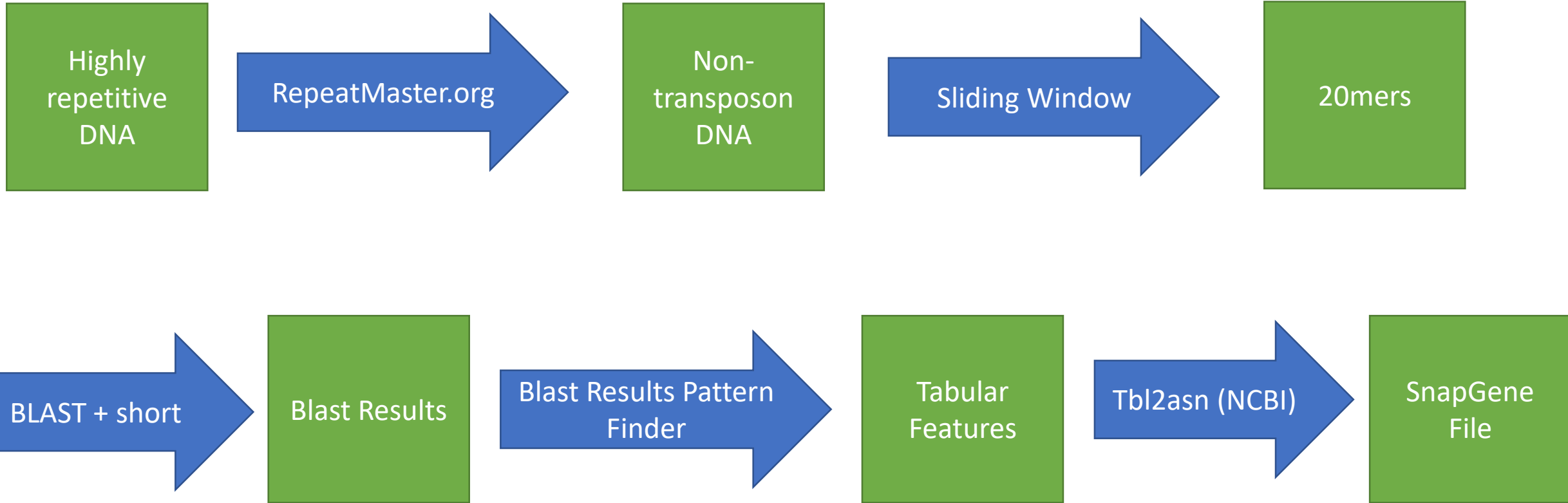


Automated Primer Region Analysis

Katherine Braught

Peterson Lab

Program Overview



Sliding Window

```
0  
1  
2 public class ItsAMAIZingFASTAformat {  
3 {  
4 final String INPUT_FILE = inputfiletxt; //must be a text file with only the bp letters  
5 final String MERS_OUTPUT_FILE = mersOutputfa; //used as query for BLAST search (.fa)  
6 final String FASTA_DNA_OUTPUT_FILE = fastaOutputfsa; //used later for tbl2asn (.fsa)  
7 final String SEQ_ID = seqID; //must match the sequence ID on fasta DNA file  
8 final int SEQ_START = seqStart; //starting location of the sequences  
9
```

Input: masked DNA file

Output: list of 20mers (FASTA)

```
> Sequence 3  
GGCCGCCCCGTTTCGCCCCG  
> Sequence 4  
GCCGCCCCGTTTCGCCCCG  
> Sequence 5  
CCGCCCCGTTTCGCCCCGCT  
> Sequence 6  
CGCCCCGTTTCGCCCCGCTC  
> Sequence 7  
GCCCCGTTTCGCCCCGCTCA  
> Sequence 8  
CCCCGTTTCGCCCCGCTCAG  
> Sequence 9  
CCCGTTTCGCCCCGCTCAGG  
> Sequence 10  
CCGTTTCGCCCCGCTCAGGG  
> Sequence 11  
CGTTTCGCCCCGCTCAGGGC  
> Sequence 12  
GTTTCGCCCCGCTCAGGGCC  
> Sequence 13  
TTCGCCCCGCTCAGGGCCG  
> Sequence 14  
TCGCCCCGCTCAGGGCCGG  
> Sequence 15  
CGCCCCGCTCAGGGCCGGC  
> Sequence 16  
GCCCCGCTCAGGGCCGGCG  
> Sequence 17  
CCCGCCCTCAGGGCCGGCGC  
> Sequence 582  
TGTTGTCTCTCTCTCAGA  
> Sequence 583  
GTTGTCTCTCTCTCAGAT  
> Sequence 584  
TTGTCTCTCTCTCAGATA  
> Sequence 585  
TGTCTCTCTCTCAGATAA  
> Sequence 586
```

Blast + short

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

Input: list of 20mers

Output: BLAST tabular results

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

```
>Features TZ3-15 matches_after_B73_Blast
2079 2132 misc_feature
      note less than 10 matches found
2133 2151 misc_feature
      note no matches found
2152 2223 misc_feature
      note less than 10 matches found
2224 2243 misc_feature
      note no matches found
2244 2252 misc_feature
      note less than 10 matches found
2253 2268 misc_feature
      note no matches found
2269 2286 misc_feature
      note less than 10 matches found
2287 2304 misc_feature
      note no matches found
2305 2308 misc_feature
      note more than 10 matches found
2309 2332 misc_feature
      note no matches found
2333 2380 misc_feature
      note less than 10 matches found
2381 2398 misc_feature
      note no matches found
2399 2566 misc_feature
      note less than 10 matches found
2567 2602 misc_feature
      note no matches found
2603 2828 misc_feature
      note less than 10 matches found
2829 2856 misc_feature
      note no matches found
2857 2875 misc_feature
      note less than 10 matches found
2980 3032 misc_feature
      note less than 10 matches found
3033 3052 misc_feature
      note no matches found
3053 3179 misc_feature
      note less than 10 matches found
3180 3187 misc_feature
      note more than 10 matches found
```

```
1 1555 misc_feature
      note less than 10 found
      note Transposon_LTR/Copia
1551 2182 misc_feature
      note Transposon_LTR/Copia
2537 3229 misc_feature
      note Transposon_DNA/hAT-Ac
3230 3290 misc_feature
      note Transposon_Simple_repeat
3291 5636 misc_feature
      note Transposon_DNA/hAT-Ac
5637 5688 misc_feature
      note Transposon_Simple_repeat
5689 7105 misc_feature
      note Transposon_DNA/hAT-Ac
7436 8125 misc_feature
      note Transposon_DNA/hAT-Ac
8126 8186 misc_feature
      note Transposon_Simple_repeat
8187 9480 misc_feature
      note Transposon_DNA/hAT-Ac
9481 9513 misc_feature
      note Transposon_Simple_repeat
10331 10414 misc_feature
      note Transposon_DNA/TcMar-Stowaway
11229 11249 misc_feature
      note Transposon_Simple_repeat
11374 11400 misc_feature
      note Transposon_Simple_repeat
14438 14946 misc_feature
      note Transposon_LTR/Gypsy
16416 16581 misc_feature
      note Transposon_LTR/Copia
16614 16639 misc_feature
      note Transposon_Simple_repeat
16740 16893 misc_feature
      note Transposon_DNA/PIF-Harbinger
16899 17038 misc_feature
      note Transposon_LTR/Copia
16901 17059 misc_feature
      note Transposon_LTR/Copia
17109 17160 misc_feature
      note Transposon_Simple_repeat
```

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

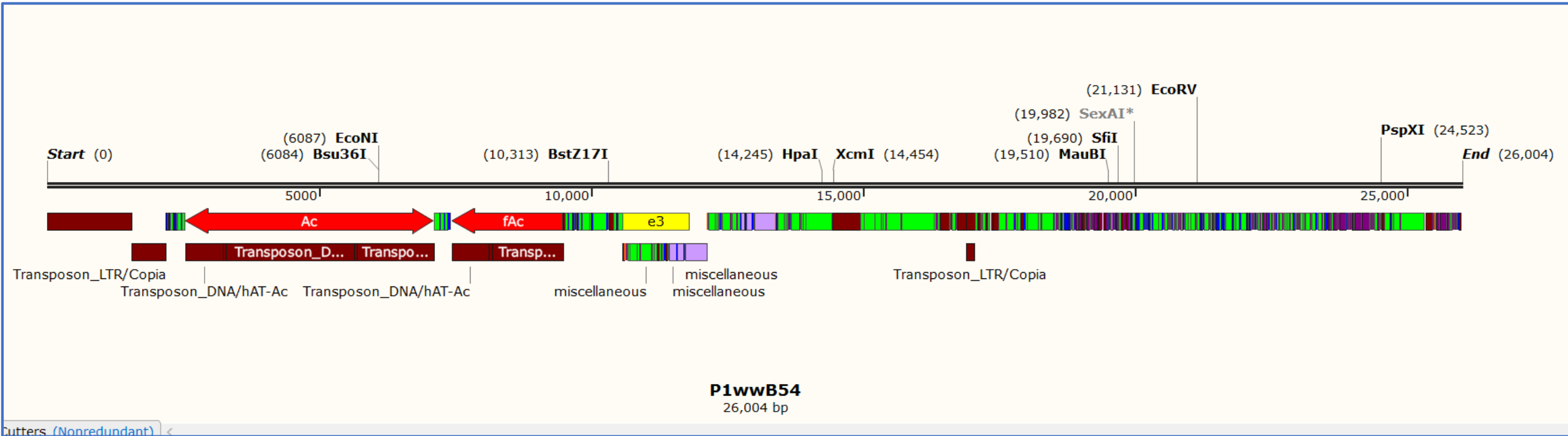
-p . = directory

-j = source qualifier

-V vb =generate GenBank file

```
LOCUS      P1_wwB54                26004 bp    DNA     linear   31-MAR-2017
DEFINITION P1-wwB54.
ACCESSION
VERSION
KEYWORDS   .
SOURCE     P1-wwB54
  ORGANISM P1-wwB54
            Unclassified.
REFERENCE  1 (bases 1 to 26004)
  AUTHORS  Test seq,T.S.
  TITLE    test seq
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 26004)
  AUTHORS  Test seq,T.S.
  TITLE    Direct Submission
  JOURNAL  Submitted (31-MAR-2017) test, Test, test, Ames, IA 50012, United
            States of America
FEATURES   Location/Qualifiers
     source          1..26004
                    /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"
     misc_feature    2205..2224
                    /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

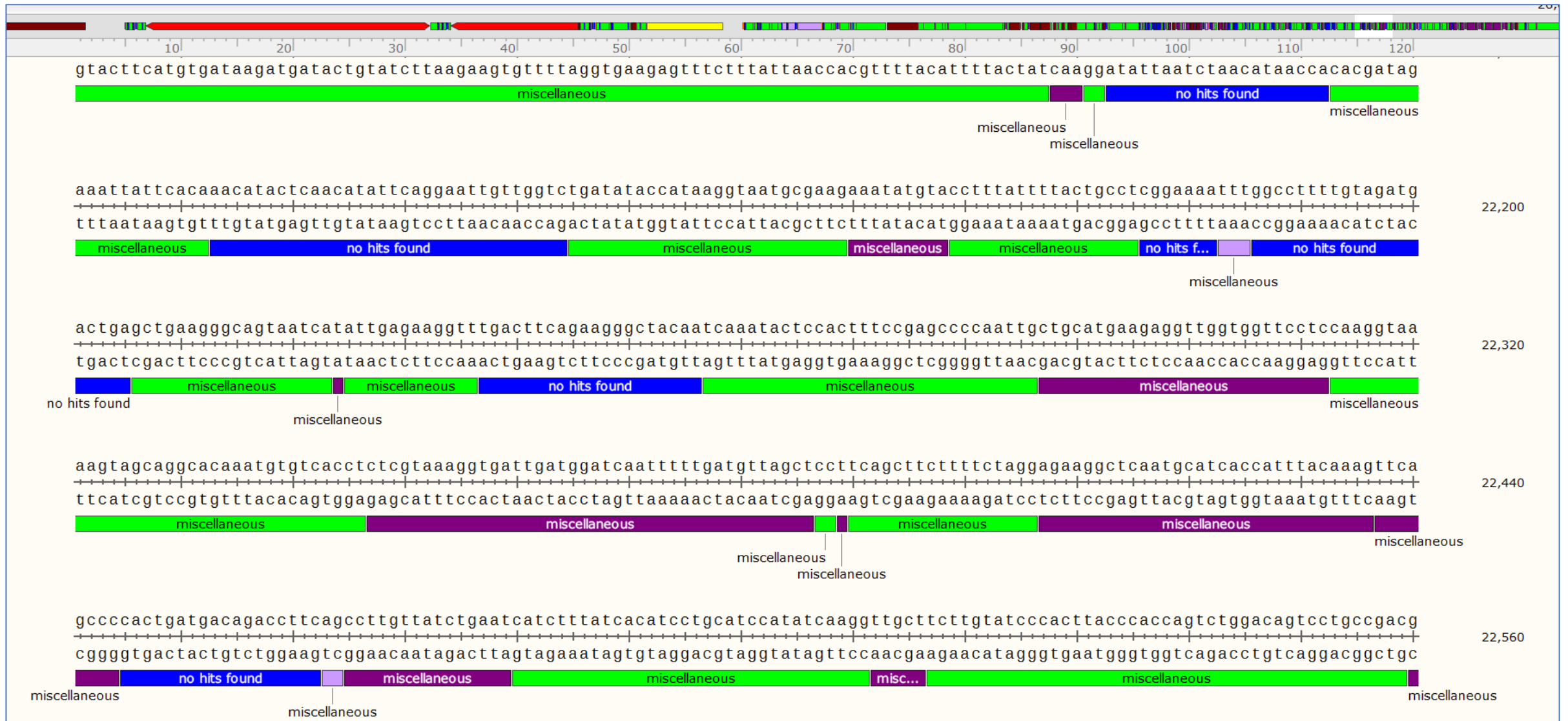
Orange= more than 10 matches

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

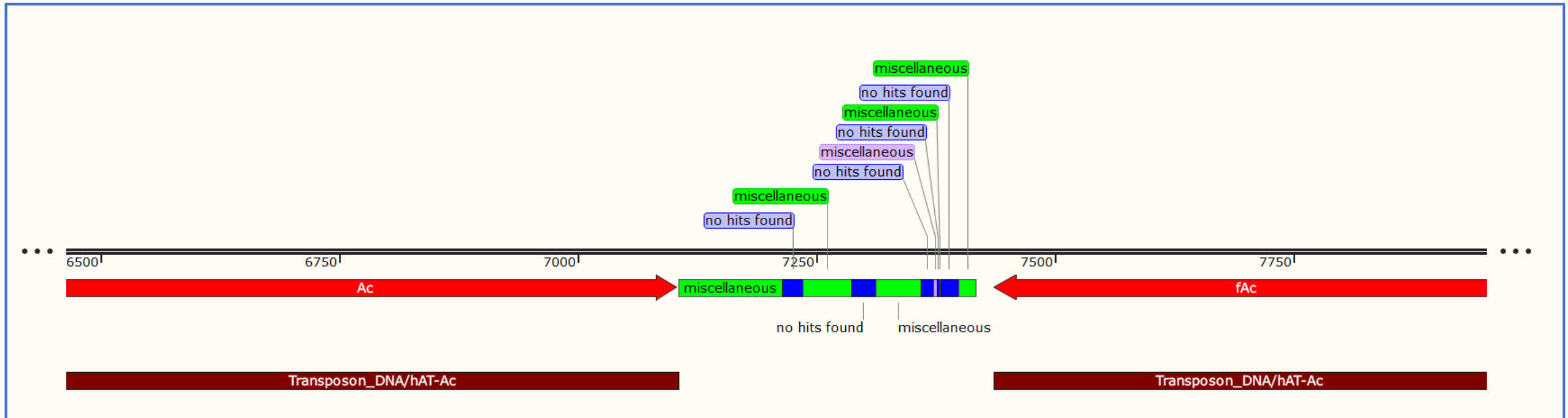
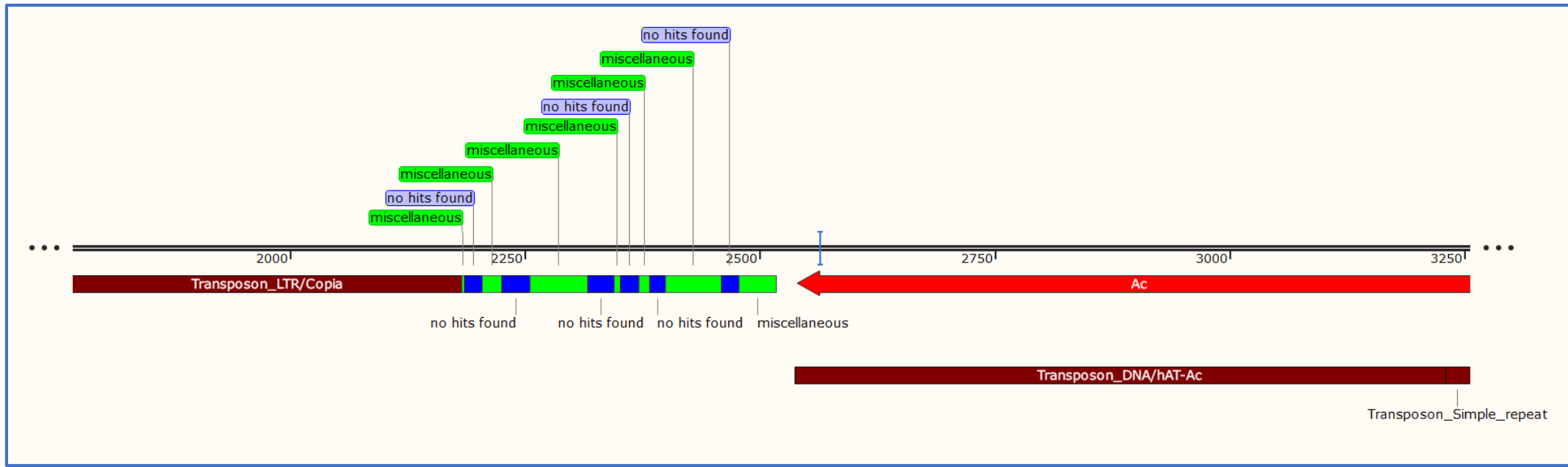
Blue= not matches in B73

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

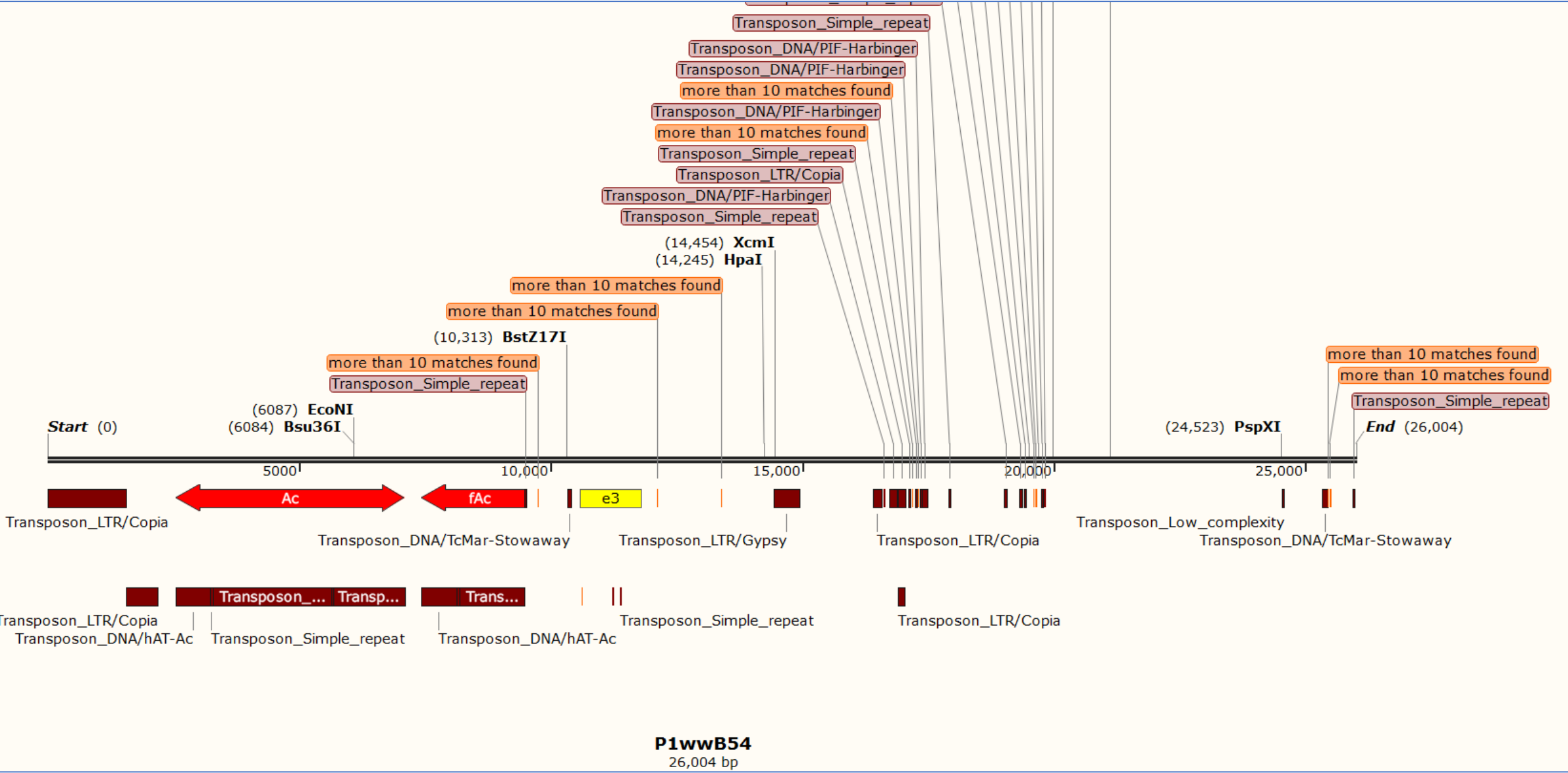
Dark purple = less than 10 matches, some in p cluster



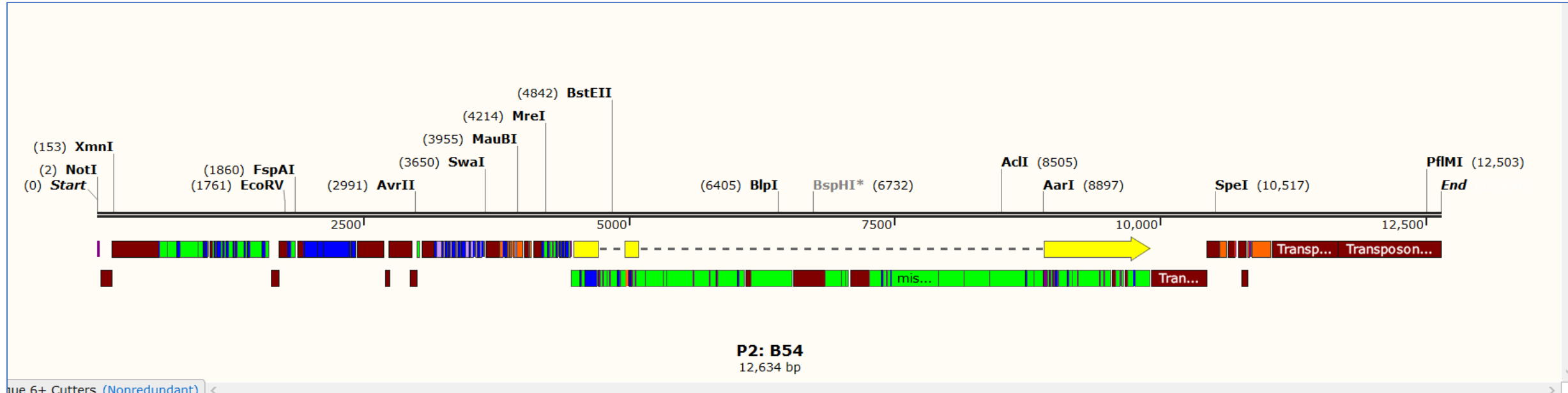
Dark Red= transposon Orange= more than 10 matches Bright Green = less than 10 matches, all matches in p cluster Blue= not matches in B73 Light purple = less than 10 matches, no matches in p cluster Dark purple = less than 10 matches, some in p cluster



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

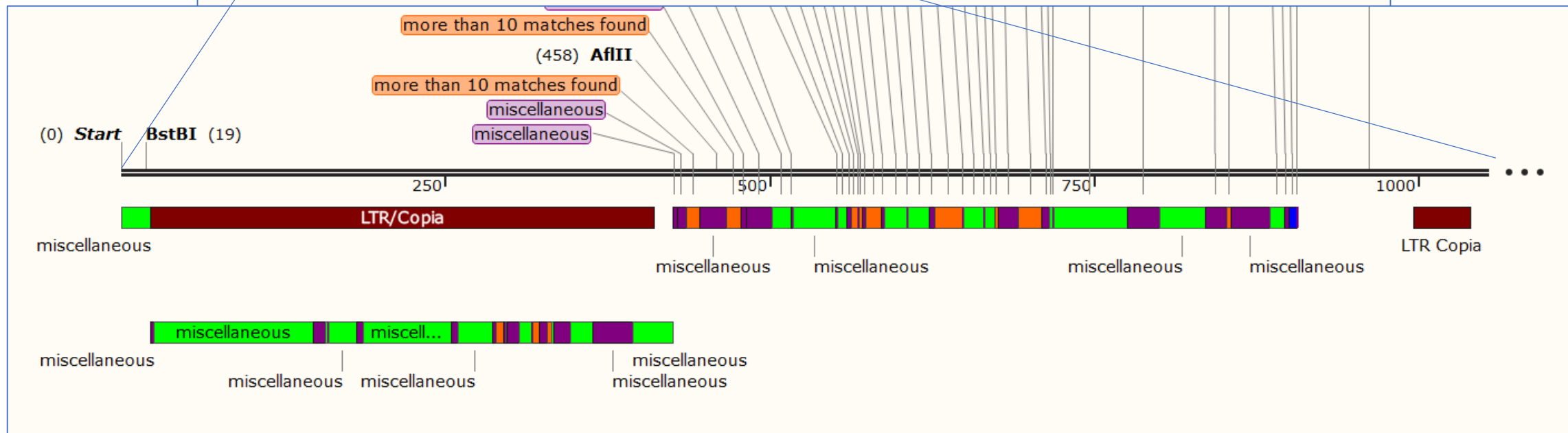
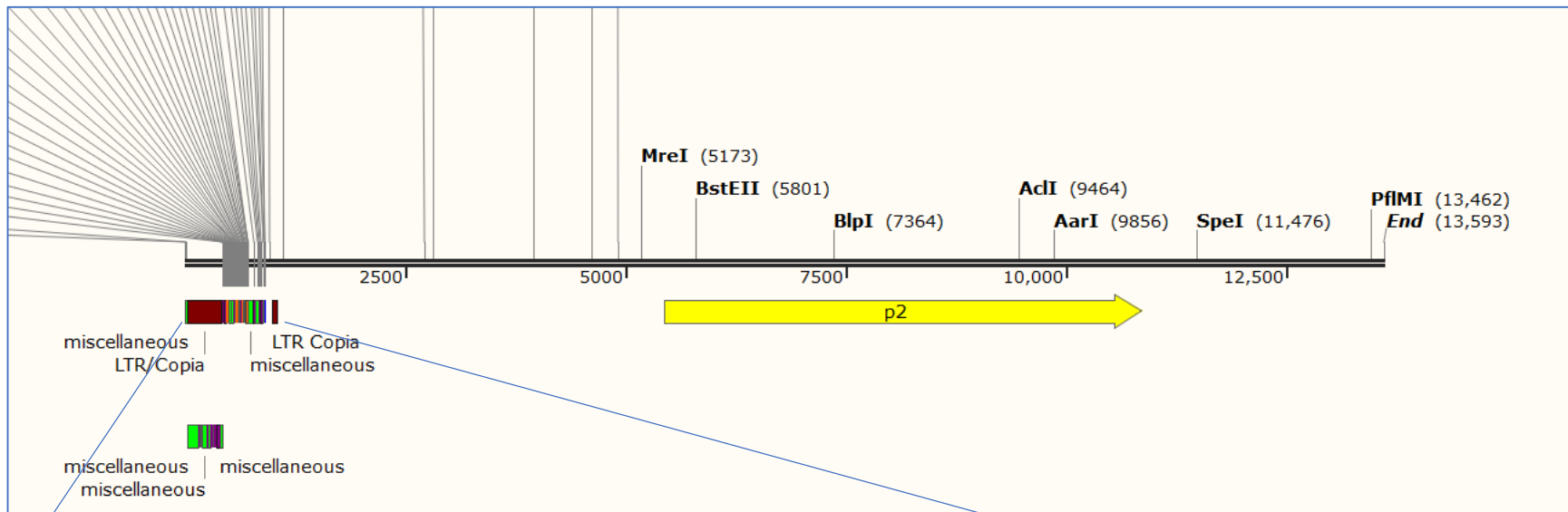
Orange= more than 10 matches

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

Blue= no matches in B73

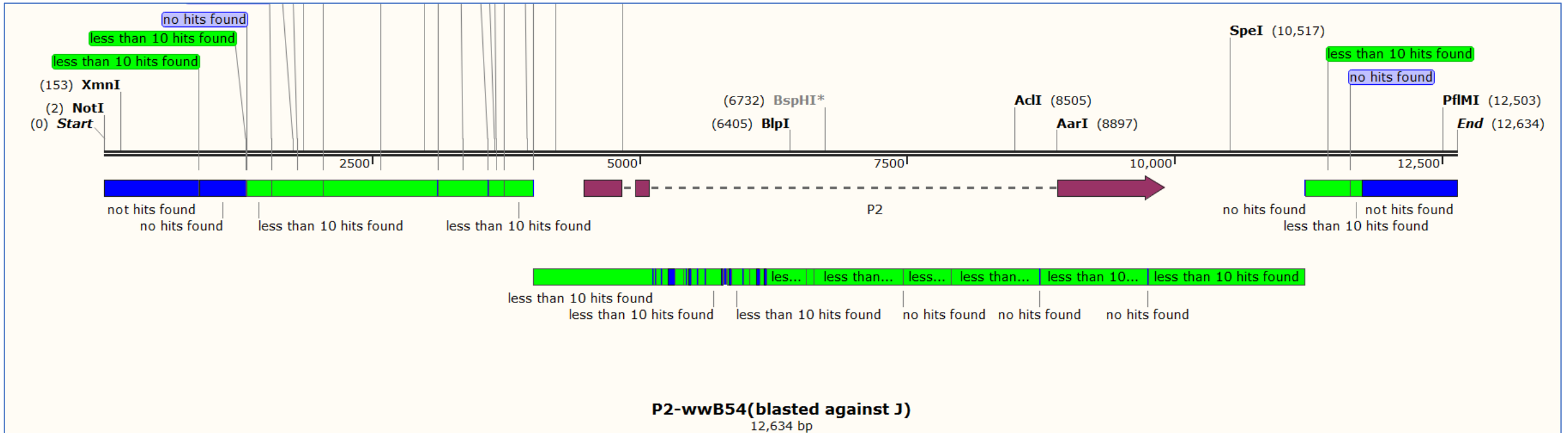
Light purple = less than 10 matches, no matches in p cluster

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



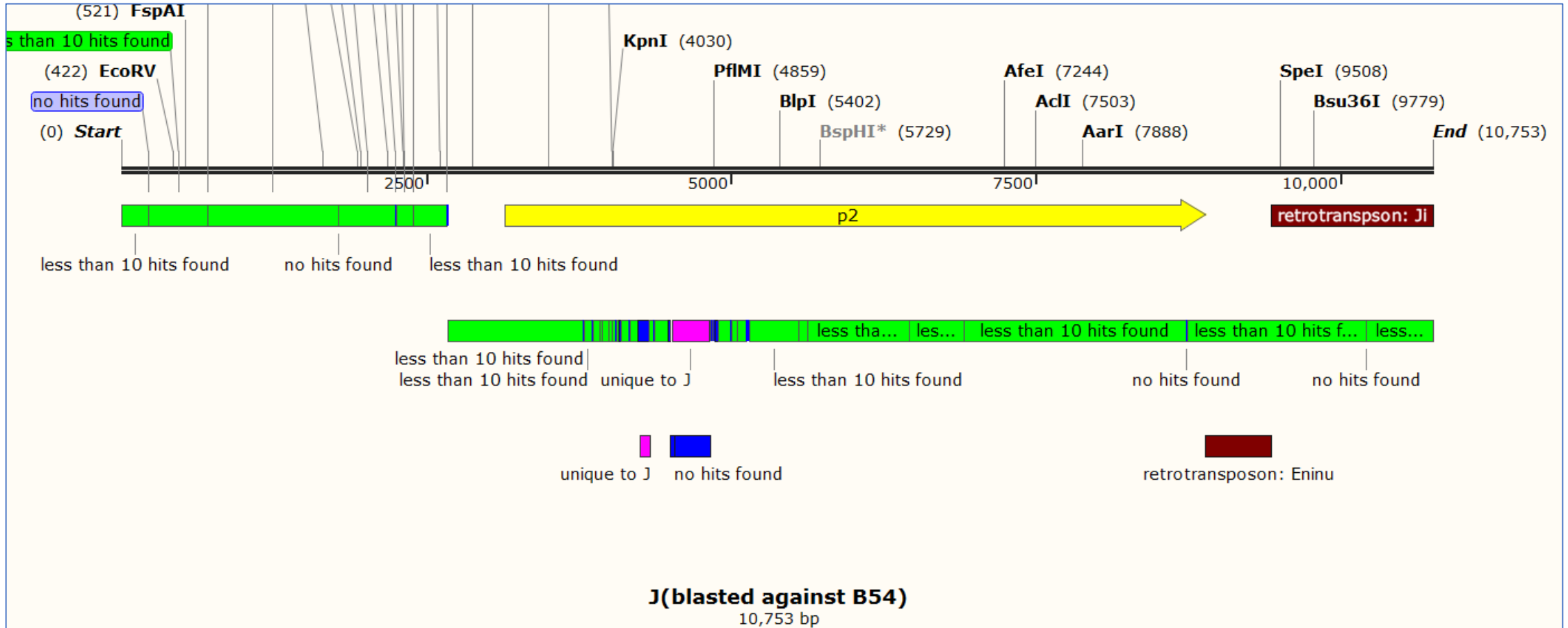
BLAST database: P2[4Co63] = P2 in J

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