	TaaCsp2ALhA0025-B5	HEAT repeat family protein	*O. sativa*
15	TaaCsp2ALhA0002-J15	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1	*O. sativa*
16	TaaCsp2ALhA0052-G3	Phosphatidyl inositol-4-phosphate 5-kinase	*O. sativa*
17	TaaCsp2ALhA0077-F03	sucrose-phosphate synthase	*O. sativa*
18	TaaCsp2ALhA0091-K14	aquaporin protein	*O. sativa*
19	TaaCsp2ALhA0050-E16	Histone-lysine N-methyltransferae	*A. thaliana*
20	TaaCsp2ALhA0027-E9	Aromatic-L-amino-acid decarboxylase	*C. roseus*

BAC End Sequence Analysis Result – 2AS

S. No.	BAC clone	Name of the gene	Organism
1	TaaCsp2AShA0025-P07	photosystem I P700 chlorophyll a apoproteinA1	*O. sativa*
2	TaaCsp2AShA0011-C10	starch synthase	*O. sativa*
3	TaaCsp2AShA0033-J02	DEAD-box ATP-dependent RNA helicase	*O. sativa*
4	TaaCsp2AShA0036-D21	zinc finger, C3HC4 type domain containingprotein	*O. sativa*
5	TaaCsp2AShA0049-G21	lectin-like receptor kinase	*O. sativa*
6	TaaCsp2AShA0049-F01	SHOOT1 protein	*O. sativa*
7	TaaCsp2AShA0004-M03	patatin-like phospholipase family protein	*O. sativa*
8	TaaCsp2AShA0041-D03	ATPA_WHEAT ATP synthase subunit alpha	*T. aestivum*
9	TaaCsp2AShA0038-L17	HSP82_MAIZE Heat shock protein	*Zea. mays*
10	TaaCsp2AShA0033-C08	WHEAT RuBisCO large subunit-binding protein	*T. aestivum*

BAC End Sequence Analysis – 2AS

S. No	BAC clone	Name of the gene	Organism
11	TaaCsp2AShA0003-L19	HSF-type DNA-binding domain containing protein	*O. sativa*
12	TaaCsp2AShA0038-A11	OsFBX362 - F-box domain containing protein	*O. sativa*
13	TaaCsp2AShA0043-K16	WEB1_ARATH Protein WEAK CHLOROPLAST MOVEMENT UNDER BLUELIGHT 1	*A. thaliana*
14	TaaCsp2AShA0015-F04	SAUR-like auxin-responsive protein family	*A. thaliana*
15	TaaCsp2AShA0039-N21	apocytochrome b	*A. thaliana*
16	TaaCsp2AShA0031-B09	Chlorophyll a-b binding protein	*O. sativa*
17	TaaCsp2AShA0032-J04	glycerol-3-phosphate transporter	*Z. mays*
18	TaaCsp2AShA0004-M04	anthranilate phosphoribosyltransferase	*O. sativa*
19	TaaCsp2AShA0005-G15	SAC3/GANP family protein	*O. sativa*
20	TaaCsp2AShA0021-I10	Cytochrome P450 protein	*O. sativa*

MAPPING POPULATIONS

Cross	Generation	Population size
T. boeoticum (pau5088) X *T. monococcum* (pau14087)	**F_{12}**	**160**
	F_5	**60**
	F_4	**105**
	F_6 (KSU)	**1200**
T. dicoccoides (pau4632) X Aconchi-89	**F_3**	**1461**
T. dicoccoides (pau4632) X PBW 114	**F_3**	**1607**
T. dicoccoides (pau4668) X Aconchi-89	**F_3**	**405**
T. dicoccoides (pau4668) X PBW 114	**F_3**	**849**
PBW 114/*Ae. Tauschii* (amphiploid) X PBW 621 (pau 14328)	**F_4**	**627**
PBW 114/*Ae. Tauschii* (amphiploid) X HD2967	**F_4**	**464**

Field view of the RIL population derived from the cross
T. boeoticum / T. monococcum

MARKER DEVELOPMENT AND MAPPING

SSRs were identified by using MISA

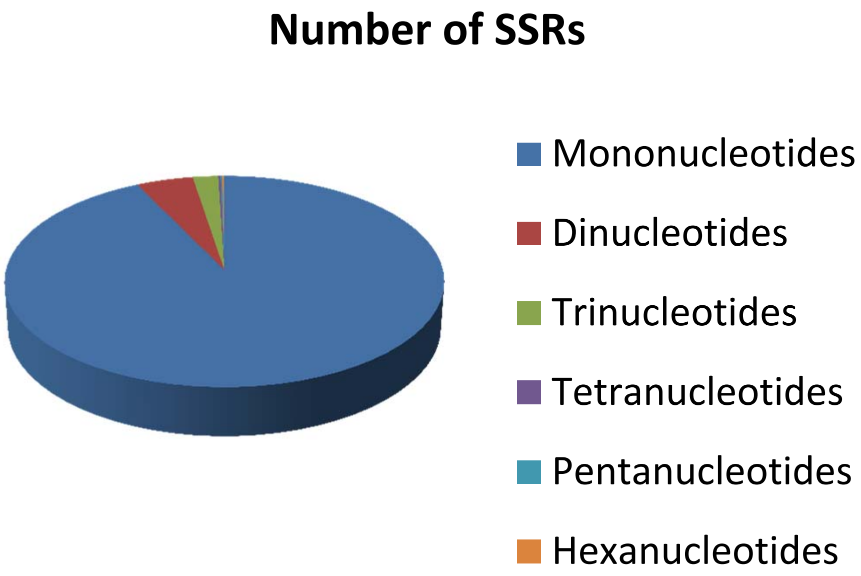
Definition of Microsatellite Marker (unit size/ minimum number of repeats) - (1/10) (2/10) (3/7) (4/6) (5/6) (6/6)

MISA Output

Total number of contigs examined:	**425821**
Total number of identified SSRs:	**22460**
Number of contigs containing SSRs:	**18376**
Number of SSRs present in compound formation:	**1859**

SSR Marker Development

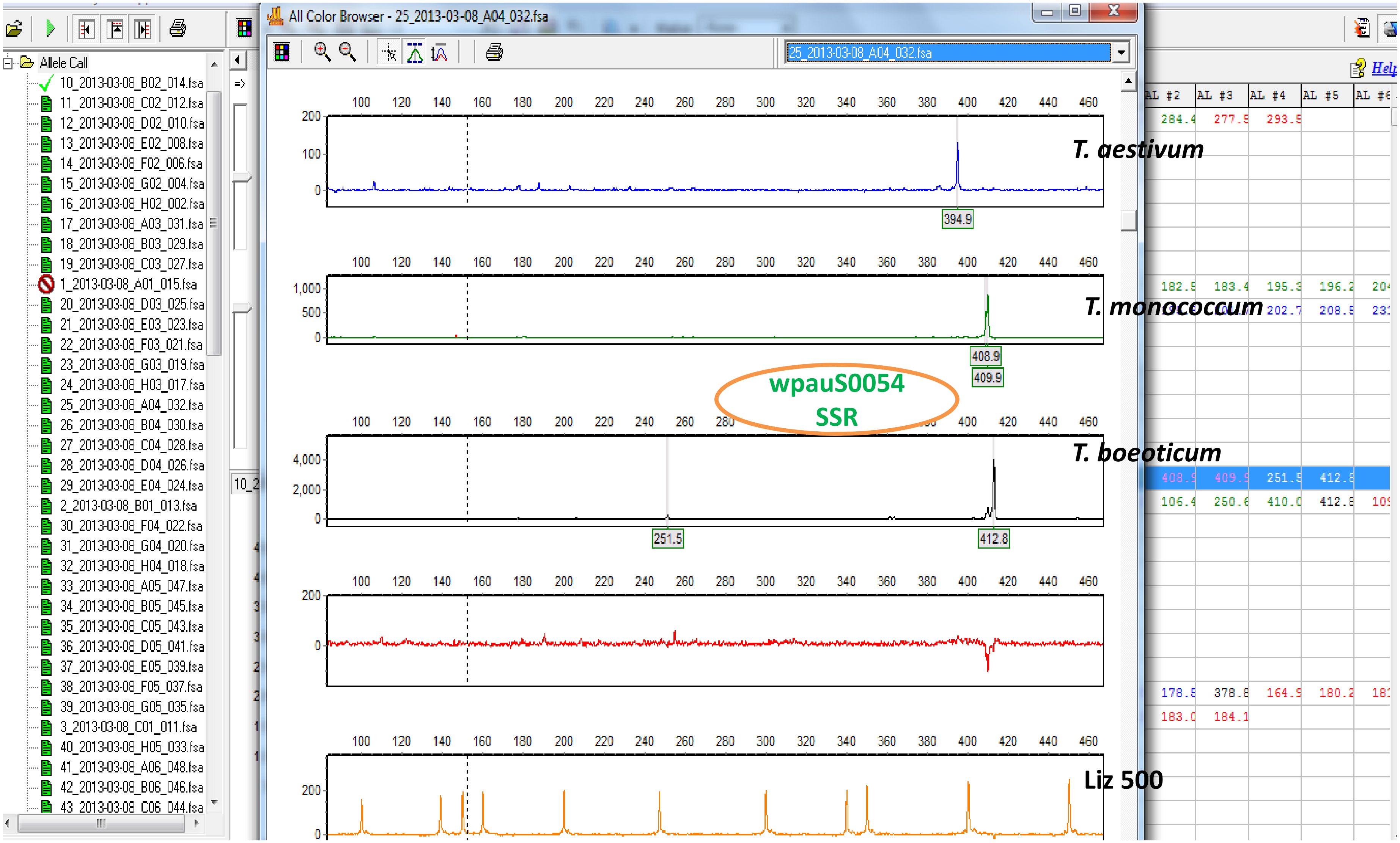
SSR prediction on 18376 sequences containing SSRs with following criteria - (1/10) (2/15) (3/10)(4/10) (5/6) (6/6)



Mononucleotides – 16918
Dinucleotides – 849
Trinucleotides – 380
Tetranucleotides – 38
Pentanucleotides – 24
Hexanucleotides – 29
No. of SSRs in compound formation - 1525

Number of SSR being mapped = 500

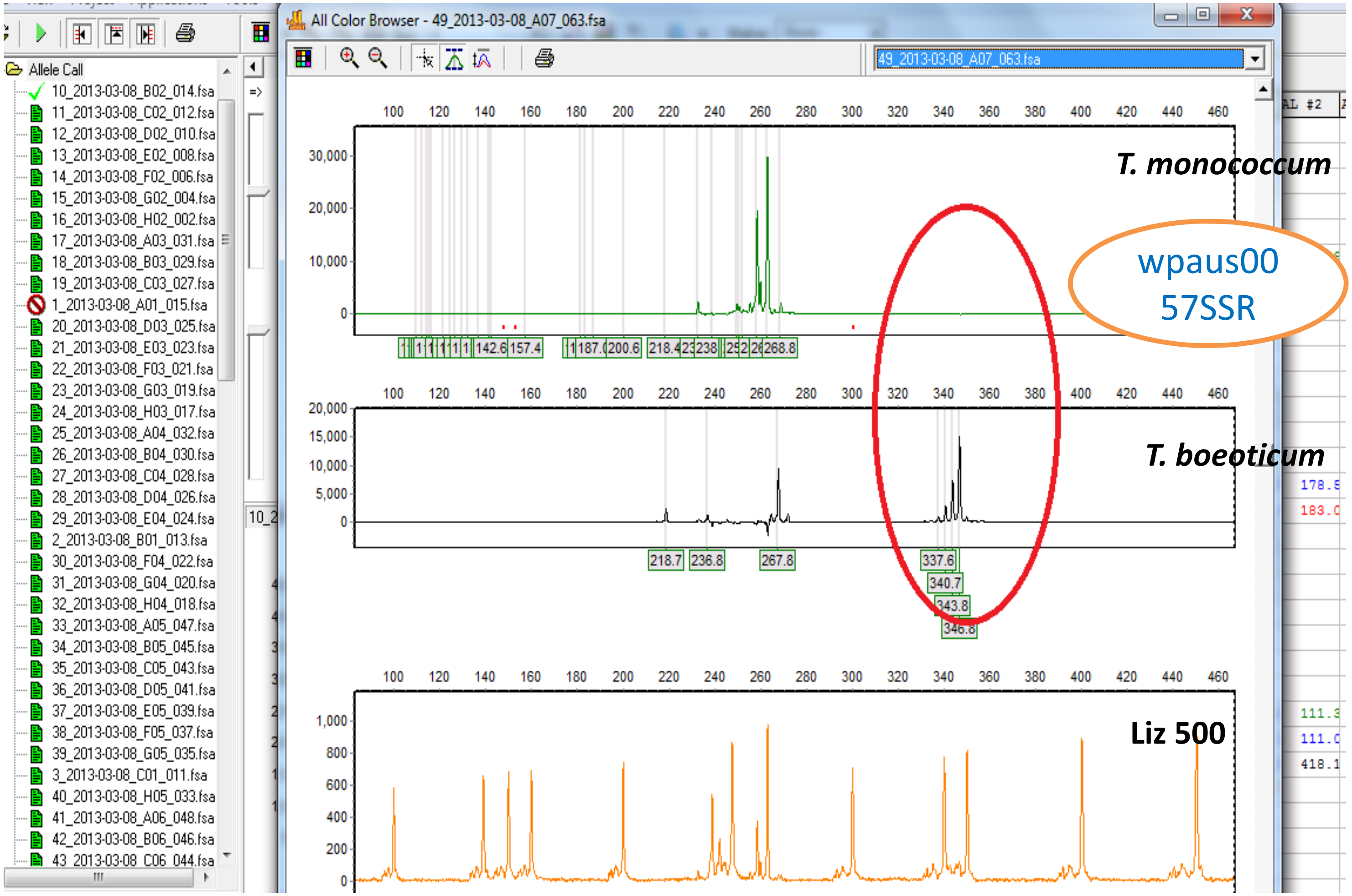
SSR Genotyping on ABI



All Color Browser - 25_2013-03-08_A04_032.fsa

Allele Call
- 10_2013-03-08_B02_014.fsa
- 11_2013-03-08_C02_012.fsa
- 12_2013-03-08_D02_010.fsa
- 13_2013-03-08_E02_008.fsa
- 14_2013-03-08_F02_006.fsa
- 15_2013-03-08_G02_004.fsa
- 16_2013-03-08_H02_002.fsa
- 17_2013-03-08_A03_031.fsa
- 18_2013-03-08_B03_029.fsa
- 19_2013-03-08_C03_027.fsa
- 1_2013-03-08_A01_015.fsa
- 20_2013-03-08_D03_025.fsa
- 21_2013-03-08_E03_023.fsa
- 22_2013-03-08_F03_021.fsa
- 23_2013-03-08_G03_019.fsa
- 24_2013-03-08_H03_017.fsa
- 25_2013-03-08_A04_032.fsa
- 26_2013-03-08_B04_030.fsa
- 27_2013-03-08_C04_028.fsa
- 28_2013-03-08_D04_026.fsa
- 29_2013-03-08_E04_024.fsa
- 2_2013-03-08_B01_013.fsa
- 30_2013-03-08_F04_022.fsa
- 31_2013-03-08_G04_020.fsa
- 32_2013-03-08_H04_018.fsa
- 33_2013-03-08_A05_047.fsa
- 34_2013-03-08_B05_045.fsa
- 35_2013-03-08_C05_043.fsa
- 36_2013-03-08_D05_041.fsa
- 37_2013-03-08_E05_039.fsa
- 38_2013-03-08_F05_037.fsa
- 39_2013-03-08_G05_035.fsa
- 3_2013-03-08_C01_011.fsa
- 40_2013-03-08_H05_033.fsa
- 41_2013-03-08_A06_048.fsa
- 42_2013-03-08_B06_046.fsa
- 43_2013-03-08_C06_044.fsa

SSR Genotyping on ABI



T. monococcum

wpaus00
57SSR

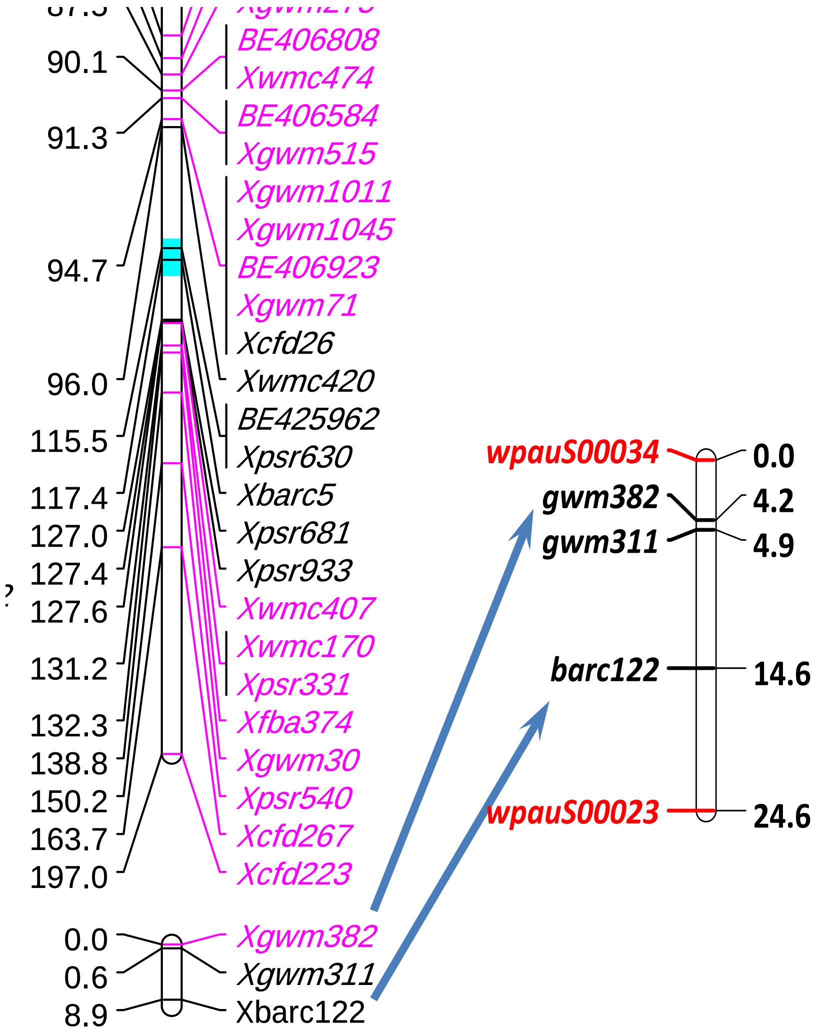
T. boeoticum

Liz 500

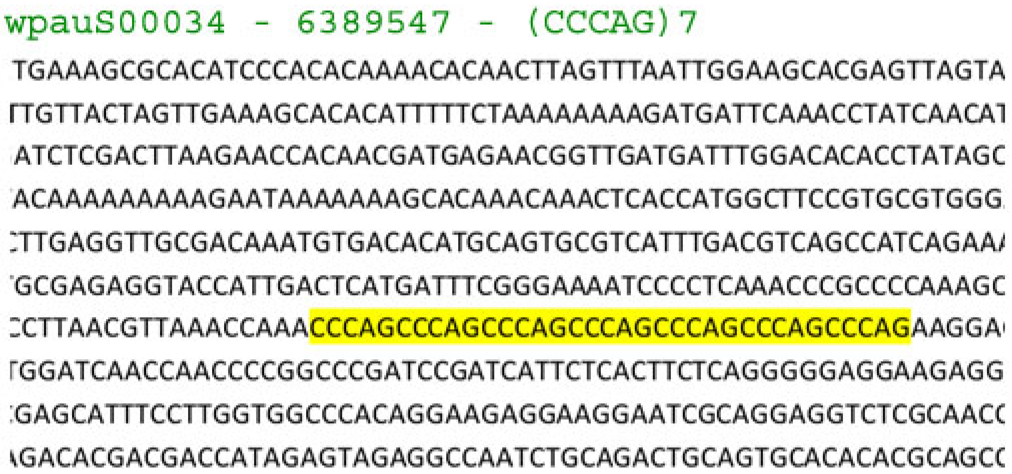
Analysis of SSR Markers

Tm	Tb	CS	No. of Markers
-	-	+	32
+	+	+	43 (Monomorphic)
+	+	+	14 (Polymorphic)
-	+	+	38 (Polymorphic)
-	-	-	01
Total			128 **(~40% Polymorphism)**

Mapping of SSRs identified from the shotgun assembly



2AL original map



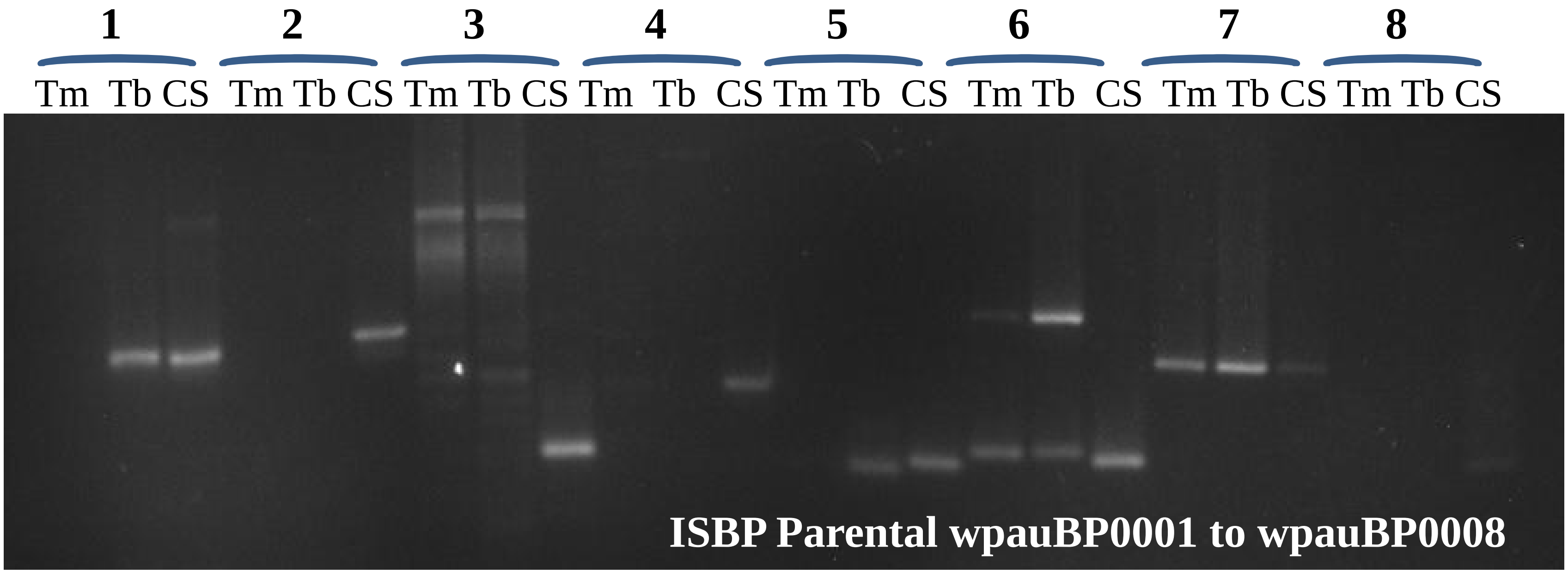
```
wpauS00034 - 6389547 - (CCCAG)7
TGAAAGCGCACATCCCACACAAAACACAACTTAGTTTAATTGGAAGCACGAGTTAGTA
TTGTTACTAGTTGAAAGCACACATTTTTCTAAAAAAAAAGATGATTCAAACCTATCAACAT
ATCTCGACTTAAGAACCACAACGATGAGAACGGTTGATGATTTGGACACACCTATAGC
ACAAAAAAAAAGAATAAAAAAAGCACAAACAAACTCACCATGGCTTCCGTGCGTGGG
TTGAGGTTGCGACAAATGTGACACATGCAGTGCGTCATTTGACGTCAGCCATCAGAA
GCGAGAGGTACCATTGACTCATGATTTCGGGAAAATCCCCTCAAACCCGCCCCAAAGC
CTTAACGTTAAACCAAACCCAGCCCAGCCCAGCCCAGCCCAGCCCAGCCCAGAAGGA
GGATCAACCAACCCCGGCCCGATCCGATCATTCTCACTTCTCAGGGGGAGGAAGAGG
GAGCATTTCCTTGGTGGCCCACAGGAAGAGGAAGGAATCGCAGGAGGTCTCGCAACC
GACACGACGACCATAGAGTAGAGGCCAATCTGCAGACTGCAGTGCACACACGCAGCC
```

```
wpauS00023 - 6340077 - (AAAG)7
ACCTAAAATCCAAACAAAATGTAAAACATTTCAAAAAAAAAATCTTATTTTGTTTTCTTG
TCTCAGTGCTTGCAACGAAATGACATTCATGGAAGTCATGGAAAAAAATAACCAAGGC
TTGAAAATAACACTTTTGGAGCATCATTCTTGTTTTTTTGCTATGACTTCCACAAATGTTAT
CGGGCACTGAGAACCTTTTGTCAATGTTTGCACCAGAAGAAAATCAGAATTTTTTAAATC
GATTTTACTGTTCGAGCTCAGAAACACCACGTCGAAAAAAAAAGTTCCTTCCTTCTTGAT
CCGGTTTCTTGATGGGTTTTGTTCGTTTGTTTGCCACACGCATGCACGCACGCACATAGG
GATCGACGCTGCTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAACGGCACCGAAAAT
ATTATATGTTTTGTGTCGGTGGTACACGTACGCTACAAGGGAGGTACGAAATAGCATAC
GGTTCAACTAAGGGCCGTTGGTGGCGTTTACGTCCGCGGTGGCCGTCACACGATGGG
GCAGCGAGGGCATTTATTATGCCAGCGACCCAGACGCTCCGTCCGTATCTGCTCATTTT
```

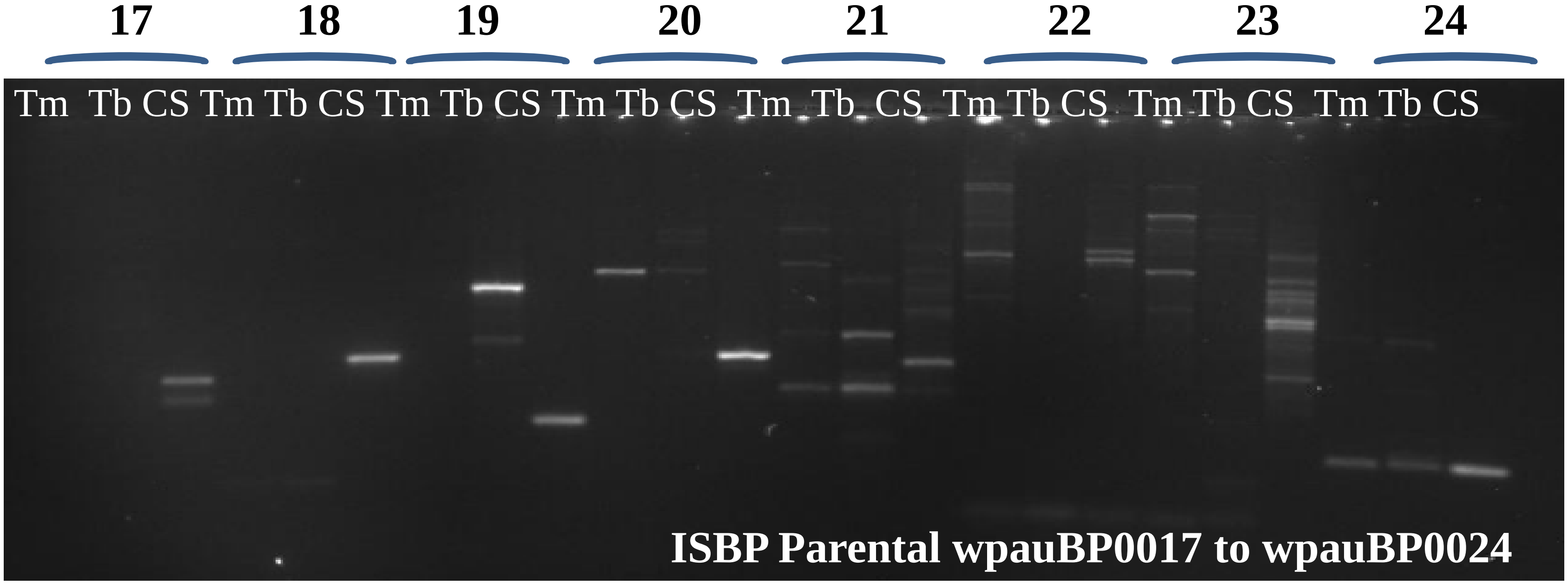
Generation of ISBP Markers

- ISBP markers were generated using ISBPFINDER.pl
- Total number of ISBPs – 216414
- Total number of ISBP markers – 12706
- Number of ISBP markers used for mapping (pilot project) – 50
- Genotyping system used – 2.5% agarose

Parental Polymorphism of ISBP Markers



ISBP Parental wpauBP0001 to wpauBP0008



ISBP Parental wpauBP0017 to wpauBP0024

Analysis of ISBP Markers

Tm	Tb	CS	Number of markers
-	-	+	11
+	+	+	8 (Polymorphic)
-	+	+	13 (Polymorphic)
+	+	+	25 (Monomorphic)
-	-	-	1 (No Amplification)
Total Loci			58 (~33% polymorphism)

Development of Gene Based Markers

HOMOLOGY SEARCHES USING 2A SPECIFIC 5’- AND 3’-EST SEQUENCES

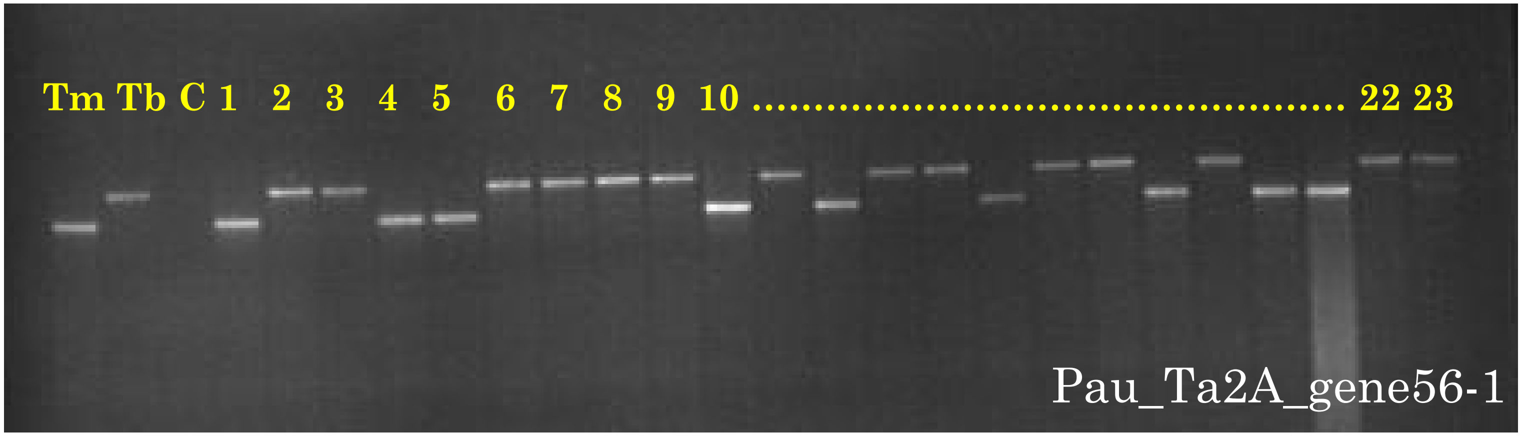
QUERY SEQUENCES	**DATABASE**	**HOMOLOGOUS SEQUENCES**	**E VALUE**
3’ AND 5’ ESTS	FlcDNAs	644	1e-10
FlcDNAs	Roche 454	**41**	1e-10
FlcDNAs	2AL-ILLUMINA	**152**	1e-10

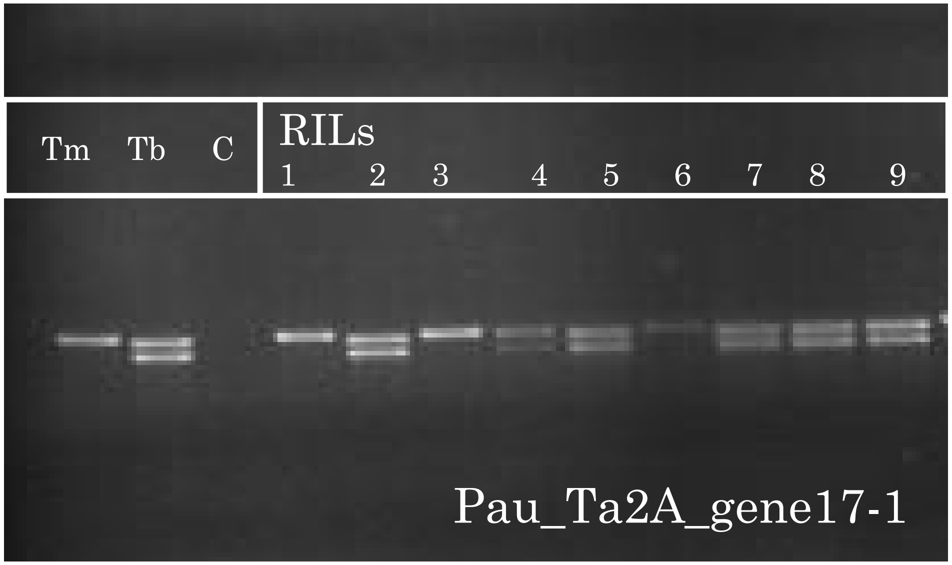
Parental Polymorphism Survey

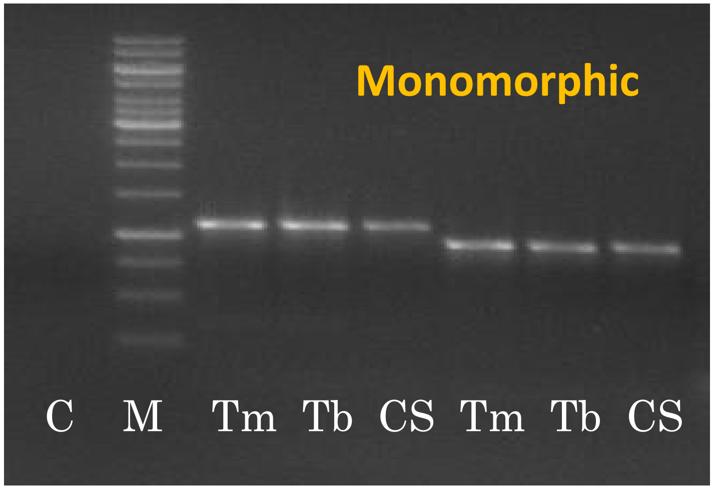
Predicted genes were used for designing of primers (Perl Primer), followed by amplification on *T. monococcum, T. boeoticum* and CS

Amplification Pattern			No. of Genes
T. monococcum	*T. boeoticum*	Chinese Spring	
+	+	+	85
-	-	-	8
+	-	+	2
-	+	+	4
+	+	-	2
-	-	+	22
Total primers			123
Size polymorphic			4
Presence/Absence			6

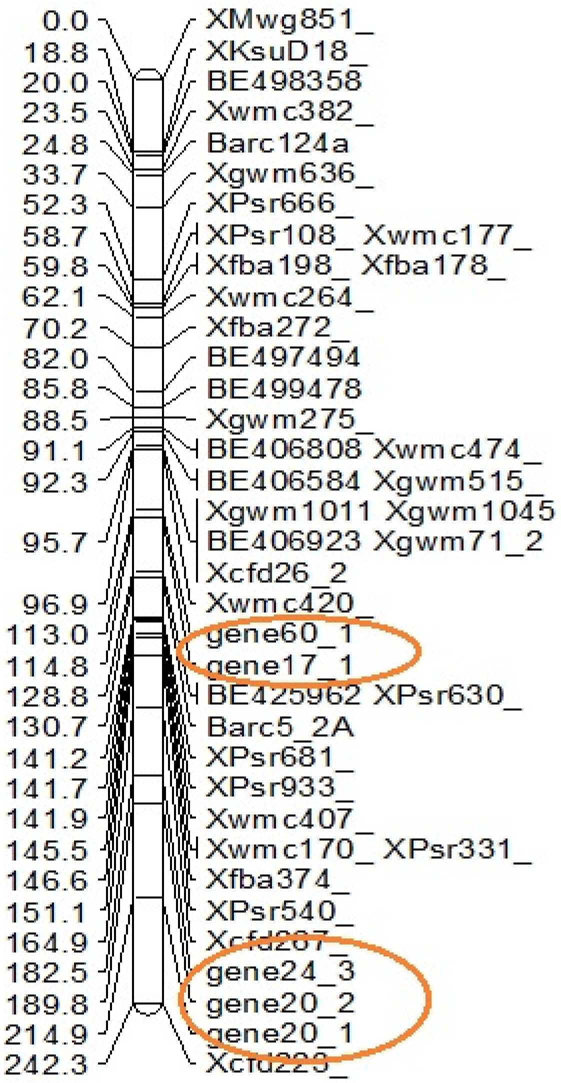
Genotyping using *T. boeoticum* X *T. monococcum* RIL population

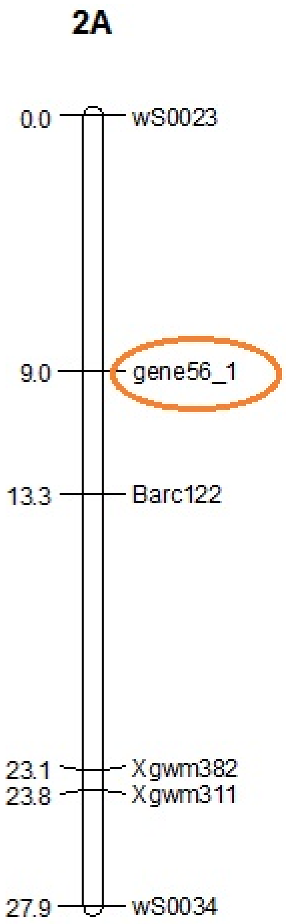




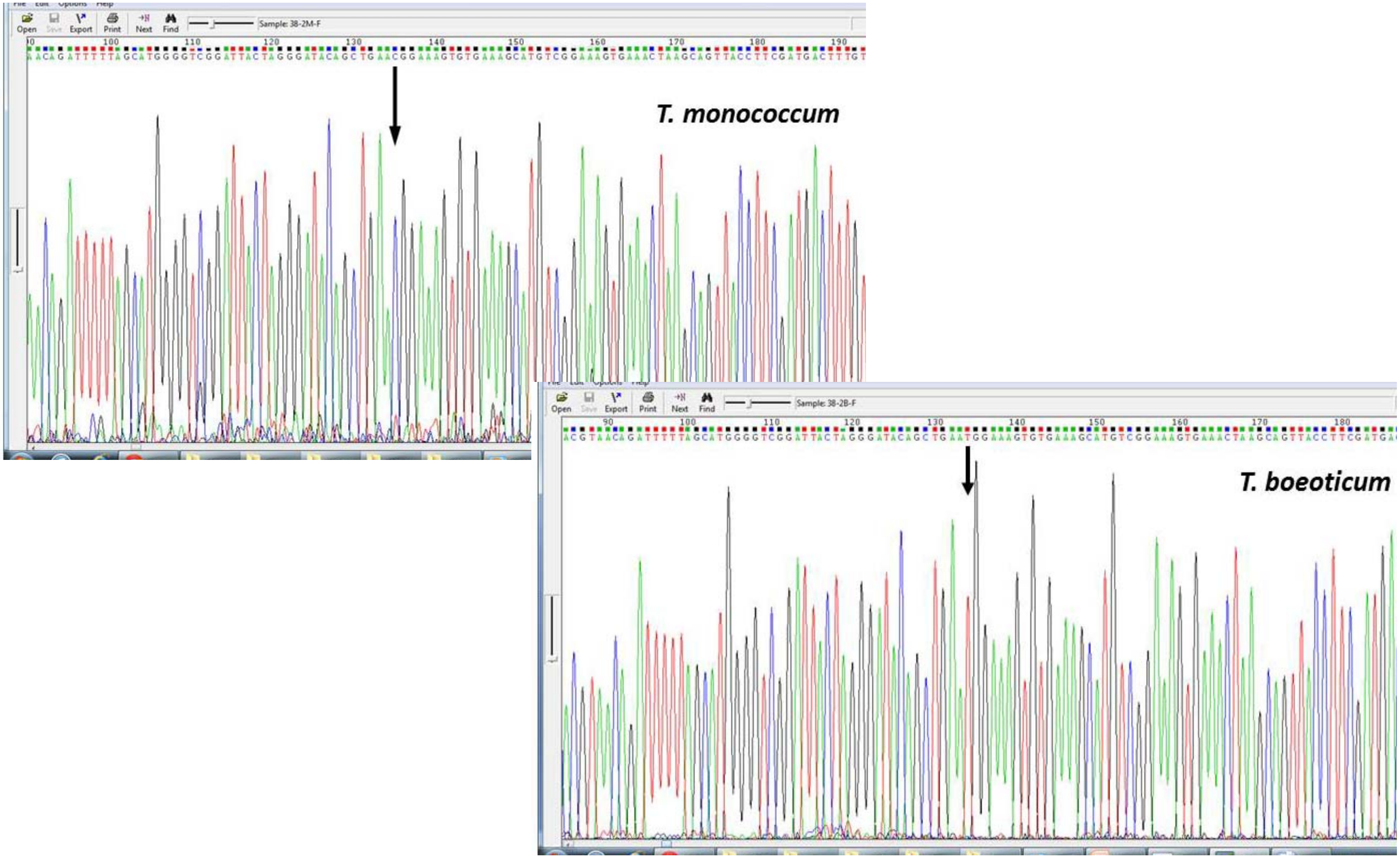


Linkage Map


0.0 XMwg851_
18.8 XKsuD18_
20.0 BE498358
23.5 Xwmc382_
24.8 Barc124a
33.7 Xgwm636_
52.3 XPsr666_
58.7 XPsr108_ Xwmc177_
59.8 Xfba198_ Xfba178_
62.1 Xwmc264_
70.2 Xfba272_
82.0 BE497494
85.8 BE499478
88.5 Xgwm275_
91.1 BE406808 Xwmc474_
92.3 BE406584 Xgwm515_
Xgwm1011 Xgwm1045
95.7 BE406923 Xgwm71_2
Xcfd26_2
96.9 Xwmc420_
113.0 gene60_1
114.8 gene17_1
128.8 BE425962 XPsr630_
130.7 Barc5_2A
141.2 XPsr681_
141.7 XPsr933_
141.9 Xwmc407_
145.5 Xwmc170_ XPsr331_
146.6 Xfba374_
151.1 XPsr540_
164.9 Xcfd267_
182.5 gene24_3
189.8 gene20_2
214.9 gene20_1
242.3 Xcfd223_


2A
0.0 wS0023
9.0 gene56_1
13.3 Barc122
23.1 Xgwm382
23.8 Xgwm311
27.9 wS0034

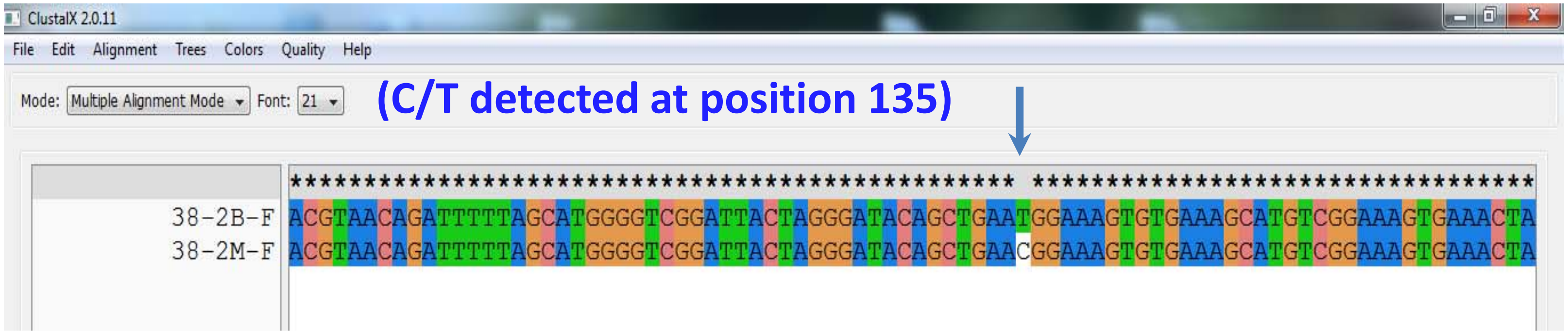
Electropherogram of an amplicon of *T. boeoticum,* T. *monococcum* (Arrow pointing to SNP, C/T)



```
Sample: 38-2M-F
90        100       110       120       130       140       150       160       170       180       190
AACAGATTTTTAGCATGGGGTCGGATTACTAGGGATACAGCTGAACGGAAAGTGTGAAAGCATGTCGGAAAGTGAAACTAAGCAGTTACCTTCGATGACTTTGT
```

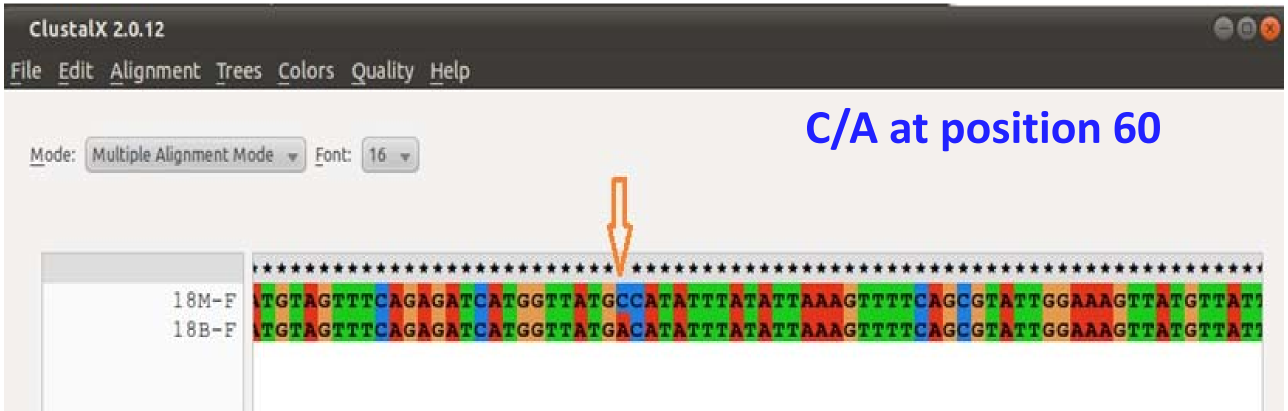
```
Sample: 38-2B-F
90        100       110       120       130       140       150       160       170       180
ACGTAACAGATTTTTAGCATGGGGTCGGATTACTAGGGATACAGCTGAATGGAAAGTGTGAAAGCATGTCGGAAAGTGAAACTAAGCAGTTACCTTCGATGA
```

SNP identified using Clustal X between *T. boeoticum* and *T. monococcum*



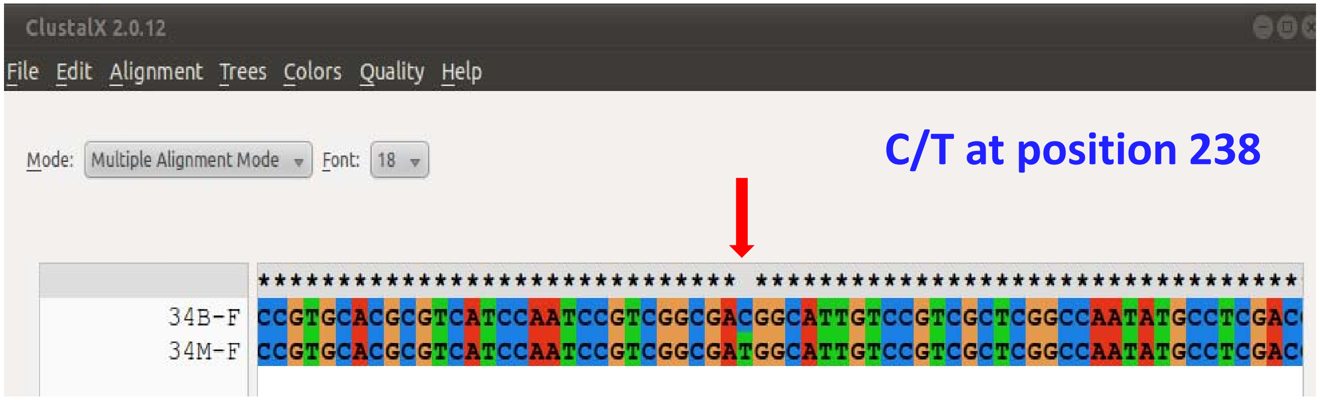
(C/T detected at position 135)

```
          ************************************************* **********************************
38-2B-F   ACGTAACAGATTTTTAGCATGGGGTCGGATTACTAGGGATACAGCTGAATGGAAAGTGTGAAAGCATGTCGGAAAGTGAAACTA
38-2M-F   ACGTAACAGATTTTTAGCATGGGGTCGGATTACTAGGGATACAGCTGAACGGAAAGTGTGAAAGCATGTCGGAAAGTGAAACTA
```



C/A at position 60

```
          ************************* ********************************************
18M-F     TGTAGTTTCAGAGATCATGGTTATGCCATATTTATATTAAAGTTTTCAGCGTATTGGAAAGTTATGTTAT
18B-F     TGTAGTTTCAGAGATCATGGTTATGACATATTTATATTAAAGTTTTCAGCGTATTGGAAAGTTATGTTAT
```



C/T at position 238

```
          ****************************** **********************************
34B-F     CCGTGCACGCGTCATCCAATCCGTCGGCGACGGCATTGTCCGTCGCTCGGCCAATATGCCTCGAC
34M-F     CCGTGCACGCGTCATCCAATCCGTCGGCGATGGCATTGTCCGTCGCTCGGCCAATATGCCTCGAC
```

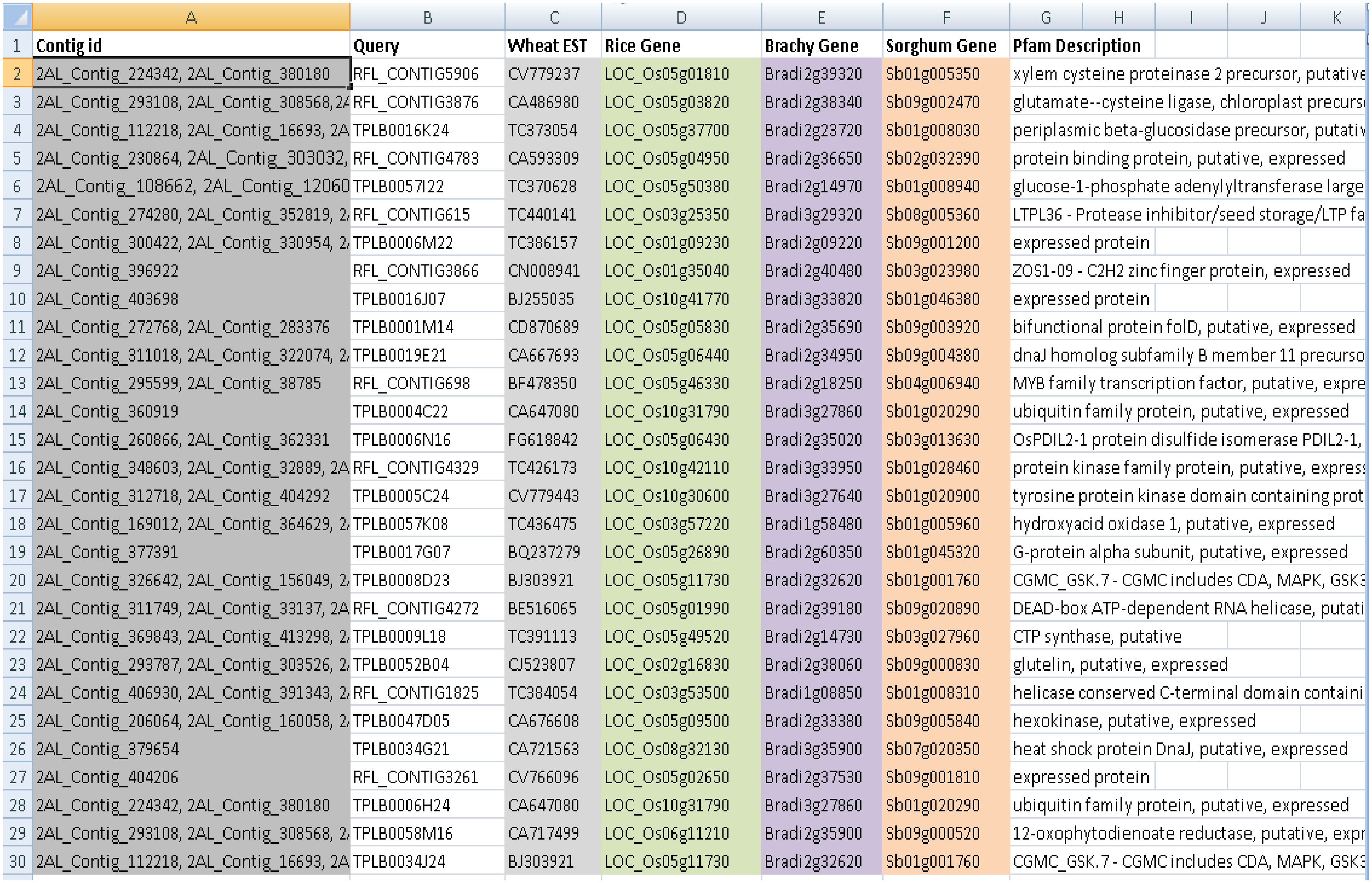
Summary

- Tm and Tb amplicons of 69 Genes have been sequenced and SNPs have been identified *(on average 1SNP/ 400bp and SNPs identified in intronic regions are more)*

- Amplicons of 16 Genes needs to be cloned for sequencing

- Genes have been identified using Fgenesh from 2AL assembly and 245 primer pairs have been designed from the 2AL contigs, which will be used for analysis of *T. monococcum* and *T. boeoticum*

Genic Similarity with other species



	A	B	C	D	E	F	G	H	I	J	K
1	**Contig id**	**Query**	**Wheat EST**	**Rice Gene**	**Brachy Gene**	**Sorghum Gene**	**Pfam Description**				
2	2AL_Contig_224342, 2AL_Contig_380180	RFL_CONTIG5906	CV779237	LOC_Os05g01810	Bradi2g39320	Sb01g005350	xylem cysteine proteinase 2 precursor, putative				
3	2AL_Contig_293108, 2AL_Contig_308568,2	RFL_CONTIG3876	CA486980	LOC_Os05g03820	Bradi2g38340	Sb09g002470	glutamate--cysteine ligase, chloroplast precurs				
4	2AL_Contig_112218, 2AL_Contig_16693, 2A	TPLB0016K24	TC373054	LOC_Os05g37700	Bradi2g23720	Sb01g008030	periplasmic beta-glucosidase precursor, putativ				
5	2AL_Contig_230864, 2AL_Contig_303032,	RFL_CONTIG4783	CA593309	LOC_Os05g04950	Bradi2g36650	Sb02g032390	protein binding protein, putative, expressed				
6	2AL_Contig_108662, 2AL_Contig_12060	TPLB0057I22	TC370628	LOC_Os05g50380	Bradi2g14970	Sb01g008940	glucose-1-phosphate adenylyltransferase large				
7	2AL_Contig_274280, 2AL_Contig_352819, 2	RFL_CONTIG615	TC440141	LOC_Os03g25350	Bradi3g29320	Sb08g005360	LTPL36 - Protease inhibitor/seed storage/LTP fa				
8	2AL_Contig_300422, 2AL_Contig_330954, 2	TPLB0006M22	TC386157	LOC_Os01g09230	Bradi2g09220	Sb09g001200	expressed protein				
9	2AL_Contig_396922	RFL_CONTIG3866	CN008941	LOC_Os01g35040	Bradi2g40480	Sb03g023980	ZOS1-09 - C2H2 zinc finger protein, expressed				
10	2AL_Contig_403698	TPLB0016J07	BJ255035	LOC_Os10g41770	Bradi3g33820	Sb01g046380	expressed protein				
11	2AL_Contig_272768, 2AL_Contig_283376	TPLB0001M14	CD870689	LOC_Os05g05830	Bradi2g35690	Sb09g003920	bifunctional protein folD, putative, expressed				
12	2AL_Contig_311018, 2AL_Contig_322074, 2	TPLB0019E21	CA667693	LOC_Os05g06440	Bradi2g34950	Sb09g004380	dnaJ homolog subfamily B member 11 precurso				
13	2AL_Contig_295599, 2AL_Contig_38785	RFL_CONTIG698	BF478350	LOC_Os05g46330	Bradi2g18250	Sb04g006940	MYB family transcription factor, putative, expre				
14	2AL_Contig_360919	TPLB0004C22	CA647080	LOC_Os10g31790	Bradi3g27860	Sb01g020290	ubiquitin family protein, putative, expressed				
15	2AL_Contig_260866, 2AL_Contig_362331	TPLB0006N16	FG618842	LOC_Os05g06430	Bradi2g35020	Sb03g013630	OsPDIL2-1 protein disulfide isomerase PDIL2-1,				
16	2AL_Contig_348603, 2AL_Contig_32889, 2A	RFL_CONTIG4329	TC426173	LOC_Os10g42110	Bradi3g33950	Sb01g028460	protein kinase family protein, putative, express				
17	2AL_Contig_312718, 2AL_Contig_404292	TPLB0005C24	CV779443	LOC_Os10g30600	Bradi3g27640	Sb01g020900	tyrosine protein kinase domain containing prot				
18	2AL_Contig_169012, 2AL_Contig_364629, 2	TPLB0057K08	TC436475	LOC_Os03g57220	Bradi1g58480	Sb01g005960	hydroxyacid oxidase 1, putative, expressed				
19	2AL_Contig_377391	TPLB0017G07	BQ237279	LOC_Os05g26890	Bradi2g60350	Sb01g045320	G-protein alpha subunit, putative, expressed				
20	2AL_Contig_326642, 2AL_Contig_156049, 2	TPLB0008D23	BJ303921	LOC_Os05g11730	Bradi2g32620	Sb01g001760	CGMC_GSK.7 - CGMC includes CDA, MAPK, GSK				
21	2AL_Contig_311749, 2AL_Contig_33137, 2A	RFL_CONTIG4272	BE516065	LOC_Os05g01990	Bradi2g39180	Sb09g020890	DEAD-box ATP-dependent RNA helicase, putati				
22	2AL_Contig_369843, 2AL_Contig_413298, 2	TPLB0009L18	TC391113	LOC_Os05g49520	Bradi2g14730	Sb03g027960	CTP synthase, putative				
23	2AL_Contig_293787, 2AL_Contig_303526, 2	TPLB0052B04	CJ523807	LOC_Os02g16830	Bradi2g38060	Sb09g000830	glutelin, putative, expressed				
24	2AL_Contig_406930, 2AL_Contig_391343, 2	RFL_CONTIG1825	TC384054	LOC_Os03g53500	Bradi1g08850	Sb01g008310	helicase conserved C-terminal domain containi				
25	2AL_Contig_206064, 2AL_Contig_160058, 2	TPLB0047D05	CA676608	LOC_Os05g09500	Bradi2g33380	Sb09g005840	hexokinase, putative, expressed				
26	2AL_Contig_379654	TPLB0034G21	CA721563	LOC_Os08g32130	Bradi3g35900	Sb07g020350	heat shock protein DnaJ, putative, expressed				
27	2AL_Contig_404206	RFL_CONTIG3261	CV766096	LOC_Os05g02650	Bradi2g37530	Sb09g001810	expressed protein				
28	2AL_Contig_224342, 2AL_Contig_380180	TPLB0006H24	CA647080	LOC_Os10g31790	Bradi3g27860	Sb01g020290	ubiquitin family protein, putative, expressed				
29	2AL_Contig_293108, 2AL_Contig_308568, 2	TPLB0058M16	CA717499	LOC_Os06g11210	Bradi2g35900	Sb09g000520	12-oxophytodienoate reductase, putative, expr				
30	2AL_Contig_112218, 2AL_Contig_16693, 2A	TPLB0034J24	BJ303921	LOC_Os05g11730	Bradi2g32620	Sb01g001760	CGMC_GSK.7 - CGMC includes CDA, MAPK, GSK				

SUMMARY

- Using HICF we have fingerprinted >50 000 BAC clones 2AL and about 60% of these yielded good fingerprint suitable for contig generation
- More than 21,000 fingerprinted BACs were used for generating contigs and ~11,000 BACs were included in contigs of varying sizes
- Clones with Minimum Tilling Path (MTP) have also been generated
- Fingerprinting and contig generation of the remaining BACs is in progress
- *De novo* hybrid assembly resulted into 425,821 contigs for 2AL, covering 63% of total genome of 2AL
- Assembly of both the arms was analyzed for the repetitive elements includes 4,27,402 repeat elements for 2AL

SUMMARY

- About 28,000 gene sequences have been identified in 2A
- SSR mining was done on assembled data and includes 22,460 SSRs for 2AL
- Bin mapped ESTs of chromosome 2A were used to fish out the full length cDNAs which were then used to BLAST against assembled data to identify corresponding gene sequences in the assembled data
- Primers were designed from the CS sequence data and used for amplifying *T. boeoticum* and *T. Monococcum* to generate SNP markers
- SSR and SNP markers will be subsequently mapped on RIL population of diploid wheat (*T. boeoticum/T*. *monococcum*) to enrich the genetic map of 2A which can then be used to align the physical map to the genetic map

Group Meetings

Dr Ananth Kumar, WSU visited PAU from Dec 22- 31, 2012

Network Group members met in May 2013

THANKS

International Collaborators

1. Dr. Jaroslav Dolezel - Czech Republic
2. Catherine Feuillet - INRA, France
3. Bikram S. Gill - KSU, USA
4. Sunish Seghal - KSU, USA
5. Beat Keller - University of Zurich, Switzerland
6. Jane Rogers – IWGSC, UK
7. Kellye Eversole - Executive Director, IWGSC