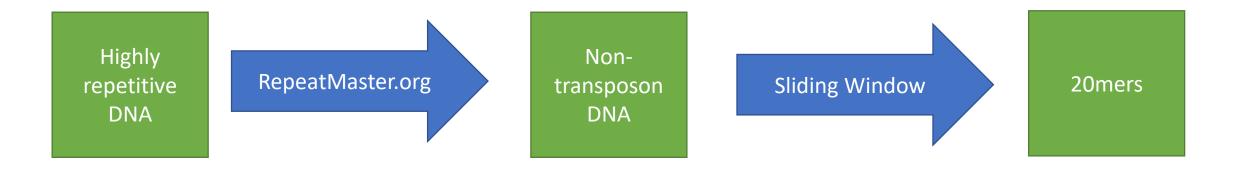
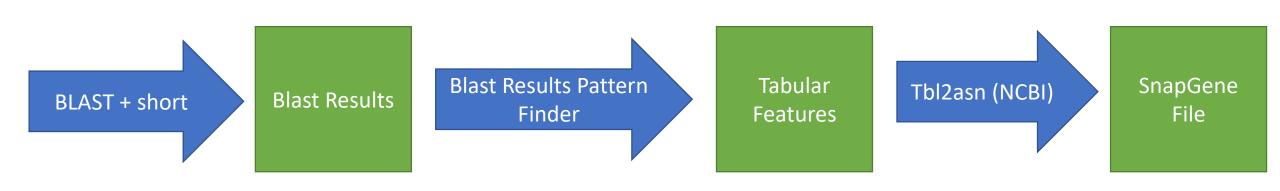
Automated Primer Region Analysis

Katherine Braught
Peterson Lab

Program Overview





Repeat Masker: (P1-wwB54)

Input: FASTA/text DNA file

Output: Masked Sequence, Table

SW	perc perc p	erc	query	positi	ion in	query		matching	repeat	posi	tion in	repeat		
score	div. del. i	ns.	sequence	begin	end	(left)		repeat	class/family	begin	end	(left)	ID	
11496		0.6	UnnamedSequence			,		Copia13-ZM_I-int	The state of the s	3423		(1471)	1	
4250		0.6	UnnamedSequence	1551		,		ZMCopia2_LTR	LTR/Copia	713		` '	2	*
20674		0.0	UnnamedSequence	2537		(22775)		_	DNA/hAT-Ac	(118)	4692		3	
27		5.1	UnnamedSequence	3230		,		(AGCTAGC)n	Simple_repeat	1	59	(0)	4	
20674		0.2	UnnamedSequence	3291		(20368)		_	DNA/hAT-Ac	(871)	3939	1594	3	
48		0.0	UnnamedSequence	5637		,		(GGTTGT)n	Simple_repeat	1	52	(0)	5	
12349		0.3	UnnamedSequence	5689		(18899)		_	DNA/hAT-Ac	(3270)	1540		6	
11298		0.0	UnnamedSequence	7436		(17879)		_	DNA/hAT-Ac	(121)	4689	3940	7	
27		5.1	UnnamedSequence	8126		,		(AGCTAGC)n	Simple_repeat	1		(0)	8	
11298		0.3	UnnamedSequence	8187		(16524)		_	DNA/hAT-Ac	(871)	3939	2646	7	
12		2.9	UnnamedSequence	9481		,		(TAGT)n	Simple_repeat	1		(0)	9	
335	25.0 0.0	0.0	UnnamedSequence	10331	10414	(15590)	+	DNA-TA-1_ZM	DNA/TcMar-Stowaway	1	84	(0)	10	
17		0.0	UnnamedSequence			(14755)		` '	Simple_repeat	1		(0)	11	
15		0.0	UnnamedSequence			` '		(GCTC)n	Simple_repeat	1		(0)	12	
874	32.9 3.1	3.4	UnnamedSequence	14438	14946	(11058)	C	Gypsy-200_ZM-I	LTR/Gypsy	(5237)	973	466	13	
300	36.8 3.6	0.0	UnnamedSequence	16416	16581	(9423)	C	ZMCOPIA3_LTR	LTR/Copia	(221)	1216	1045	14	
12	4.7 0.0	8.3	UnnamedSequence	16614	16639	(9365)	+	(AATA)n	Simple_repeat	1	24	(0)	15	
624	19.3 9.1	2.4	UnnamedSequence	16740	16893	(9111)	+	HARB-N11_ZM	DNA/PIF-Harbinger	4	167	(215)	16	
927	11.5 1.4	0.7	UnnamedSequence	16899	17038	(8966)	C	PREM1A_ZM_LTR	LTR/Copia	(1594)	1814	1674	17	
406	21.2 8.2 1	1.0	UnnamedSequence	16901	17059	(8945)	+	PREM1A_ZM_LTR	LTR/Copia	1674	1828	(1580)	17	*
16	19.9 5.8	0.0	UnnamedSequence	17109	17160	(8844)	+	(ATTTATA)n	Simple_repeat	1	55	(0)	18	
234	11.4 0.0	0.0	UnnamedSequence	17256	17290	(8714)	+	HARB-N14_ZM	DNA/PIF-Harbinger	187	221	(43)	19	
234	11.4 0.0	0.0	UnnamedSequence	17352	17386	(8618)	+	HARB-N14_ZM	DNA/PIF-Harbinger	187	221	(43)	20	
692	14.9 0.0	0.0	UnnamedSequence	17389	17495	(8509)	+	DNA-6_ZM	DNA/PIF-Harbinger	1	107	(23)	21	
15	19.4 2.3	2.3	UnnamedSequence	17904	17946	(8058)	+	(CCG)n	Simple_repeat	1	43	(0)	22	
17	20.8 0.0	2.2	UnnamedSequence	19024	19069	(6935)	+	(GCCCCG)n	Simple_repeat	1	45	(0)	23	
23	4.2 12.2	0.0	UnnamedSequence	19323	19371	(6633)	+	(GACGAG)n	Simple_repeat	1	55	(0)	24	
16	16.3 0.0	2.8	UnnamedSequence	19414	19450	(6554)	+	G-rich	Low_complexity	1	36	(0)	25	
13	21.7 2.6	0.0	UnnamedSequence	19754	19791	(6213)	+	GA-rich	Low_complexity	1	39	(0)	26	
15	0.0 0.0	0.0	UnnamedSequence	19820	19838	(6166)	+	(GCC)n	Simple_repeat	1	19	(0)	27	
12	20.7 5.1	0.0	UnnamedSequence	24544	24582	(1422)	+	GA-rich	Low_complexity	1	41	(0)	28	
394	13.3 3.5	9.1	UnnamedSequence	25339	25454	(550)	+	STOWAWAY1_ZM	DNA/TcMar-Stowaway	71	180	(55)	29	
12	17.6 3.1	0.0	UnnamedSequence	25953	25984	(20)	+	(GCGCCC)n	Simple_repeat	1	33	(0)	30	
			<u> </u>			` ,		•				. ,		

NNNNNNNNNNNNNAGCTGCTTTCACAAACCATGTTTCCATGACTGTGTGG CACCGCTAGCTCTCGCAACACCATGTGGGTTGTAGTCCCACATGGTGTTG CGAGAGCTAGCGGTGCCACACAGTCATGGAAACATGGTTTGTGAAAGCAG CTTAACTAATTACTAGCTAGCTGTGAGAGAGTCGTCAGCGGCTAGCGGAG TGCAAAGCGCGCATTGGCTAGCTTCCCTGTTGCCTTGATCTGATCATAGA GGGGTCCCTGTCCGTGGCAAACAGTGGGATTAATGACGGCCCACGCCGGC CCTCCCTAATGATTTCCCCCGCCTTAACGTTGACACTGCGGACGGTGAGA GGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCCCATCGTCCTTC GCGAACGATGCGTACATTTAGCGCCATATATAAACACACGCCGGCACATG ATTACTACGGCCTATATACGCGCGTCGTGAGCGGCGCGATCAATGAGTGA GCTAGTTCGTTGCACCAAATTGTCGATCAAGTTGCATCAGAGATCAGACA NNNNNNNNNNNNTGTACTAGGCAGGCTAAGAAACACGATAGATACCAC GCATTCGGTTCTCATTACGAAGCTGCAGCTGCCCAAGTGCCCAACCAGCA GCGATGATCACGTACGCTCACCATCCTGCGTCCTTGCGGTTTAAATTAAT TCCCCGGCCGAACAGACAACGAGATCAAGAACTACTGGAACTCGCACCTC AGCCGGCAGATCCACACGTACCGCCGGAAATACACCGCCGGGCCTGACGA CACCGCCATCGCCATCGACATGAGCAAGCTGCAGAGCGCCGACAGGCGGC GCGGCGGCAGGACCCCGGGCCGCCGAAGGCTAGCGCCAGCAGGACC

Sliding Window

```
> Seqence 7
GCCCGTTCGCCCGCCCTCA
> Seqence 8

Description of the sequence se
```

Input: masked DNA file

Output: list of 20mers (FASTA)

```
> Segence 3
GGCCGCCCGTTCGCCCGCC
> Segence 4
GCCGCCCGTTCGCCCGCCC
> Segence 5
CCGCCCGTTCGCCCGCCCT
> Segence 6
CGCCCCGTTCGCCCGCCCTC
> Segence 7
GCCCCGTTCGCCCGCCCTCA
> Segence 8
> Segence 9
CCCGTTCGCCCGCCCTCAGG
> Segence 10
CCGTTCGCCCGCCCTCAGGG
> Segence 11
CGTTCGCCCGCCCTCAGGGC
> Segence 12
GTTCGCCCGCCCTCAGGGCC
> Segence 13
TTCGCCCGCCCTCAGGGCCG
> Segence 14
TCGCCCGCCCTCAGGGCCGG
> Segence 15
CGCCCGCCTCAGGGCCGGC
> Segence 16
CCCCCCCCTCAGGGCCGGCG
> Segence 17
CCCGCCCTCAGGGCCGGCGC
> Segence 582
TGTTGTCTCTCTCTCAGA
> Segence 583
GTTGTCTCTCTCTCAGAT
> Segence 584
TTGTCTCTCTCTCAGATA
> Segence 585
TGTCTCTCTCTCAGATAA
> Segence 586
```

CGGCCGCCCCGTTCGCCCGC

P2-wwB54

Blast + short

blastn -db maize.fa -query 20mer.fa -out results.out -task blastn-short -outfmt "7 sstart" -evalue .006

Input: list of 20mers

Output: BLAST tabular results

```
# 1 hits found
48356408
# BLASTN 2.6.0+
# Query: Segence 2210
# Database: maize.fa
# Fields: s. start
# 1 hits found
48356407
# BLASTN 2.6.0+
# Query: Seqence 2211
# Database: maize.fa
# Fields: s. start
# 1 hits found
48356406
# BLASTN 2.6.0+
# Query: Segence 2212
# Database: maize.fa
# Fields: s. start
# 5 hits found
48356405
48436464
48449099
48576294
48588931
# BLASTN 2.6.0+
# Query: Segence 2213
# Database: maize.fa
# Fields: s. start
# 5 hits found
48356404
48436463
48449098
48576293
48588930
# BLASTN 2.6.0+
# Query: Segence 2214
# Database: maize.fa
# Fields: s. start
# 5 hits found
48356403
48436462
48449097
```

Tabular Features Creator: .tbl

Input:

BLAST results

transposon table

Output:

Feature Table

Summary File

100		l'kaataki thi			
re	!5	Creator: .tbl	1551	2182	misc_fe
			2537	3229	misc_fe
		15 matches_after_B73_Blast	2020	2200	
2079	2132	<pre>misc_feature note less than 10 matches found</pre>	3230	3290	misc_fe
2133	2151	misc feature	2201	5626	micc fo
2133	2232	note no matches found	3291	5636	misc_fe
2152	2223	misc_feature	5627	E600	micc fo
		note less than 10 matches found	5637	5688	misc_fe
2224	2243	misc_feature	5600	74.05	
2244	2252	note no matches found	5689	7105	misc_fe
2244	2252	<pre>misc_feature note less than 10 matches found</pre>		0405	
2253	2268	misc feature	7436	8125	misc_fe
LLJJ	2200	note no matches found			
2269	2286	misc_feature	8126	8186	misc_fe
		note less than 10 matches found			
2287	2304	misc_feature	8187	9480	misc_fe
l		note no matches found			
2305	2308	misc_feature	9481	9513	misc_fe
2309	2332	note more than 10 matches found misc_feature			
2309	2332	note no matches found	10331	10414	misc_fe
2333	2380	misc feature			
i		note less than 10 matches found	11229	11249	misc_fe
2381	2398	misc_feature			
		note no matches found	11374	11400	misc_fe
2399	2566	misc_feature			_
2567	2602	note less than 10 matches found misc feature	14438	14946	misc fe
2307	2002	note no matches found			_
2603	2828	misc feature	16416	16581	misc_fe
		note less than 10 matches found			_
2829	2856	misc_feature	16614	16639	misc fe
		note no matches found			_
2857	2875	misc_feature	16740	16893	misc fe
2000	2022	note less than 10 matches found	10740	10055	250
2980	3032	misc_feature note less than 10 matches found	16899	17038	misc fe
3033	3052	misc feature	10055	17030	m136_16
		note no matches found	16901	17059	misc fe
3053	3179	misc_feature	10301	11000	m±36_16
		note less than 10 matches found	17109	17160	misc_fe
3180	3187	misc_feature	1/109	1/100	mrsc_16
		note more than 10 matches found			

		note	less than 10 found
1	1555	misc_feature	
		note	Transposon_LTR/Copia
1551	2182	<pre>misc_feature</pre>	
		note	Transposon_LTR/Copia
2537	3229	misc_feature	
1		note	Transposon_DNA/hAT-Ac
3230	3290	misc_feature	
			Transposon_Simple_repeat
3291	5636	misc_feature	
l			Transposon_DNA/hAT-Ac
5637	5688	misc_feature	
	7405	_	Transposon_Simple_repeat
5689	7105	misc_feature	Turney DNA /bat as
7426	0435		Transposon_DNA/hAT-Ac
7436	8125	misc_feature	Transpasan DNA/bAT As
8126	8186	misc feature	Transposon_DNA/hAT-Ac
0120	0100	_	Transposon Simple repeat
8187	9480	misc feature	Transposon_simple_repeat
0107	5400	_	Transposon DNA/hAT-Ac
9481	9513	misc feature	Transposon_binA/ nAT Ac
3401	3313	_	Transposon Simple repeat
10331	10414	misc feature	
		_	Transposon DNA/TcMar-Stowaway
11229	11249	misc_feature	, = ,
		note	Transposon_Simple_repeat
11374	11400	misc_feature	
		note	Transposon_Simple_repeat
14438	14946	misc_feature	
			Transposon_LTR/Gypsy
16416	16581	misc_feature	
			Transposon_LTR/Copia
16614	16639	misc_feature	
			Transposon_Simple_repeat
16740	16893	misc_feature	T
46000	47020		Transposon_DNA/PIF-Harbinger
16899	17038	misc_feature	Transpager LTD/Conic
16901	17059		Transposon_LTR/Copia
10901	1/039	misc_feature	Transposon LTR/Copia
17109	17160	misc feature	II alisposoli_LIK/Copia
1/103	1/100	_	Transposon Simple repeat
		IIVIE	THE PERSON OF TH

Secondary Output

```
Summary Of Created Features For P1 wwB54
Input File: P1B54blastresults.OUT
Transposon File: transposontableData.txt
Created Features:
Number Of Transposons: 33
Number of Areas with no matches in B73: 101
Number of Areas with 10+ matches in B73: 10
Number of Regions with less than 10 matches: 425
        Number of Regions with all matches found in P cluster of B73: 227
        Number of Regions with no matches found in the P cluster of B73: 38
        Number of Regions with some matches found in P cluster of B73: 160
Total Features Created: 569
Blast Results Analyzed: 14953
```

tbl2asn (NCBI)

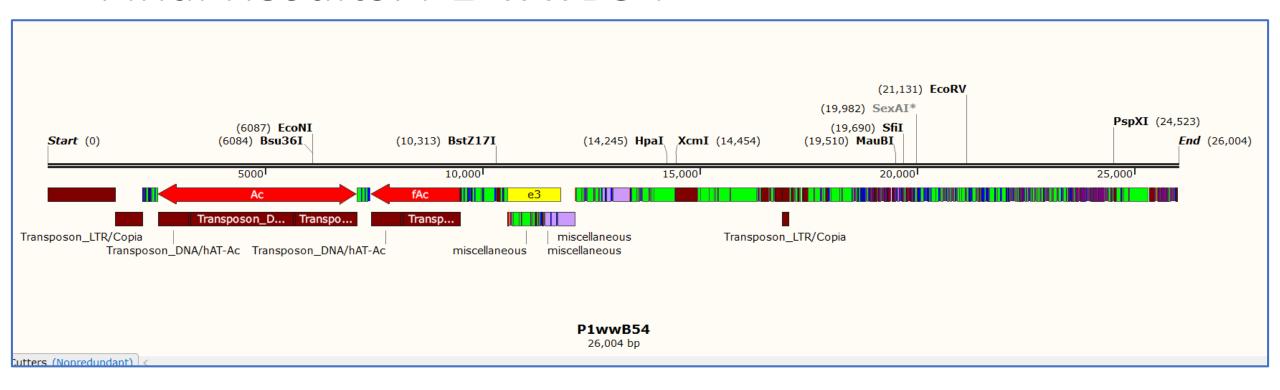
tbl2asn -t template.sbt -p . -j "[organism=P1_wwB54]" -V vb

t= template file

- -p . = directory
- -j = source qualifier
- -V vb =generate GenBank file

```
LOCUS
                                                        linear
                                                                     31-MAR-2017
            P1 wwB54
                                   26004 bp
                                               DNA
DEFINITION P1-wwB54.
ACCESSION
VERSION
KEYWORDS
SOURCE
            P1-wwB54
  ORGANISM P1-wwB54
            Unclassified.
REFERENCE
           1 (bases 1 to 26004)
  AUTHORS
            Test seq,T.S.
  TTTLE
            test seq
  JOURNAL
            Unpublished
REFERENCE
            2 (bases 1 to 26004)
  AUTHORS
            Test seq,T.S.
            Direct Submission
  TTTLE
  JOURNAL
            Submitted (31-MAR-2017) test, Test, test, Ames, IA 50012, United
            States of America
FEATURES
                     Location/Qualifiers
                     1..26004
     source
                     /organism="P1-wwB54"
                     /mol type="genomic DNA"
     misc feature
                     1..1555
                     /note="Transposon LTR/Copia"
     misc feature
                     1551..2182
                     /note="Transposon LTR/Copia"
     misc feature
                     2183..2185
                     /note="less than 10 found and all matches are in the
                     pcluster"
     misc feature
                     2186..2204
                     /note="no hits found"
                     2205..2224
     misc feature
                     /note="less than 10 found and all matches are in the
                     pcluster"
     misc feature
                     2225..2254
                     /note="no hits found"
     misc feature
                     2255..2316
                     /note="less than 10 found and all matches are in the
                     pcluster"
     misc feature
                     2317..2344
                     /note="no hits found"
     misc feature
                     2345..2351
                     /note="less than 10 found and all matches are in the
                     pcluster"
```

Final Results: P1-wwB54



Dark Red= transposon

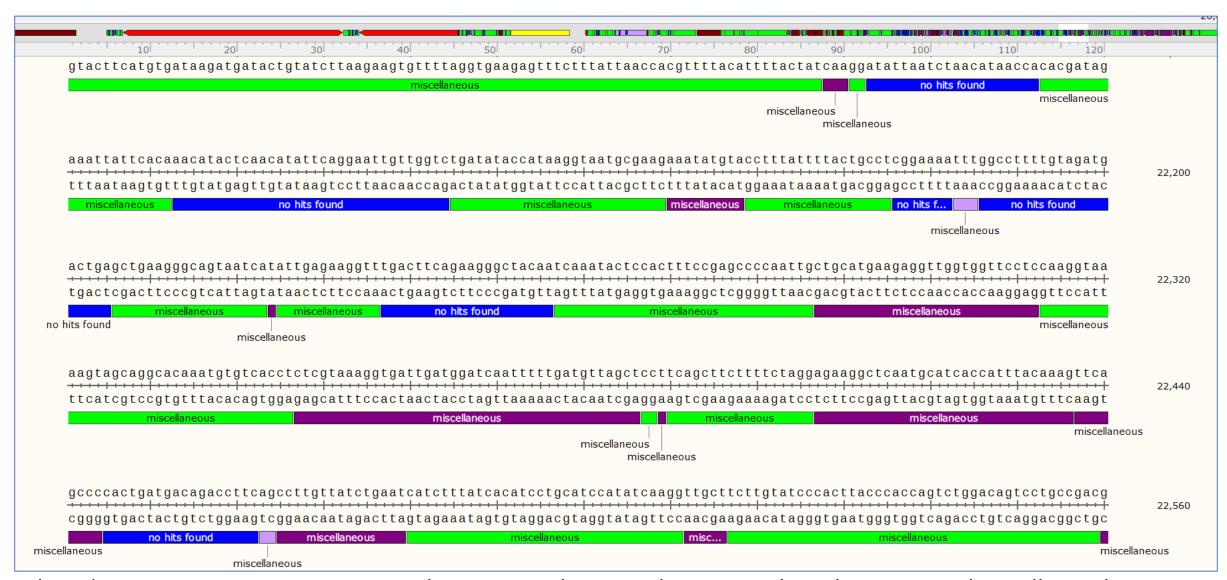
Orange= more than 10 matches

Bright Green = less than 10 matches, all matches in p cluster

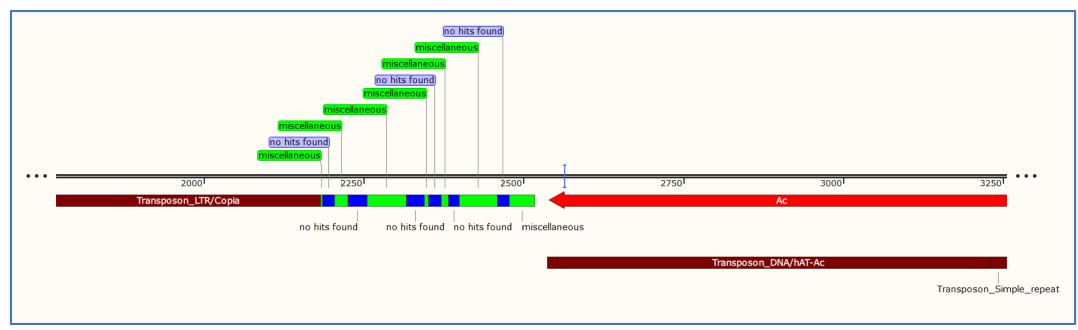
Blue= not matches in B73

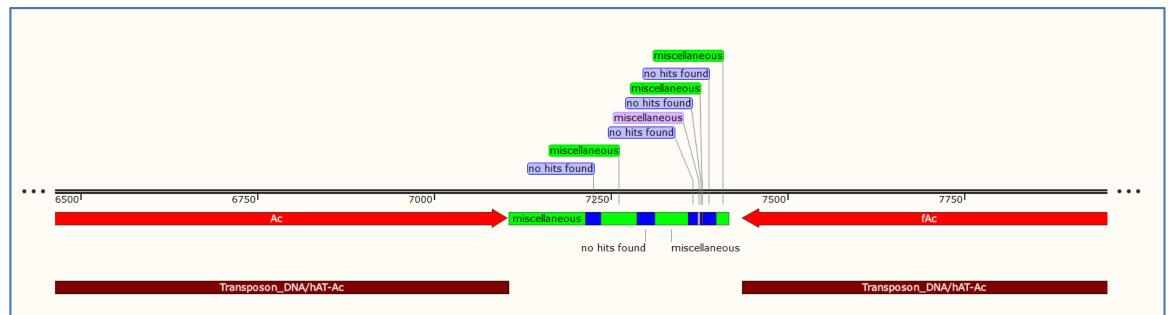
Light purple = less than 10 matches, no matches in p cluster

Dark purple = less than 10 matches, some in p cluster

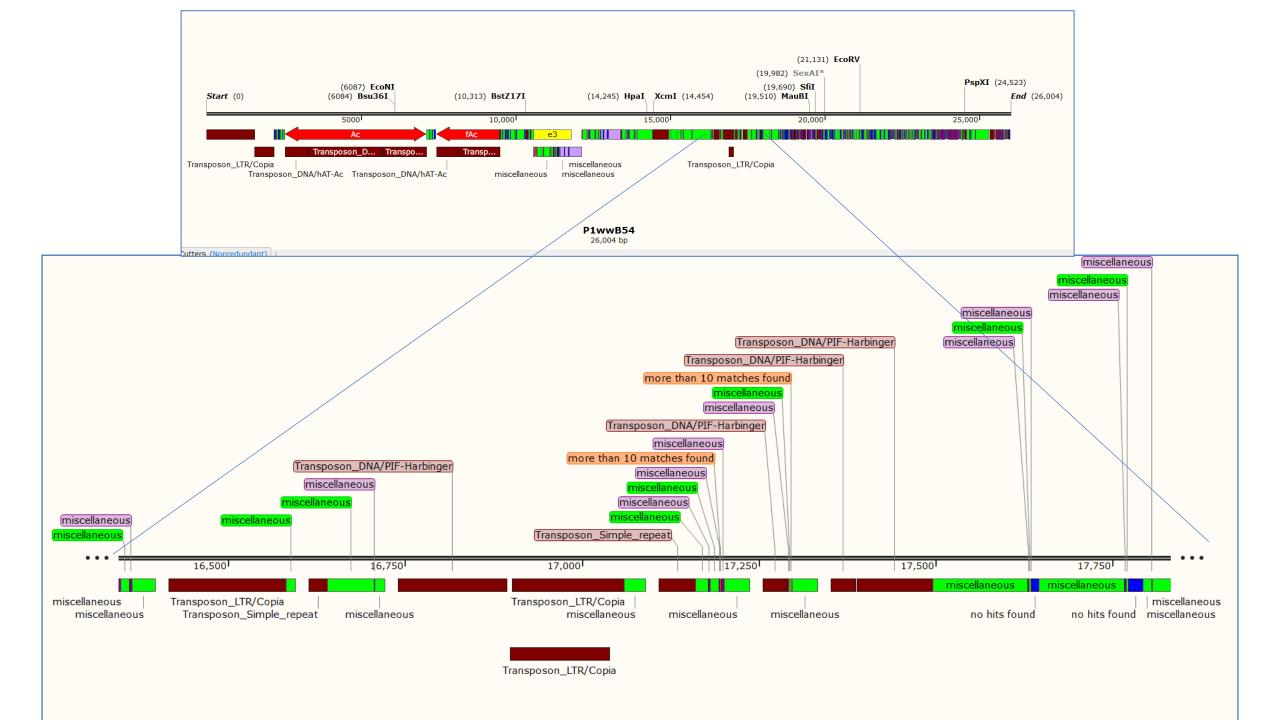


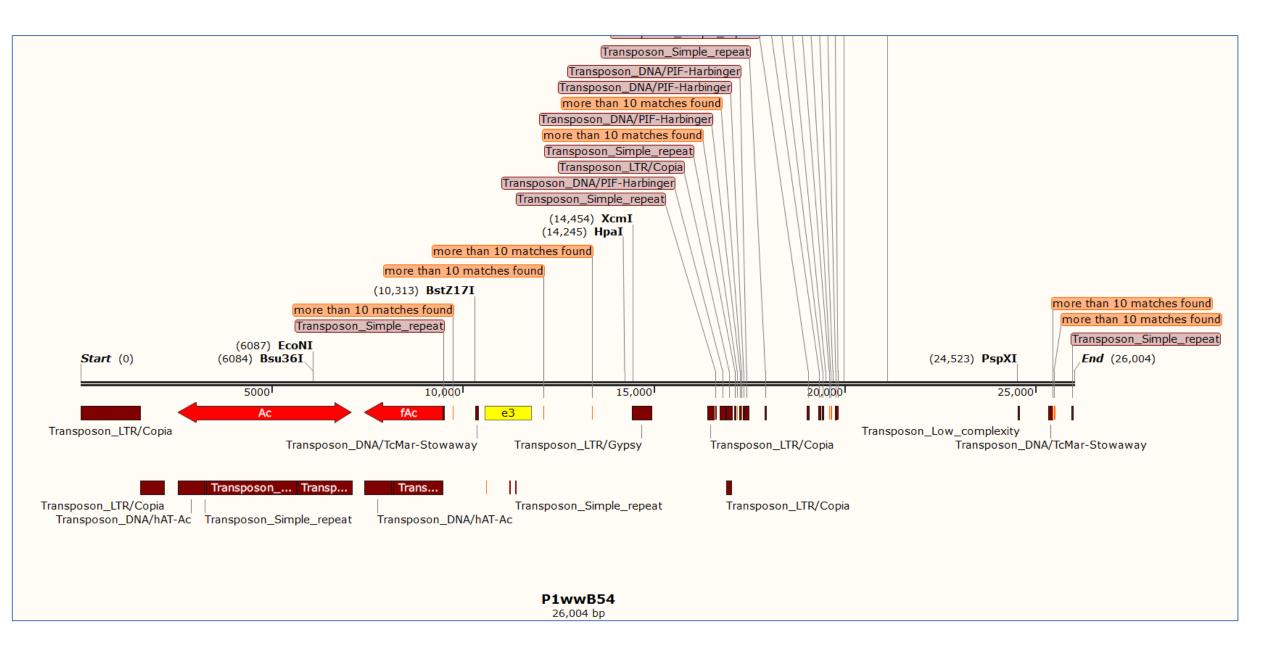
Dark Red= transposon Orange= more than 10 matches Bright Green = less than 10 matches, all matches in p cluster Blue= not matches in B73 Light purple = less than 10 matches, no matches in p cluster Dark purple = less than 10 matches, some in p cluster



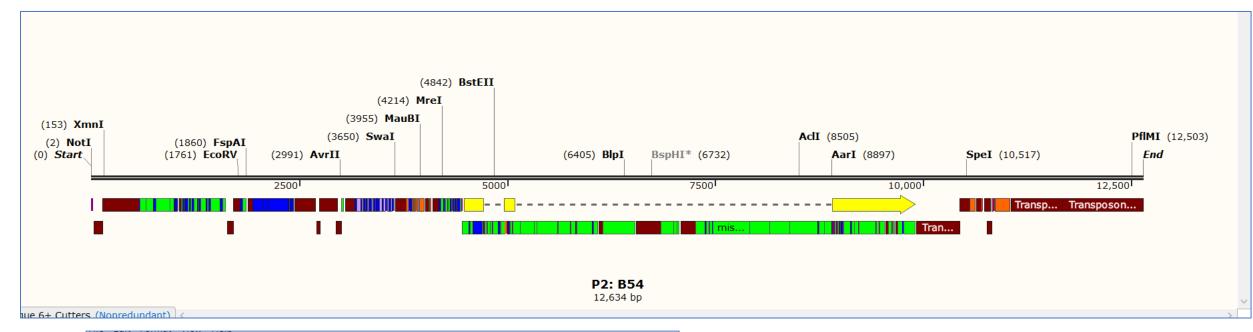


Dark Red= transposon Bright Green = less than 10 matches, all matches in p cluster Blue= no matches in B73 Light purple = less than 10 matches, no matches in p cluster





P2- wwB54



Summary Of Created Features For P2_wwB54
Input File: P2B54BlastResults.OUT
Transposon File: P2-B54repeatmaskerfile.txt

Created Features:
Number Of Transposons: 26
Number of Areas with no matches in B73: 57
Number of Areas with 10+ matches in B73: 18
Number of Regions with less than 10 matches: 162

Number of Regions with all matches found in P cluster of B73: 82
Number of Regions with no matches found in the P cluster of B73: 29
Number of Regions with some matches found in P cluster of B73: 51
Total Features Created: 263
Blast Results Analyzed: 7644

Dark Red= transposon

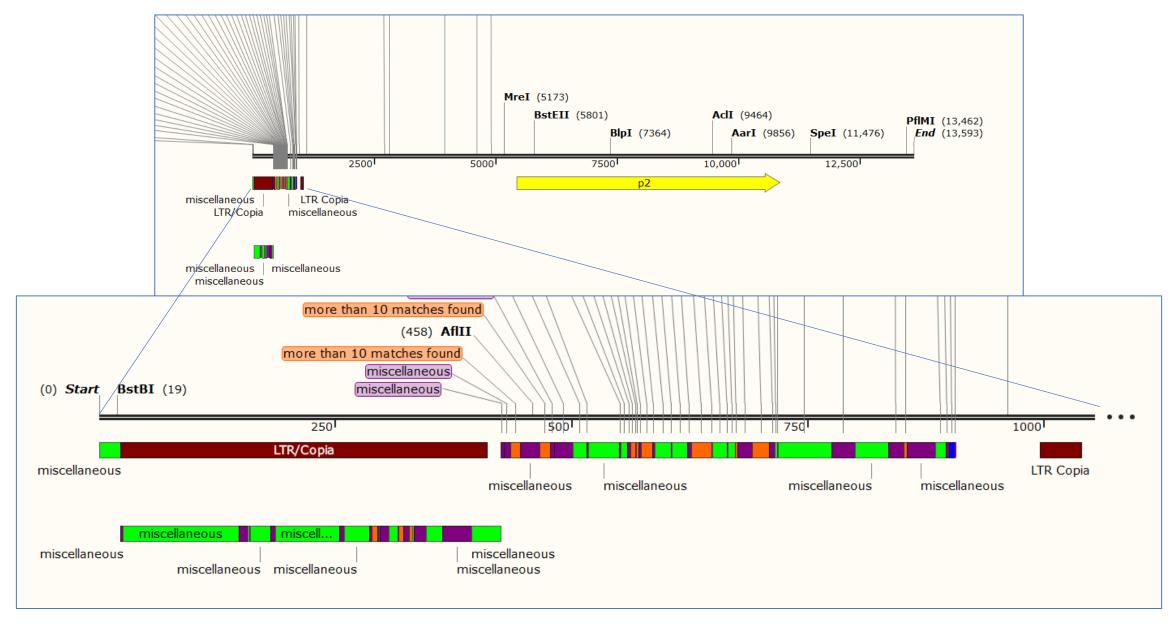
Orange= more than 10 matches

Bright Green = less than 10 matches, all matches in p cluster

Blue= no matches in B73

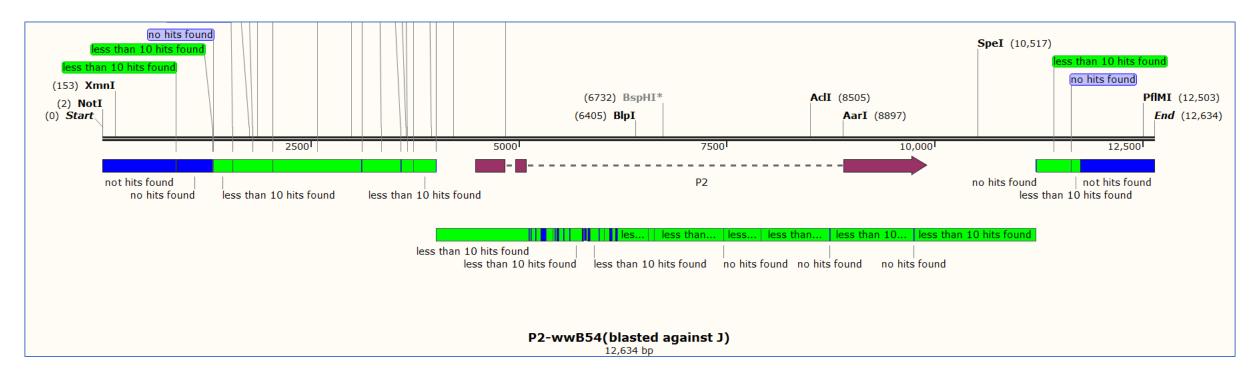
Light purple = less than 10 matches, no matches in p cluster

Dark purple = less than 10 matches, some in p cluster



Dark Red= transposon Bright Green = less than 10 matches, all matches in p cluster Blue= no matches in B73 Dark purple = less than 10 matches, some matches in p cluster Orange = more than 10 matches

J vs B54



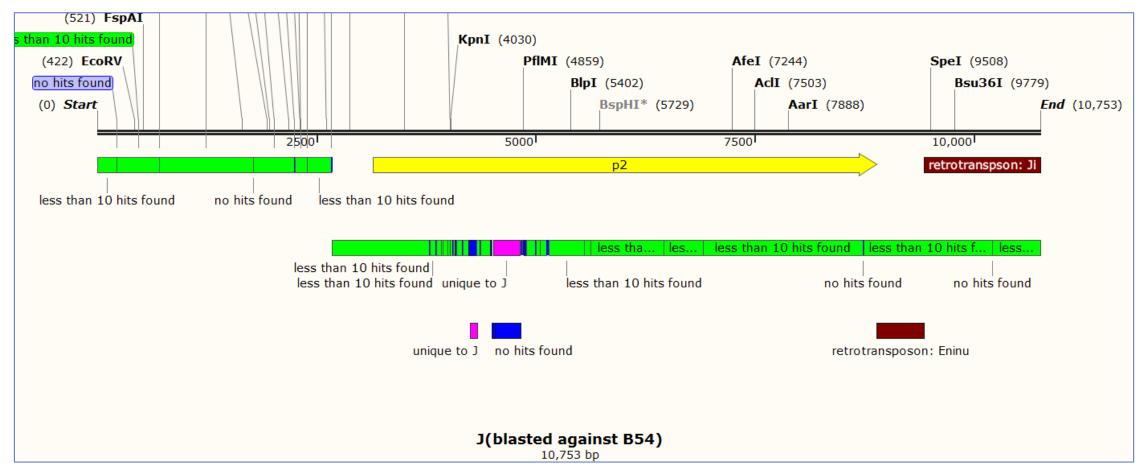
BLAST database: P2[4Co63] = P2 in J

Sequence: P2-wwB54

Green = less than 10 matches in p2 of J Blue= no matches in p2 of J

BLAST database: P2-wwB54

Sequence: P2[4Co63] = P2 in J



Green = less than 10 matches in p2 of J Blue= no matches in p2 of J Pink=regions unique to J