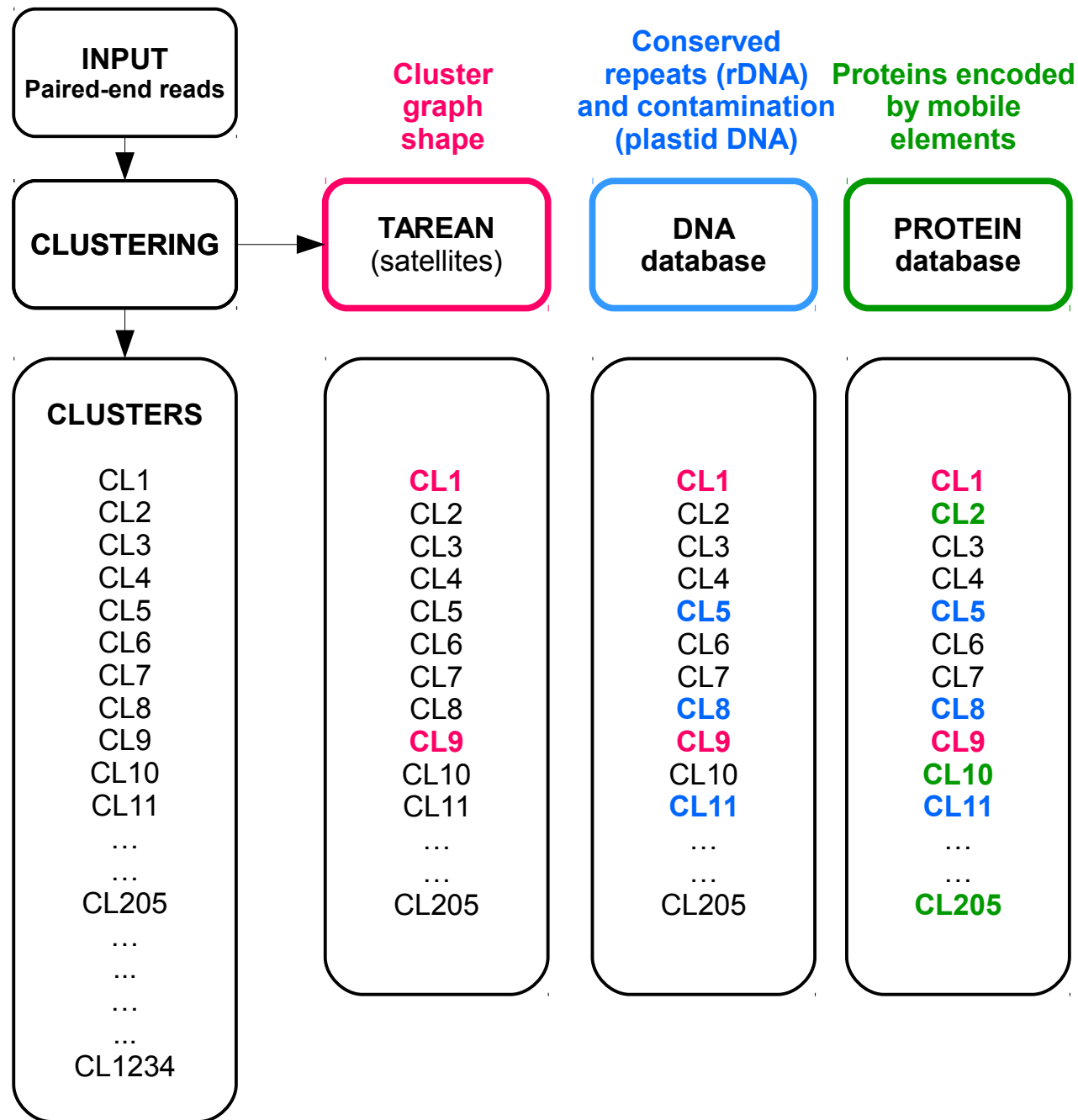
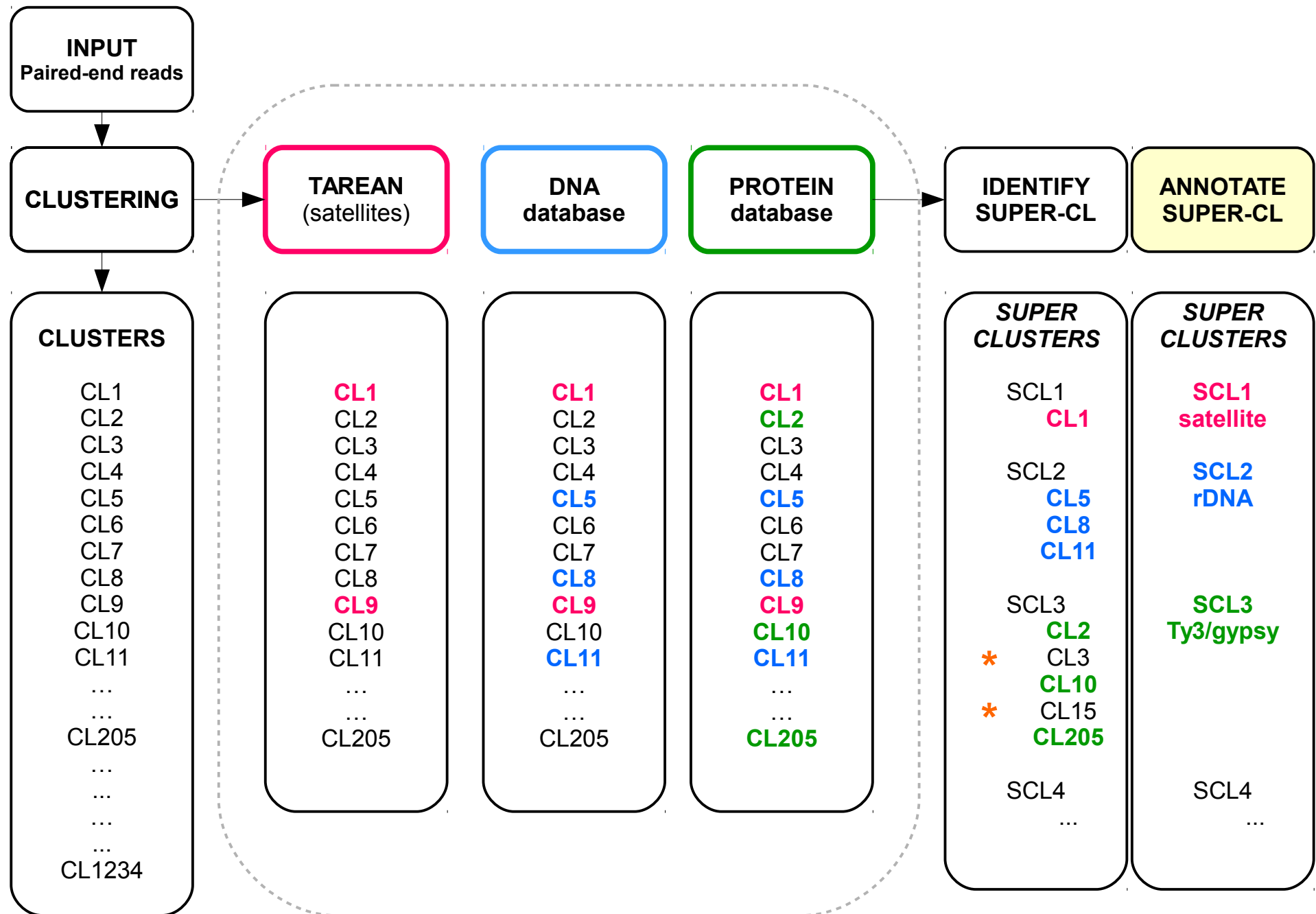


Using *RepeatExplorer* output for repeat
annotation and quantification

RepeatExplorer ver. 2



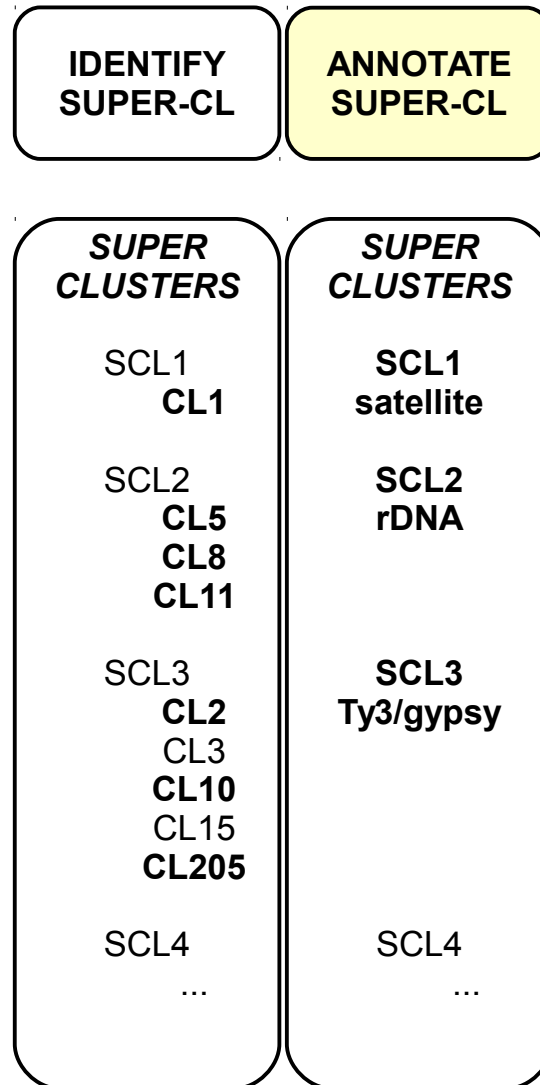
RepeatExplorer ver. 2



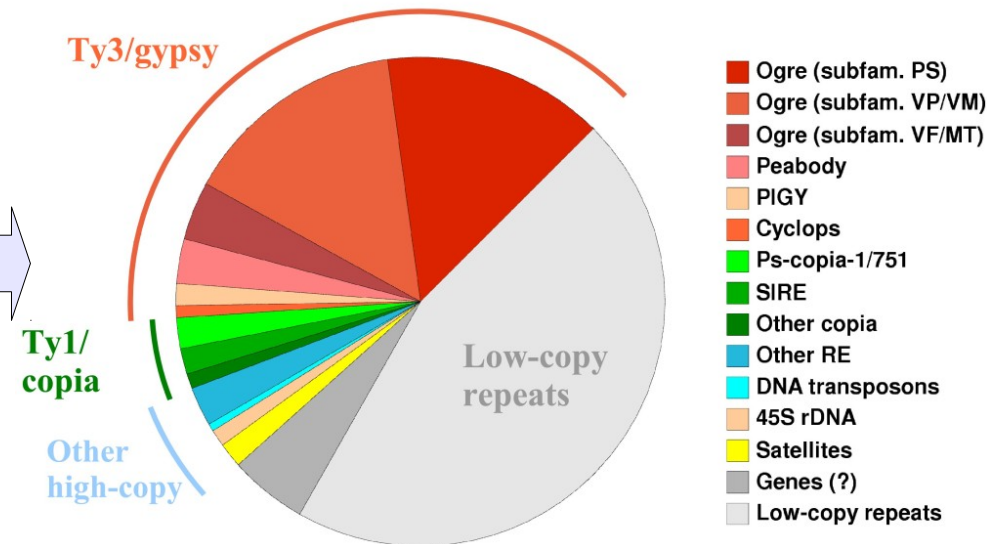
Repeat quantification

Cluster sizes (numbers of reads)
are proportional to genomic
abundance of corresponding repeats

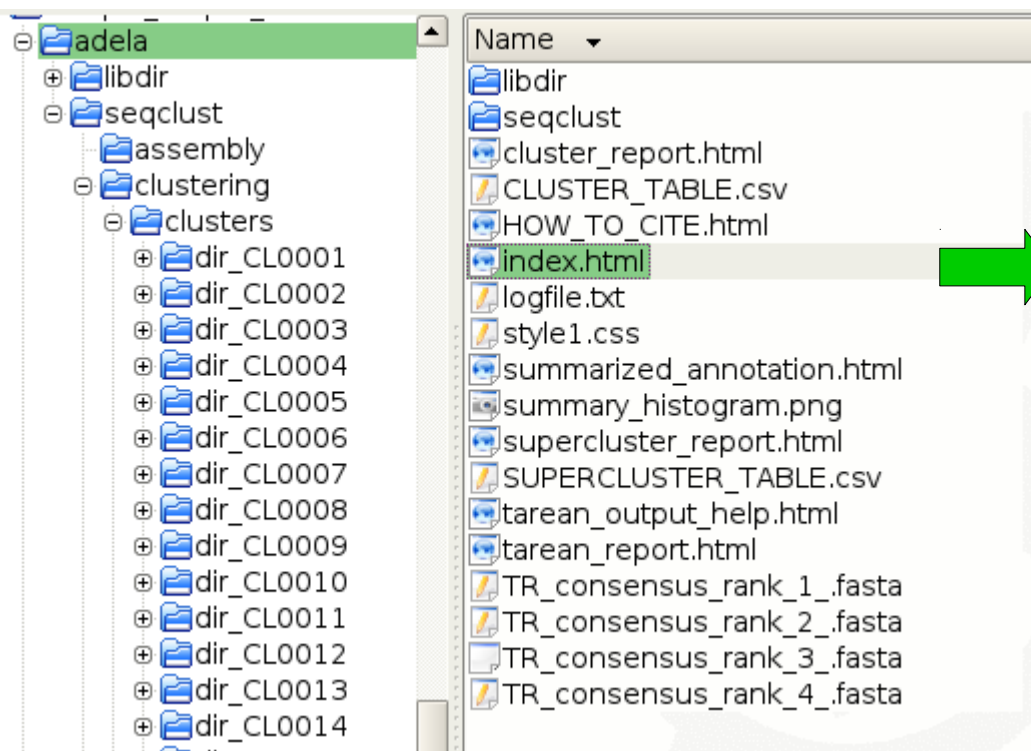
	7192836 (= 100%)	64.1
CL	reads	genome %
1	304159	4.229
2	234749	3.264
3	216307	3.007
4	202822	2.820
5	149693	2.081
6	145911	2.029
7	143766	1.999
8	142608	1.983
9	141836	1.972
10	123886	1.722
11	79345	1.103
12	72781	1.012
13	67096	0.933
14	65455	0.910
15	62334	0.867
16	53845	0.749
17	49341	0.686
18	45062	0.626
19	44762	0.622
20	43332	0.602
21	42344	0.589
22	40125	0.558
23	39923	0.555
24	36353	0.505
25	35977	0.500
26	35674	0.496
27	34829	0.484
28	34534	0.480
29	34302	0.477
30	33114	0.460
31	32930	0.458



Proportions of various repeat types in a genome

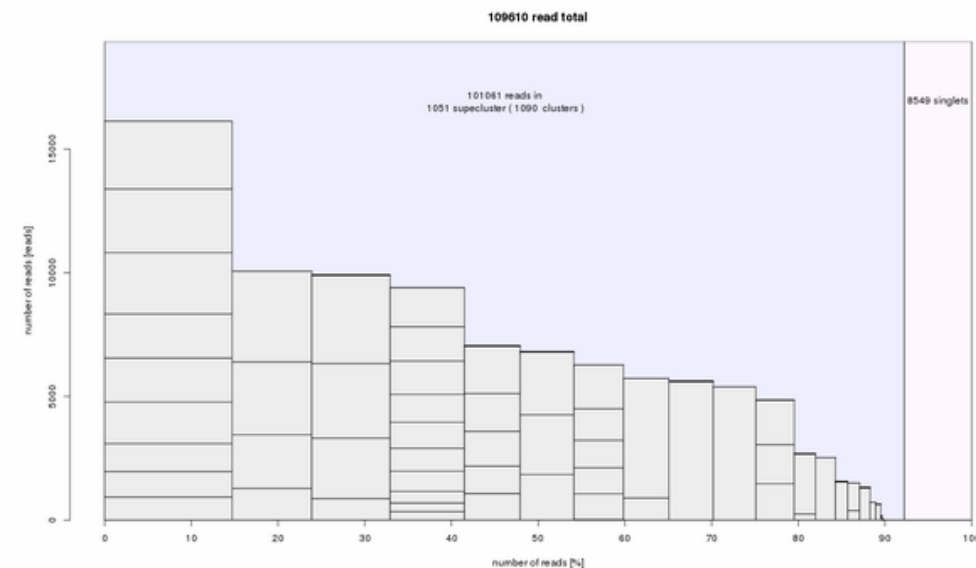


RepeatExplorer output files



Sequence Clustering Summary

PIPELINE VERSION : domains-v0.2.1-1795(8fff502)
PROTEIN DATABASE VERSION : protein_database_v2.2.fasta



Run statistics:

Number of input sequences: 109610

Number of analyzed sequences: 109610

Proportion of sequences in analyzed clusters : 90 %

Cluster merging: No

Paired end sequences: Yes

Available analyses:

[Tandem repeat analysis](#)

[Cluster annotation](#)

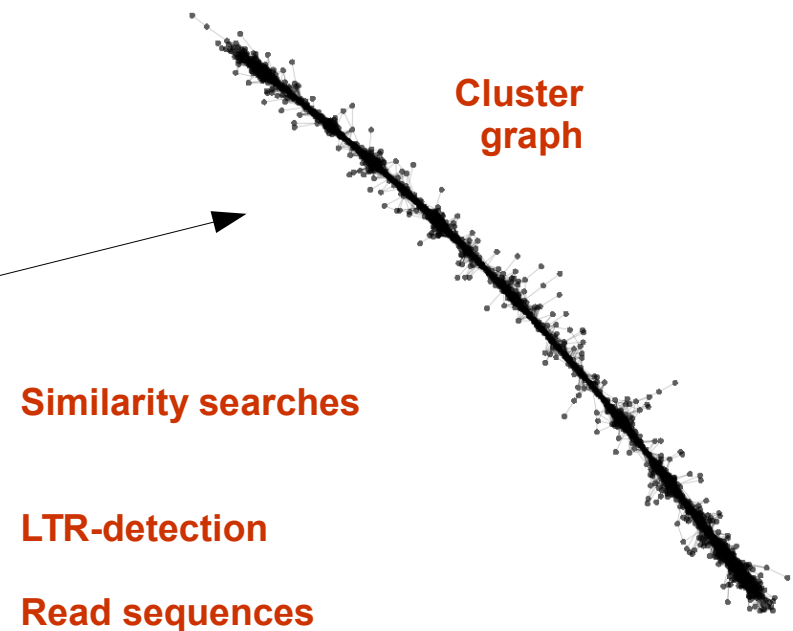
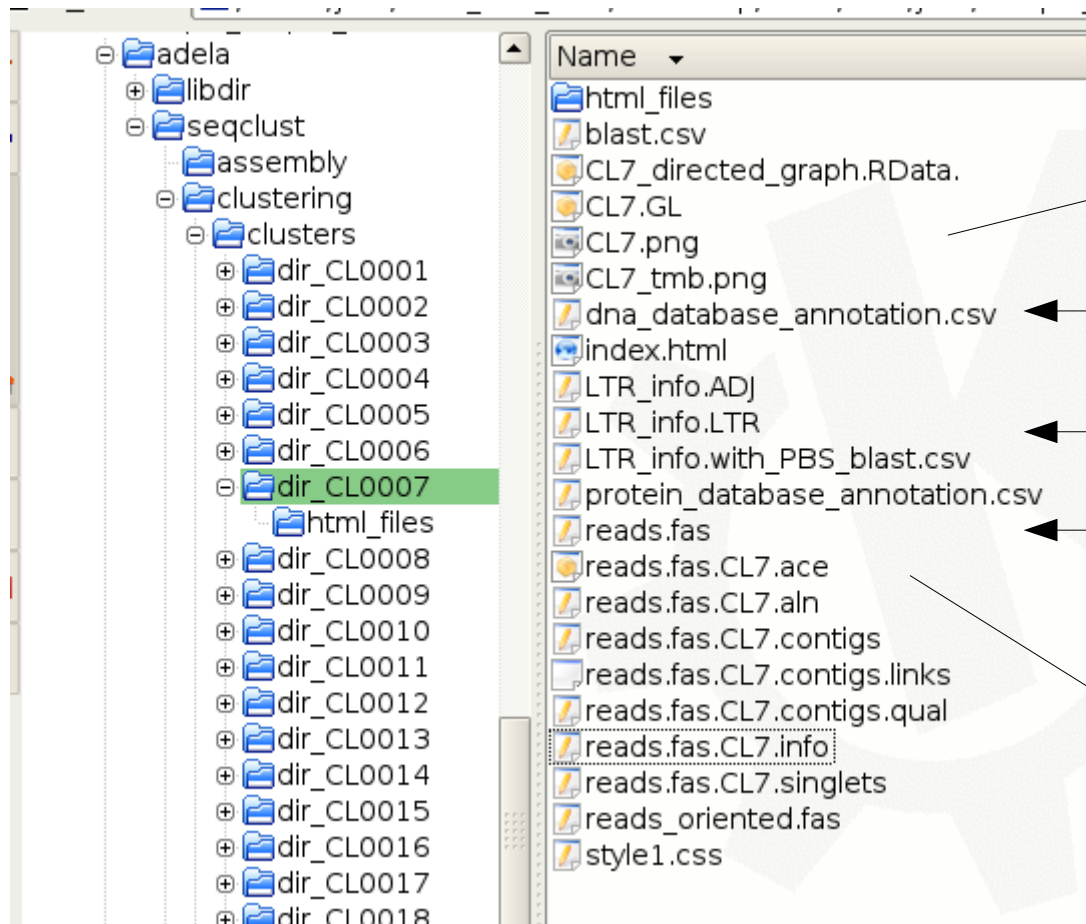
[Supercluster annotation](#)

[Repeat annotation summary](#)

Always download and use archive
with complete output of the analysis
(do not work with HTML only)

RepeatExplorer output files

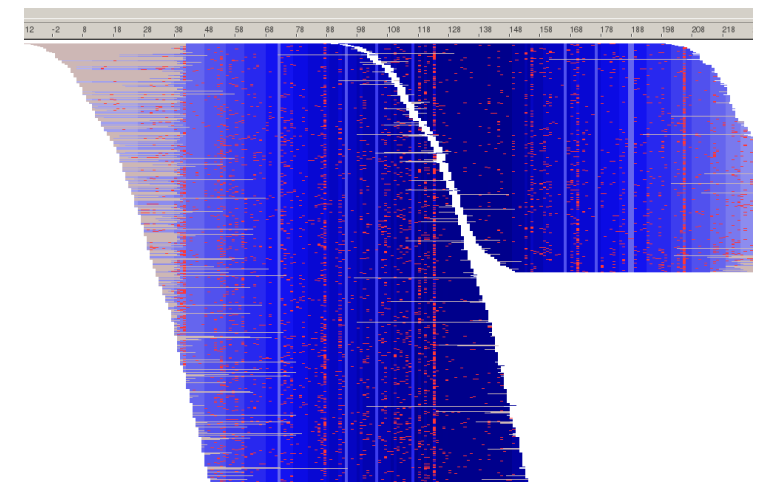
Cluster directories



Similarity searches

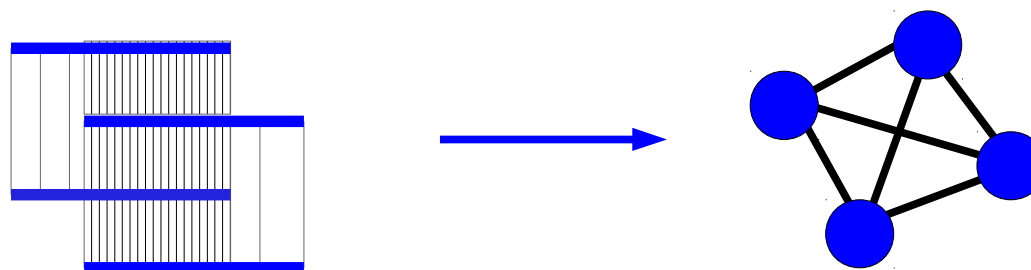
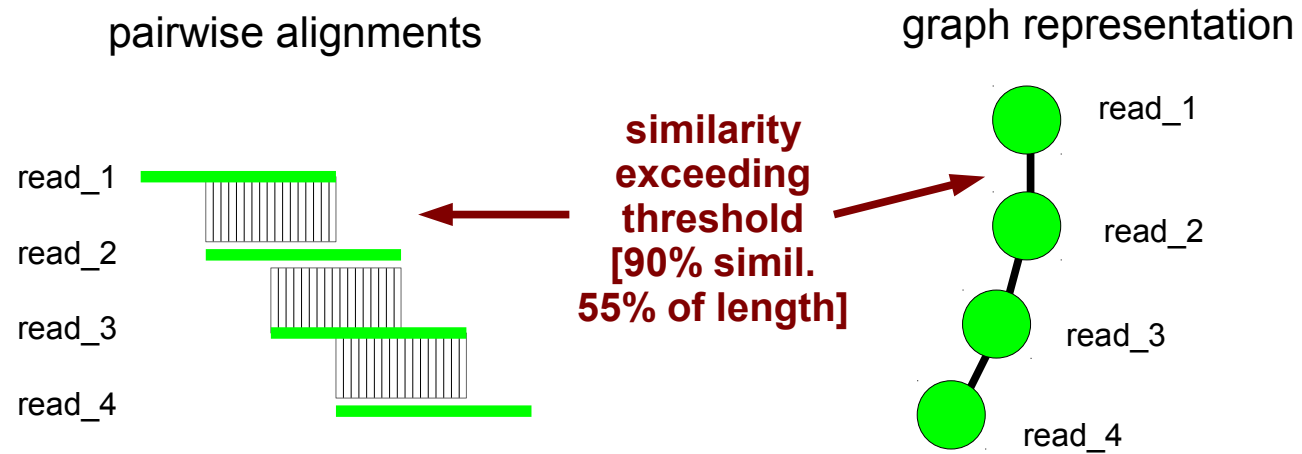
LTR-detection

Read sequences

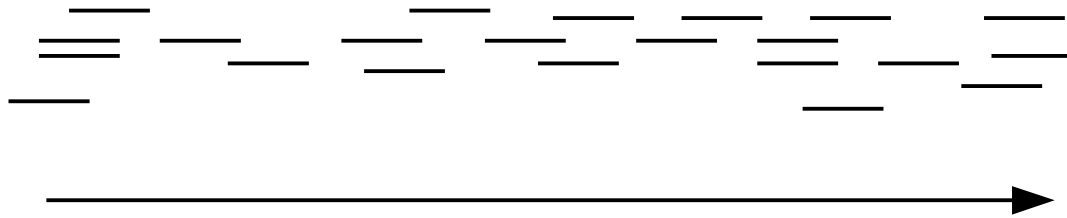


assembly

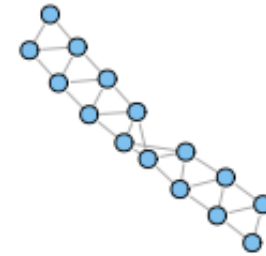
Graph shapes



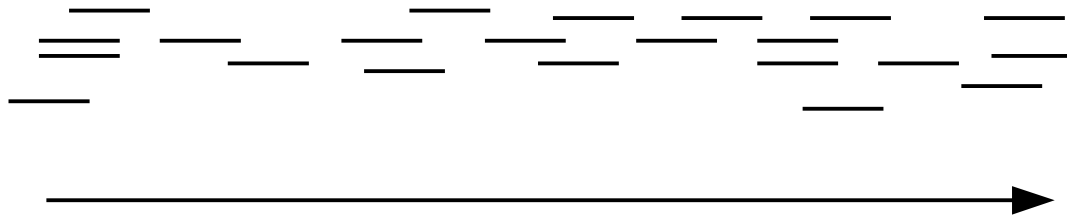
Linear graphs



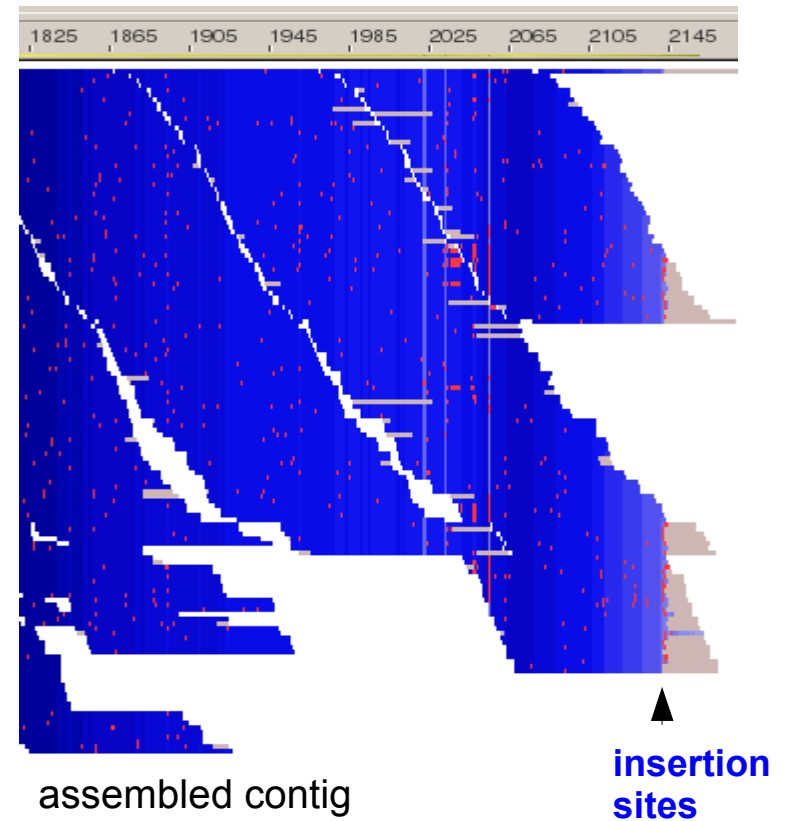
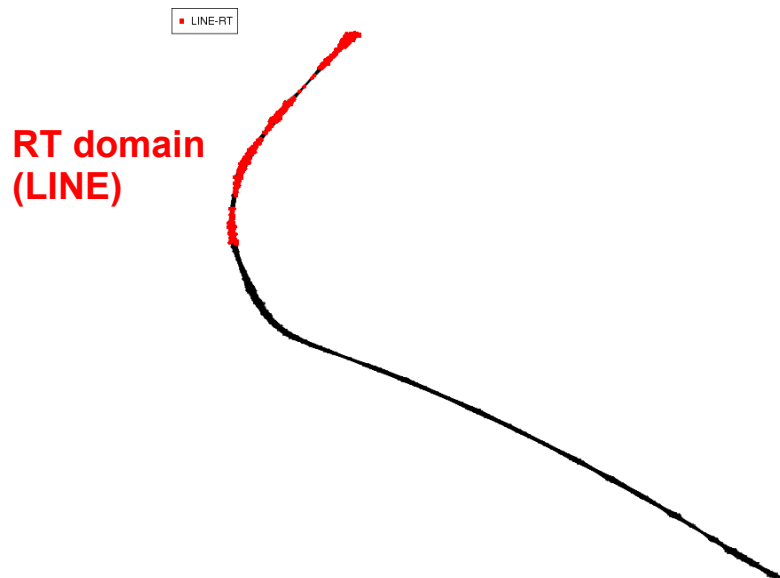
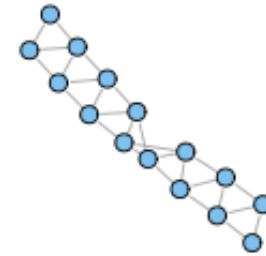
DNA transposons, LINEs, ...



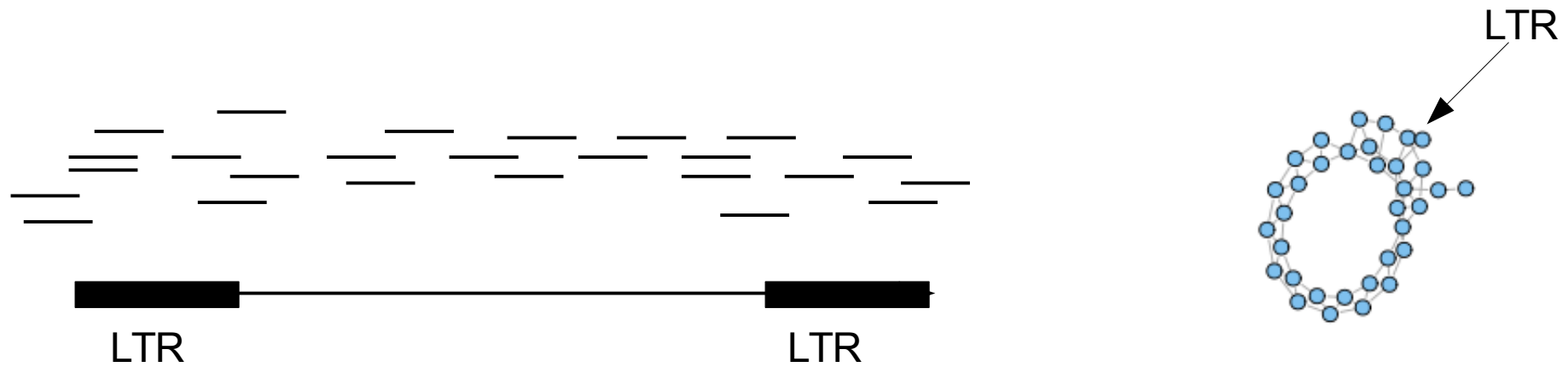
Linear graphs



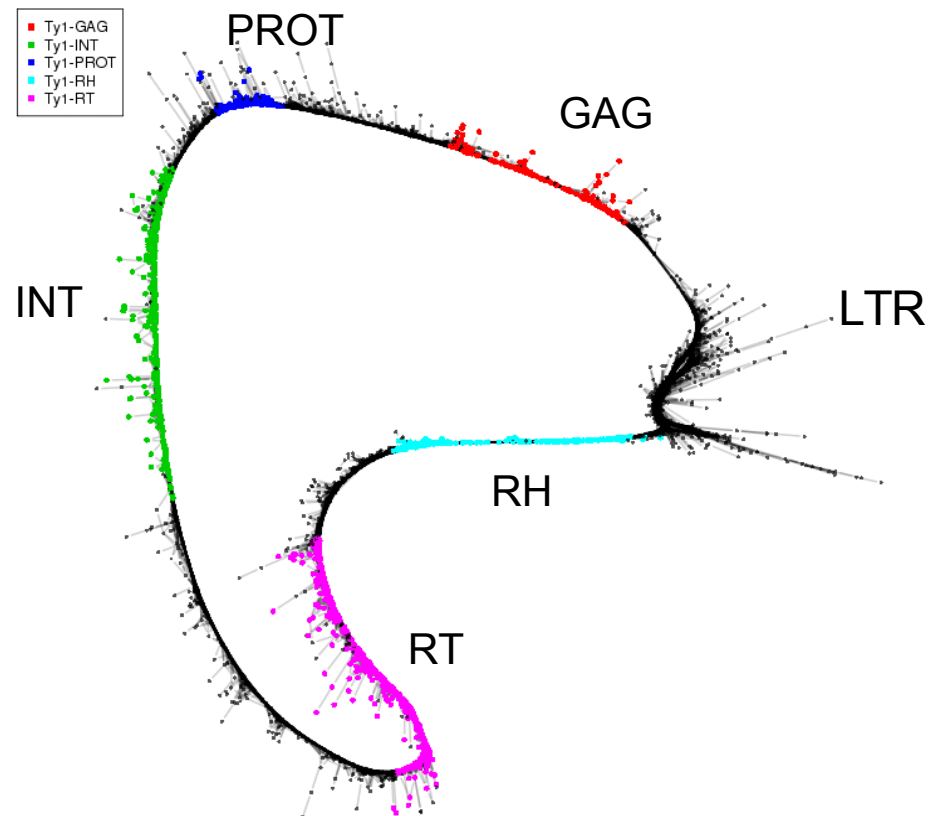
DNA transposons, LINEs, ...



Linear / large circular graphs

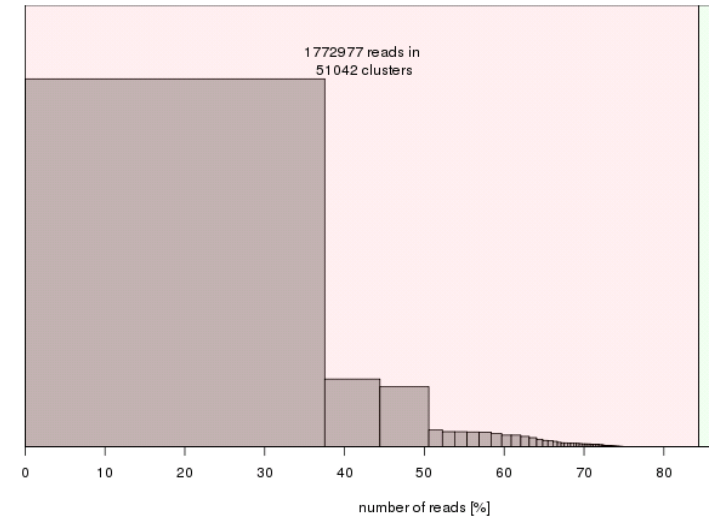
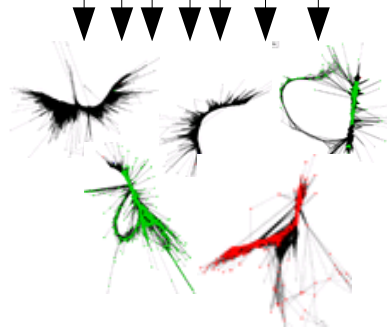
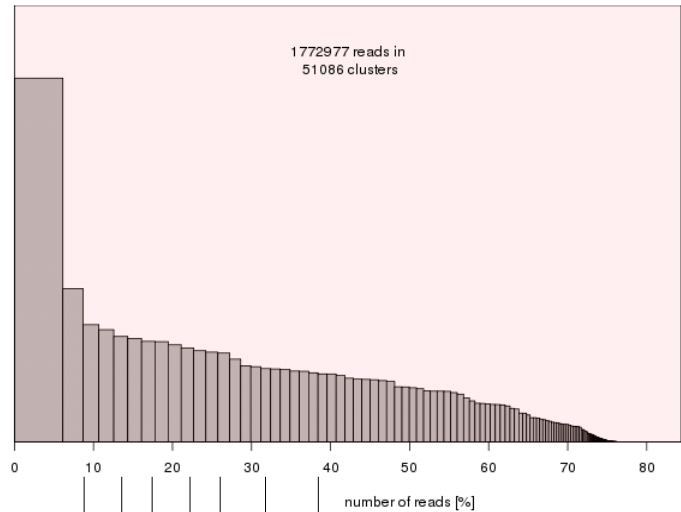


LTR-retrotransposons

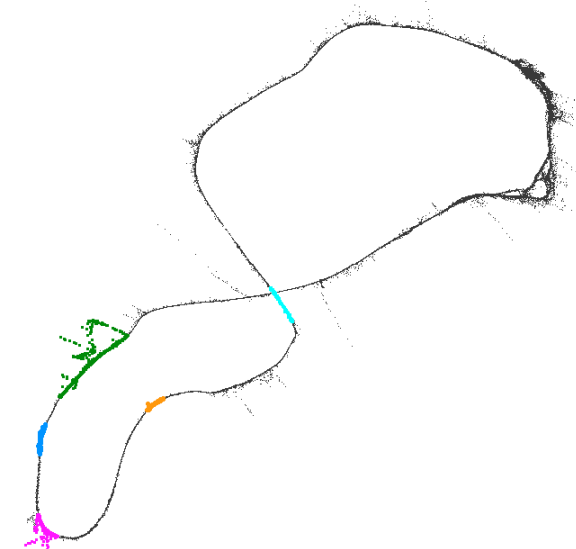


Cluster fragmentation

Repeats can be split to multiple clusters !



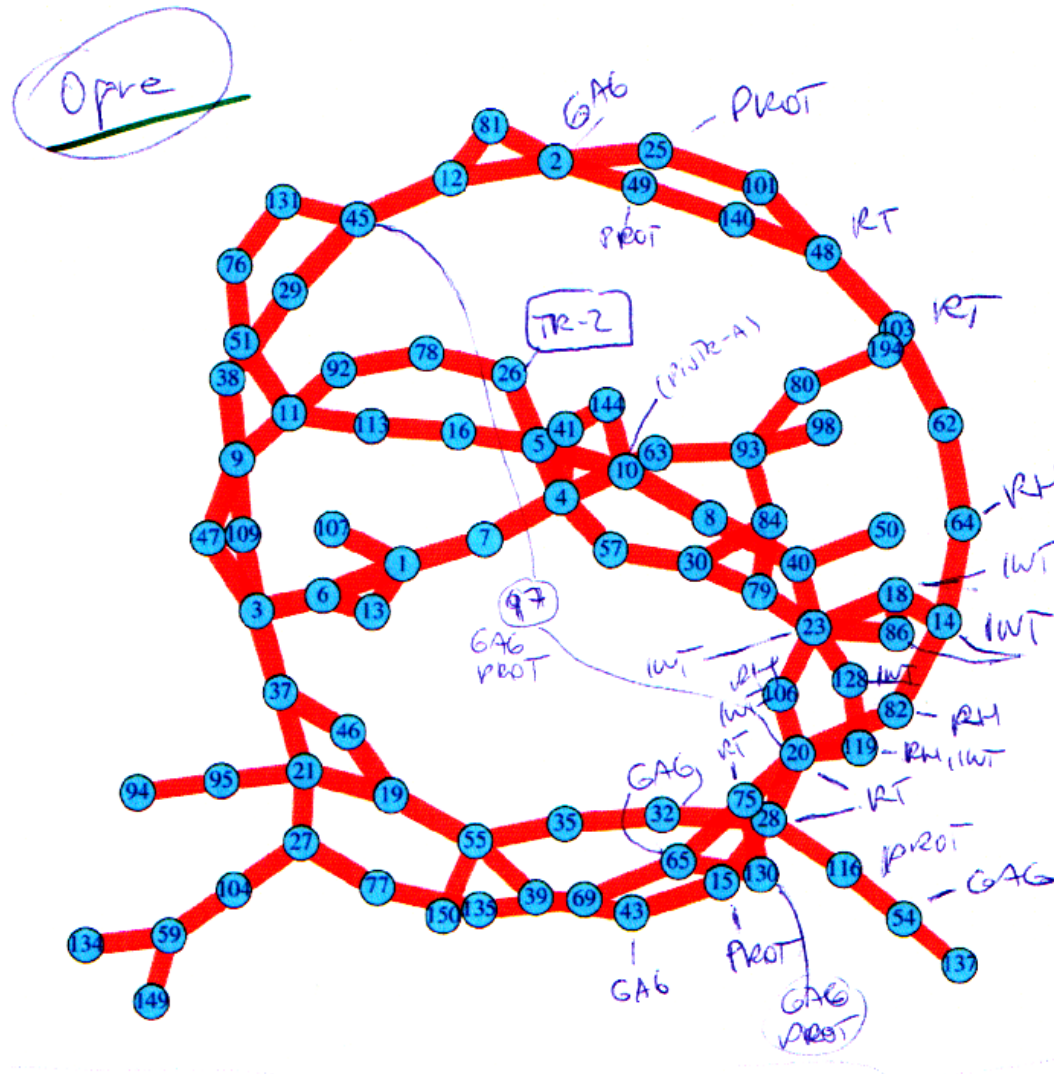
Evaluation of cluster connections via paired-end reads



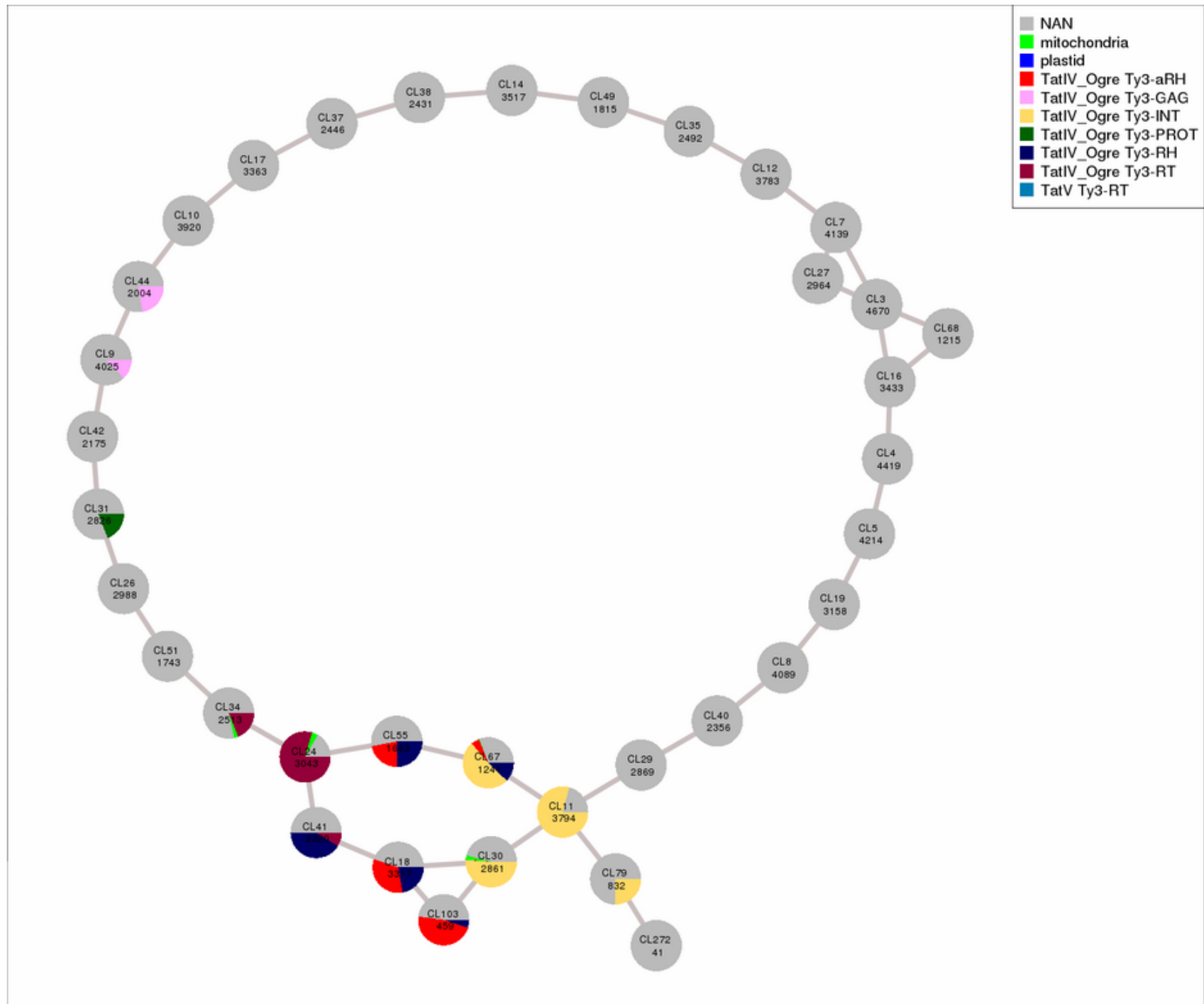
LTR-retrotransposon Ogre in *Vicia pannonica* (~ 40% of the genome)

Superclusters in the past...

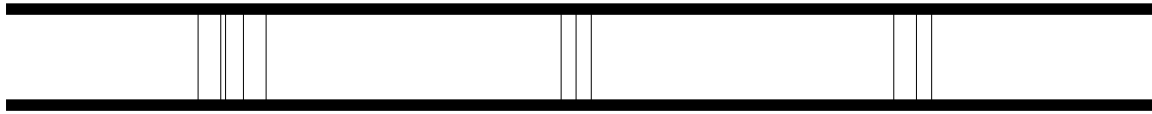
PFL 0.1



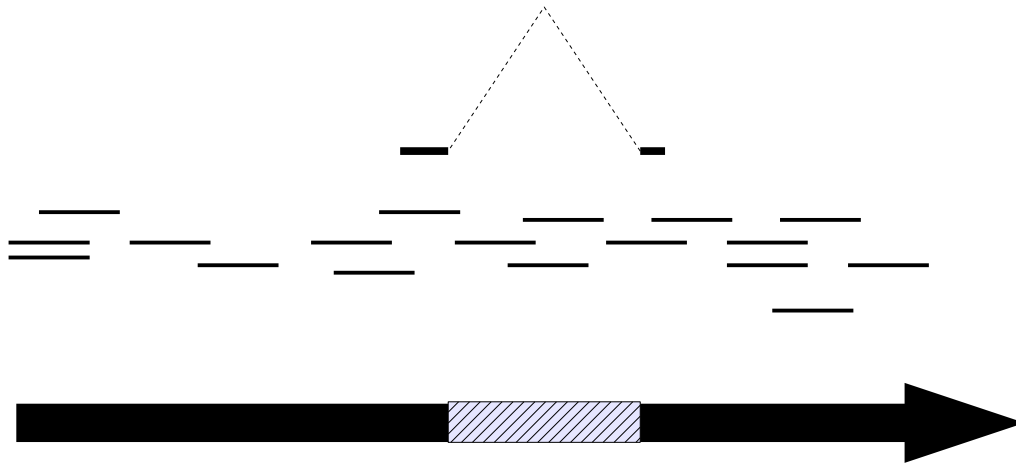
...and now in RE 2



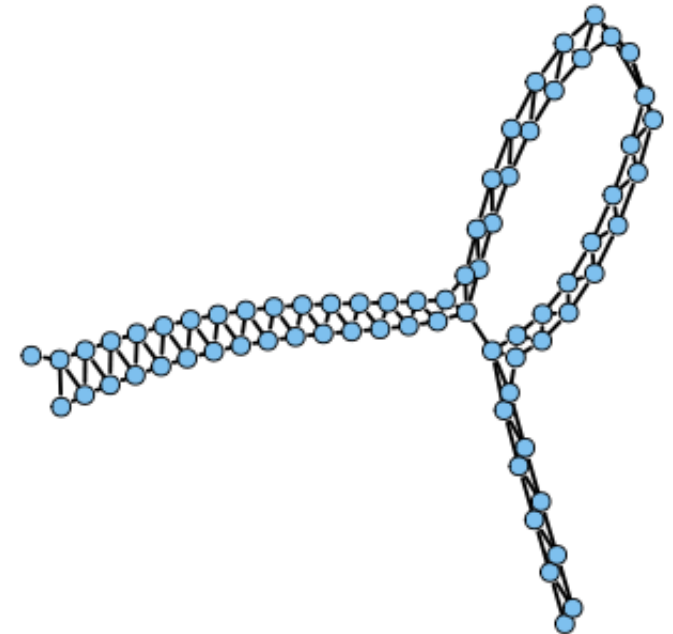
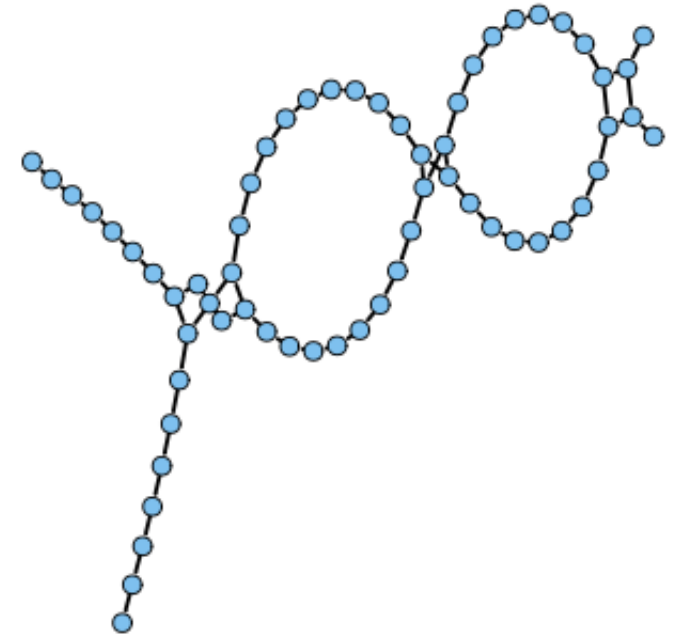
It's getting complicated...



Repeat variants with partial similarity
(depends on similarity threshold)



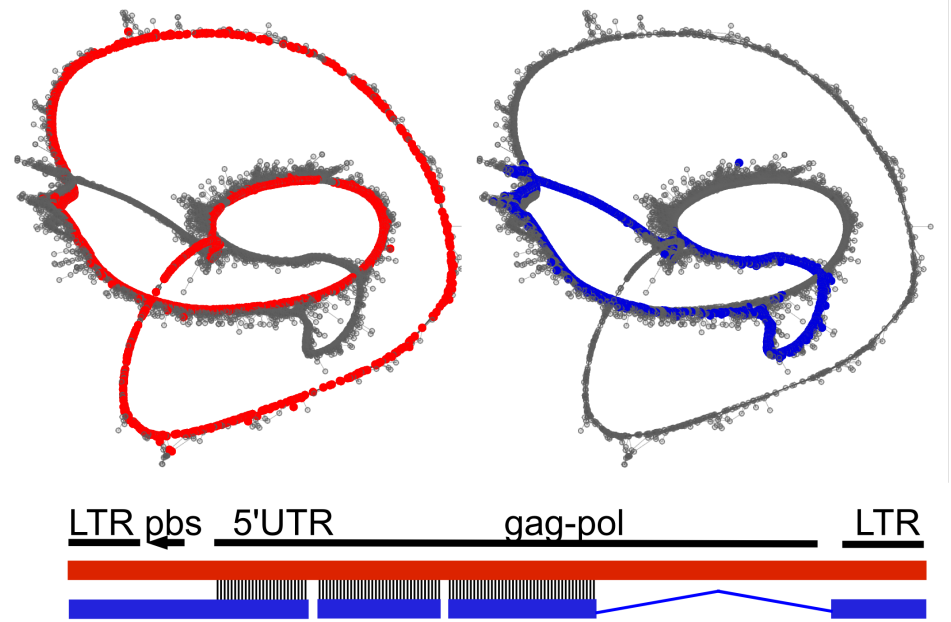
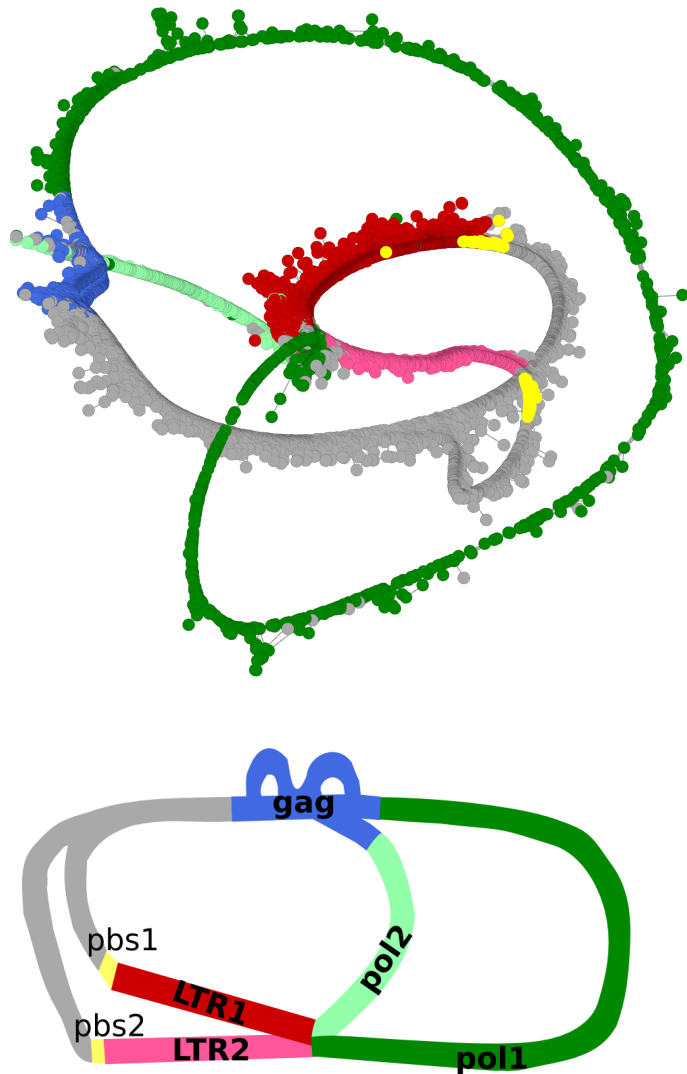
Variants differing in indel



It's getting complicated...

...but it has some meaning

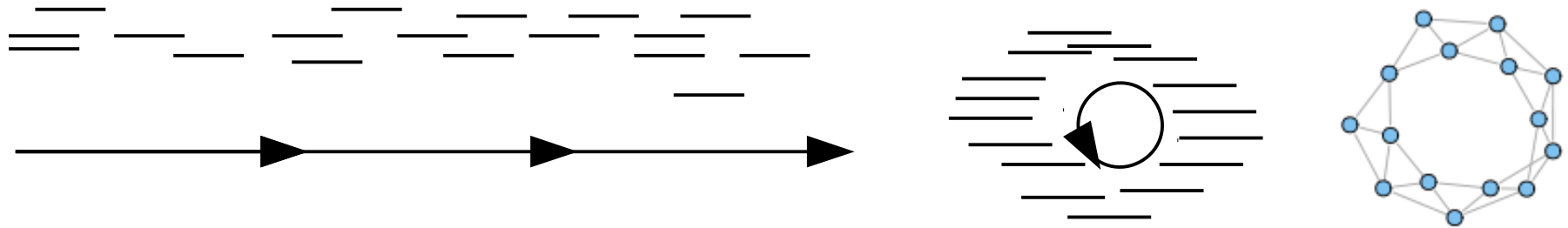
(in this case, presence of element variants differing in LTR sequence and in deletion within gag-pol region)



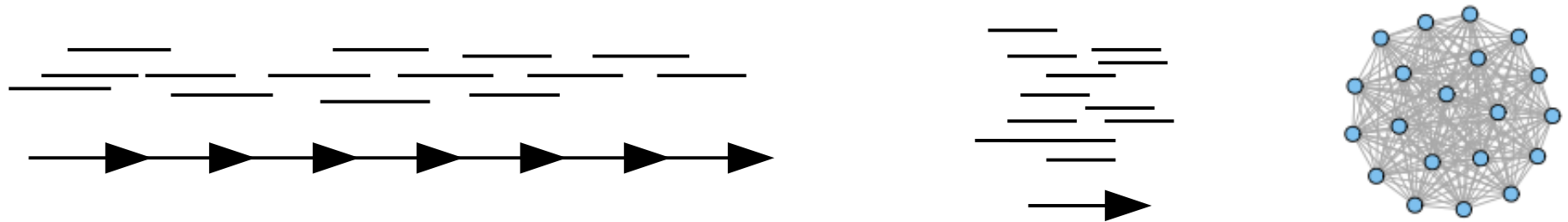
Circular graphs

TANDEM REPEATS

Read length $\ll$ monomer

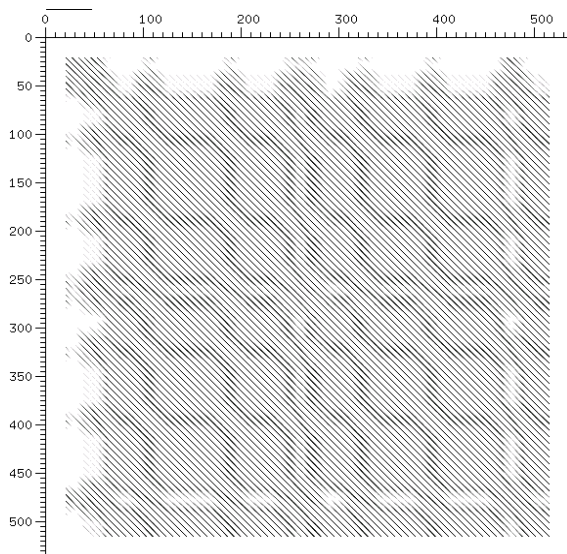
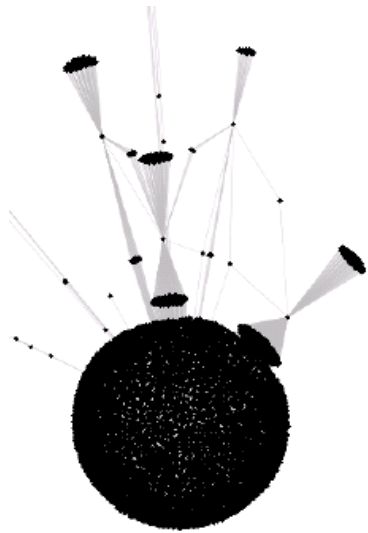


Read length $\geq$ monomer

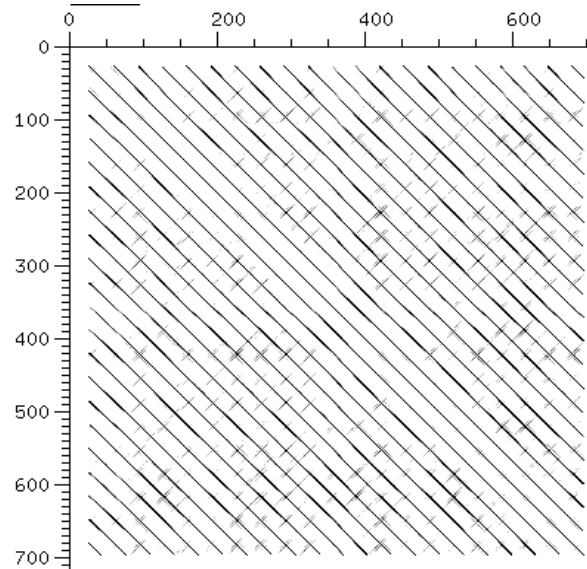
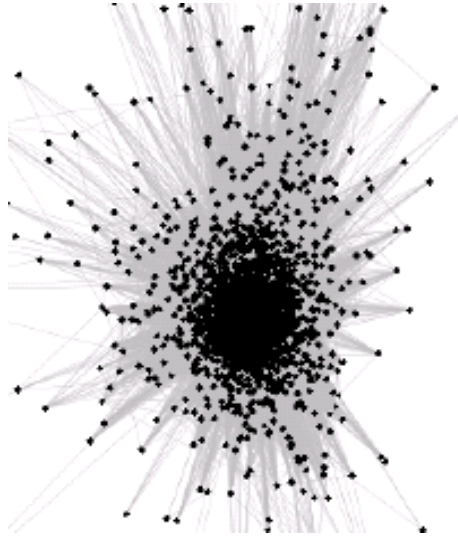


Circular graphs

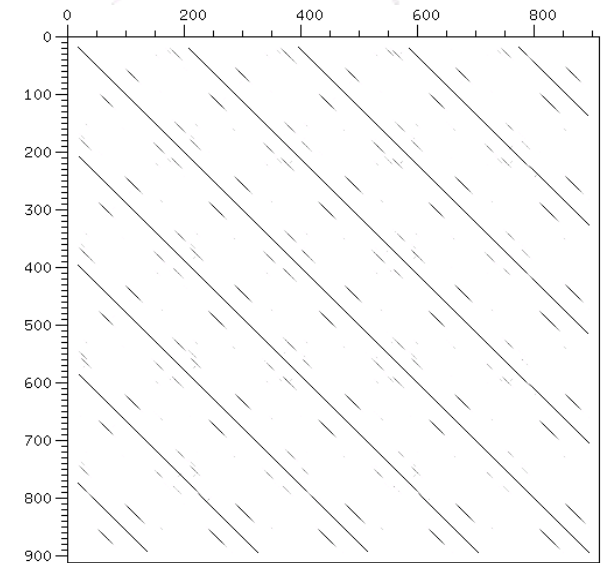
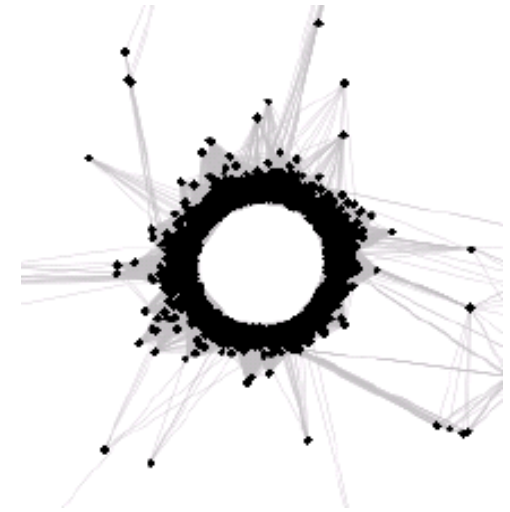
microsatellite (6bp)
(GAACCT)_n



satellite 35 bp

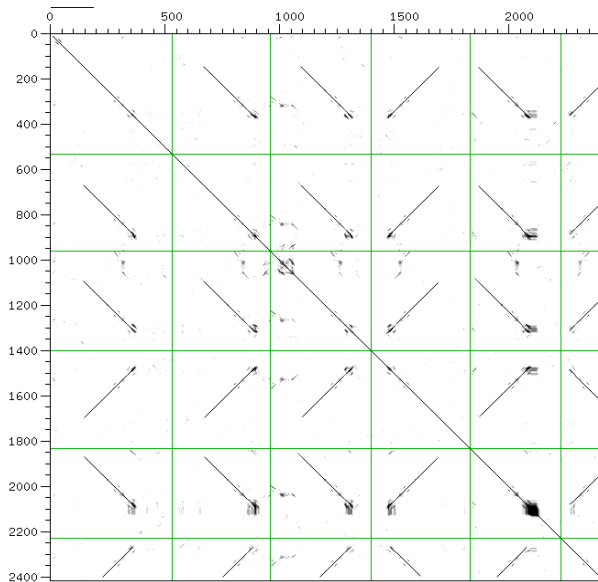
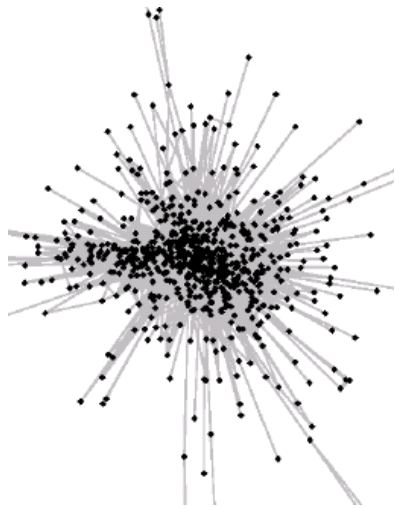


satellite 190 bp

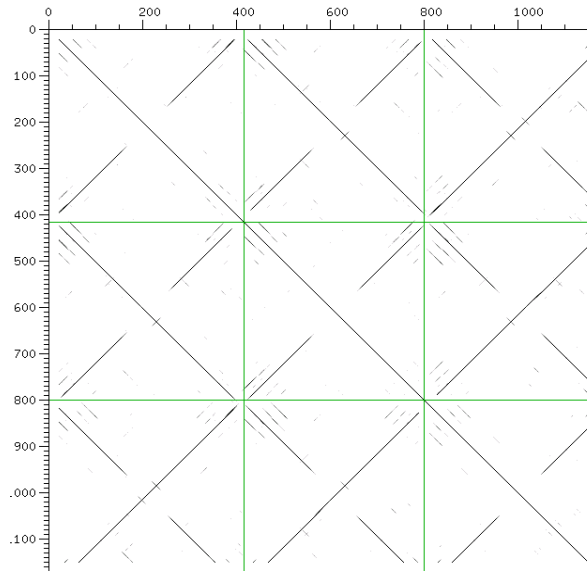


Circular graphs

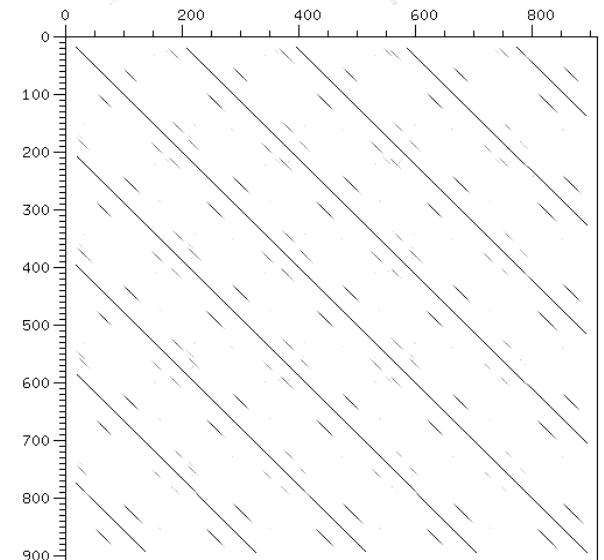
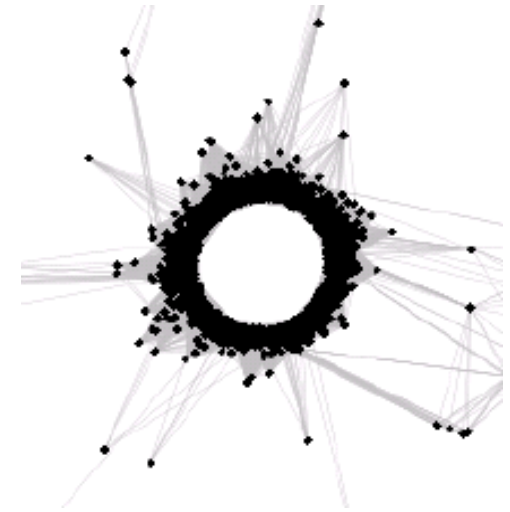
SINE



MITE
(foldback)

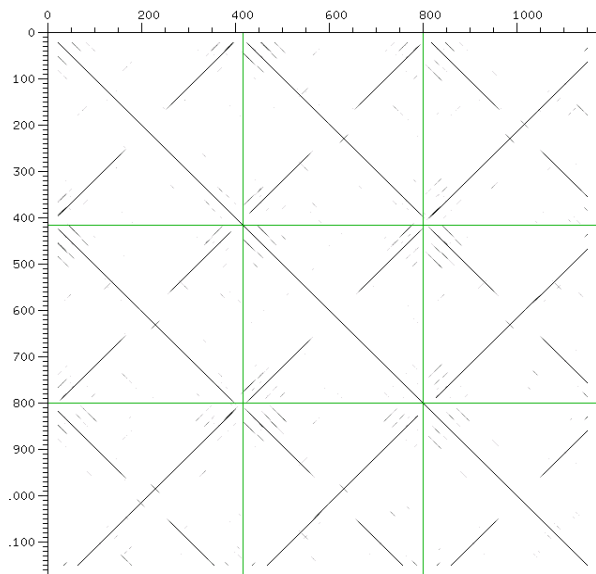


satellite 190 bp



Insertion sites of mobile elements

MITE (foldback)

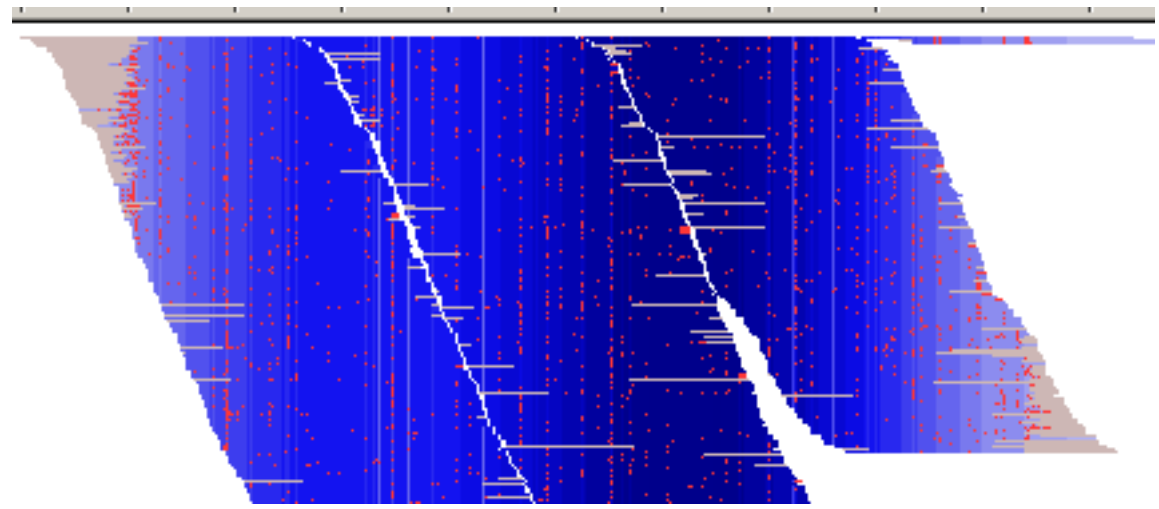


```
ATTCAATAATATAATTTTTTTTAGGGTTTTCAACAATTTTAGTCATATT  
TTTATTAATAATTTAAATAATAAAGGGTTTTGAACAATTTTAGTCATATT  
CTTAATTCAATTATAAAAGATTTAAAGGGTTTTTAACAATTTTAGTCATATT  
CCTCGTAATAATAATTAATAAATTTTTGGATTTTCAACCATTTTAGTCATATT  
AAAATTGTATTTTAAATTTTTTTTAGGGTTTTCAACAATTTTAGTCATATT  
TTGAGCCAAACTAATAGTTAAATTTTAGGGTTTTCAACAATTTTAGTCATATT  
ACACATGCCATTTTATGATAGAAAAAGGGTTTTCAACAATTTTAGTCATATT  
TAGTAGTAGAAATGGTTTTAGGGTTTAGGGTTTTCAACAATTTTAGTCATATT  
TTTTATTATTTTGTGATTTTAAATAGGGTTTTGAACAATTTTAGTCATATT  
CTTAATTCAATTATAAAAGATTTAAAGGGTTTTTAACAATTTTAGTCATATT  
AAAATTGTATTTTAAATTTTTTTTAGGGTTTTCAACAATTTTAGTCATATT  
GCAGTTTATGGTATAAAATTTTAAAGGGTTTTCAATAATTTTAGTCATATT  
TGTAATTTTATTTGCATATTTTAAAGGGTTTTCAATCTTTTGGTCATATT  
TTTGAAATAAAATAGAGTATAAATAGGGTTTTCAACAATTTTAAATCATATT  
TTAATGGACTAAATGTGTATTTTAAAGGGTTTTCAACAATTTTAGTCATATT  
TTTTAACTAAATTCATGTATTTTAGGGTTTTCAACAATTTTAGTCATATT  
TTCTATTTCAATCATGTATAAATTAAGGGTTTTCAACAATTTTAGTCATATT
```

← TIR

```
TGAAACAAATATGACCAAAAAAGTTAAAAACCCCTTAAATAAAACATAAAGAGTATAAA  
TGAAACAAATATGACCAAAAAAGTTAAAAACCCATTTTAAATTTAAACA  
TGAAACAAATATGACCAAAAAAGTTAAAAATCCATAGATTTAATGTGAAAA  
TAAACAAATATGACCAAAAAAGTTAAAAACCCATATATATATATGTGA  
TGAAACAAATATACCAAAAAAGTTAAAAACCCCTTAAATAAAATATAGTTA  
TGAAACAAATATGACCAAAAAAGTTAAAAACCTATTTTTTTTTATAAGATTAT  
TGAAACAAATATACCAAAAAAGTTAAAAACCCCTTAAATTATACCCATTTT  
TGAAACAAATATGACCAAAAAAGTTAAAAACCCATTTTAAATTTAAACAATTAT  
TGAAATAAATATGAACAAAAATTTTAAACCCCTATGAATATTGTTATAAGGG  
CGAAACAAATATGACCAACAAATTTAAAAACCCCTTCTCTCTATATTTTTTA  
TGAAACAAATATGACCAAAAAAGTTAAAAACCCCTTAAATGAATTACAAATATGCGTG  
TAAACAAATATGACCAAAAAAGTTTAAACCCCTTAAATAAAACATAAAGAGTATAAA  
TAAACAAATATGACAAAAAAAGTGAAAAACCCAAATTTGAATCATACAAAAAGAAG  
TGAAATAAATATGACCAAAAAAGTTAAAAACCCGGTAATAAATAGAGTAAGCATATT  
TGAAACAAATATGACCAAAAAAGTTAAAAATCCATAGATTTAATGTGAAAAATACGATAT  
TGAAACAAATATGACCAAAAAAGTTAAAAACCCGTATATTTAATGTGAAAAATACGCT  
TGAAACAAATATACCAAAAAAGTTAAAAACCCCTTAAATTATACCCATTTTGTTTT
```

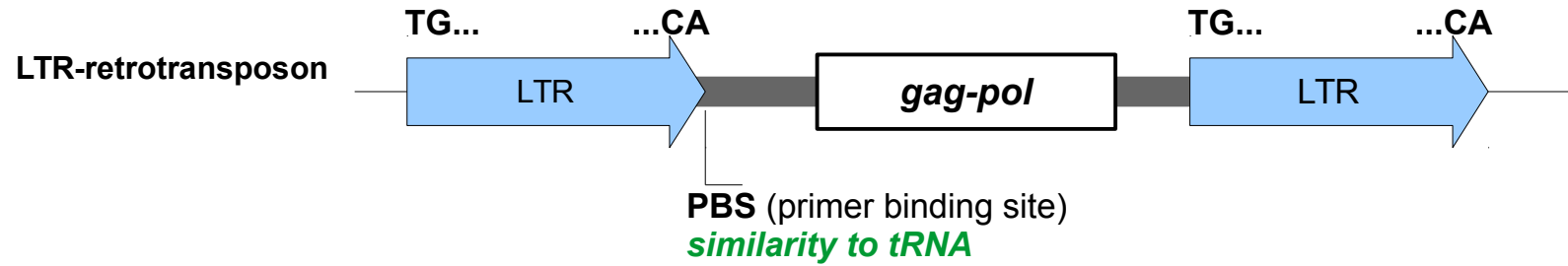
TIR →



contig

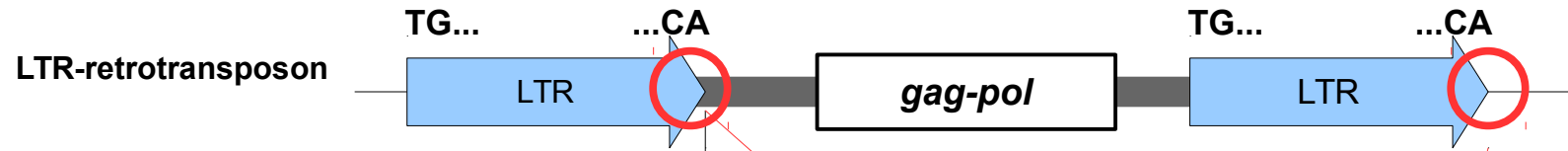
Insertion sites of mobile elements

A new tool for detection of LTR / PBS sites

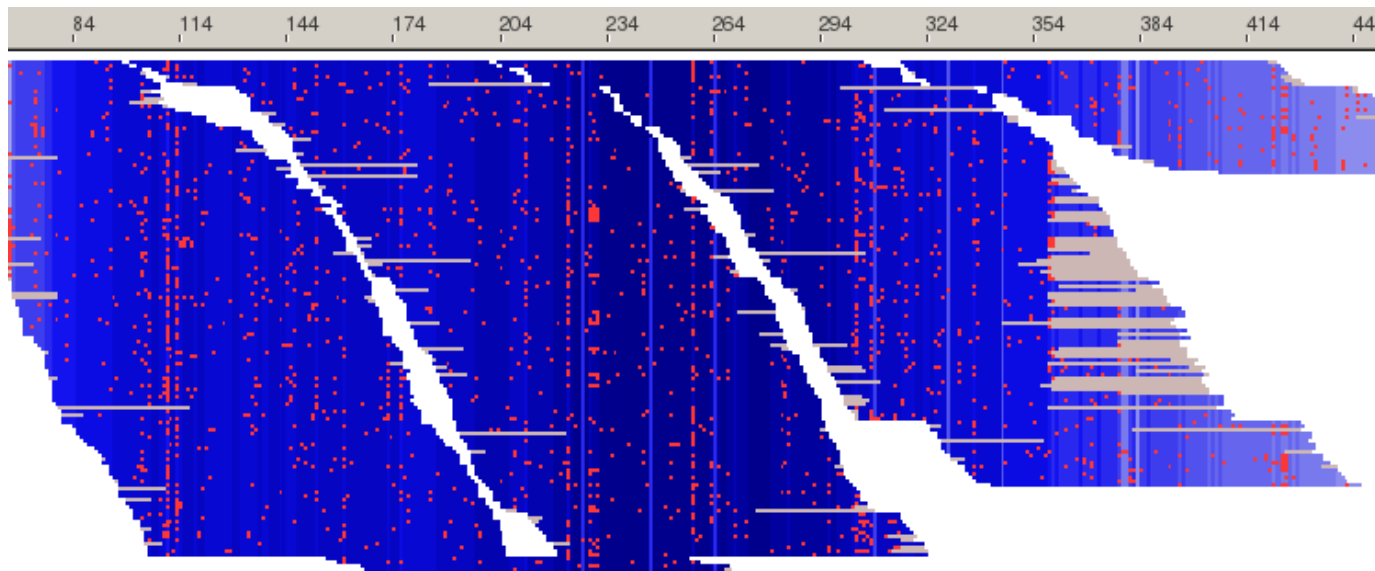


Insertion sites of mobile elements

A new tool for detection of LTR / PBS sites

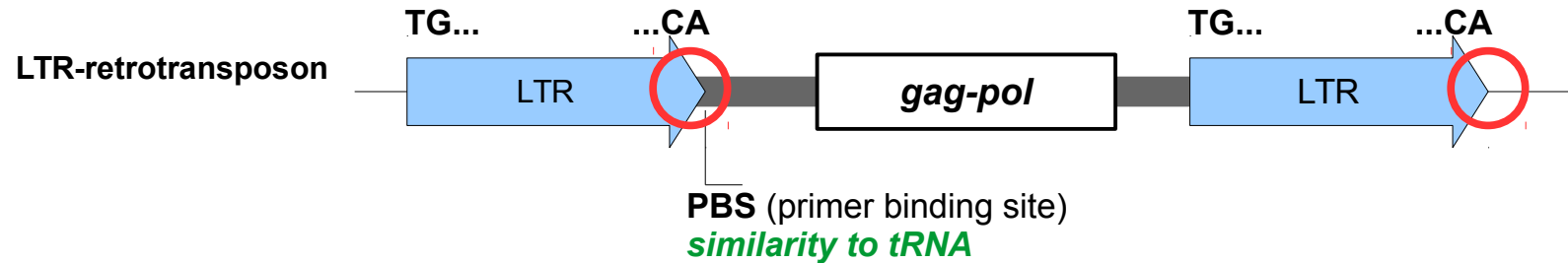


PBS (primer binding site)
similarity to tRNA



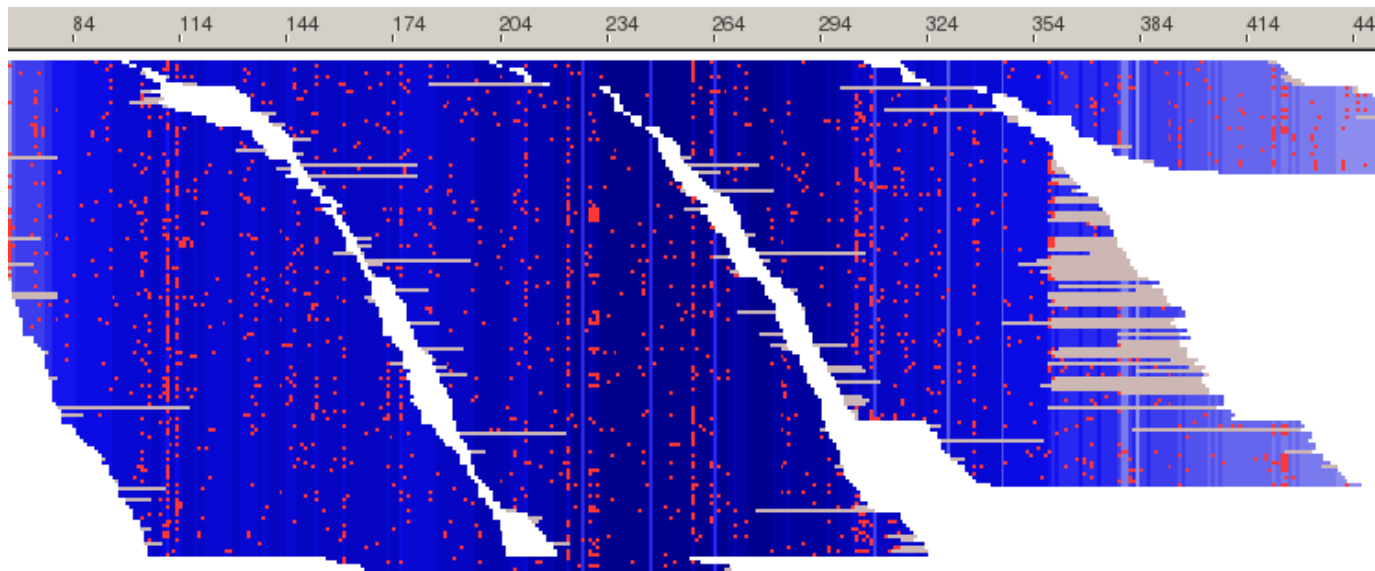
Insertion sites of mobile elements

A new tool for detection of LTR / PBS sites

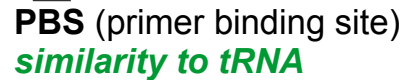


The tool searches assembled contigs (ACE files) for **differences in proportions of masked reads** accompanied by **occurrence of TG/CA**

↓
...and **similarity to tRNA**

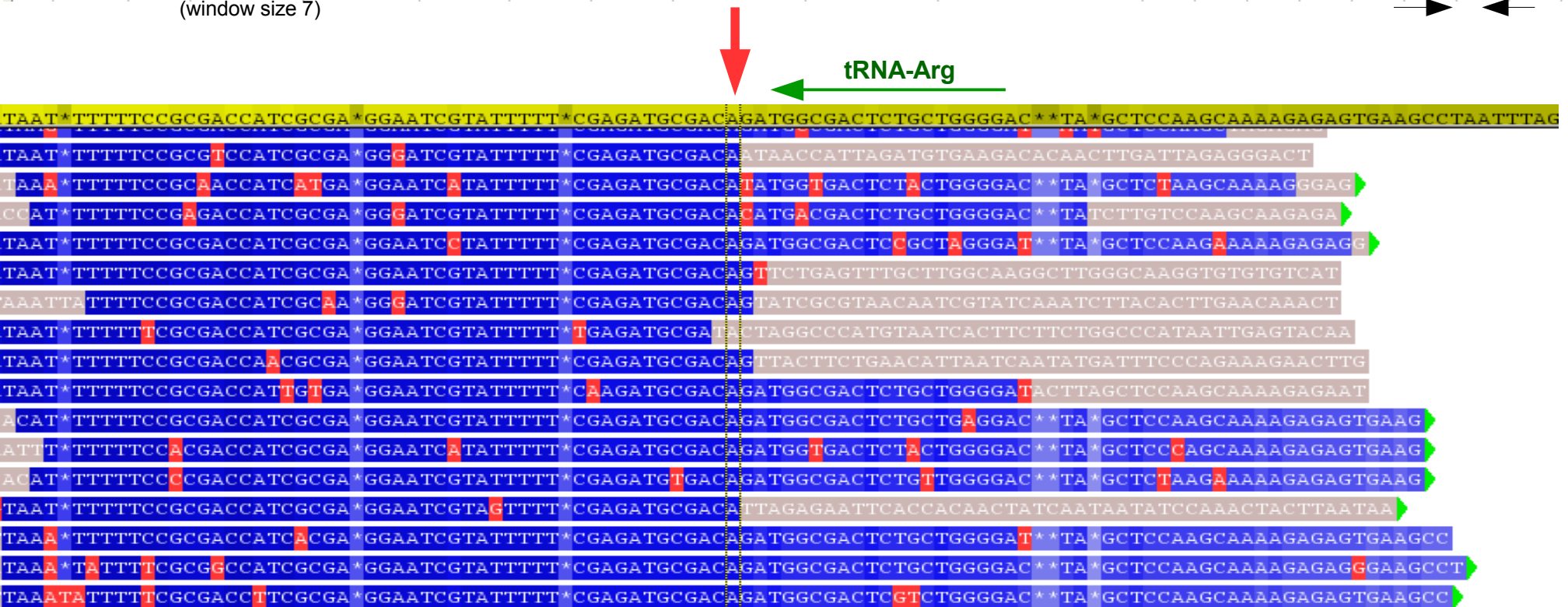


A new tool for detection of LTR / PBS sites



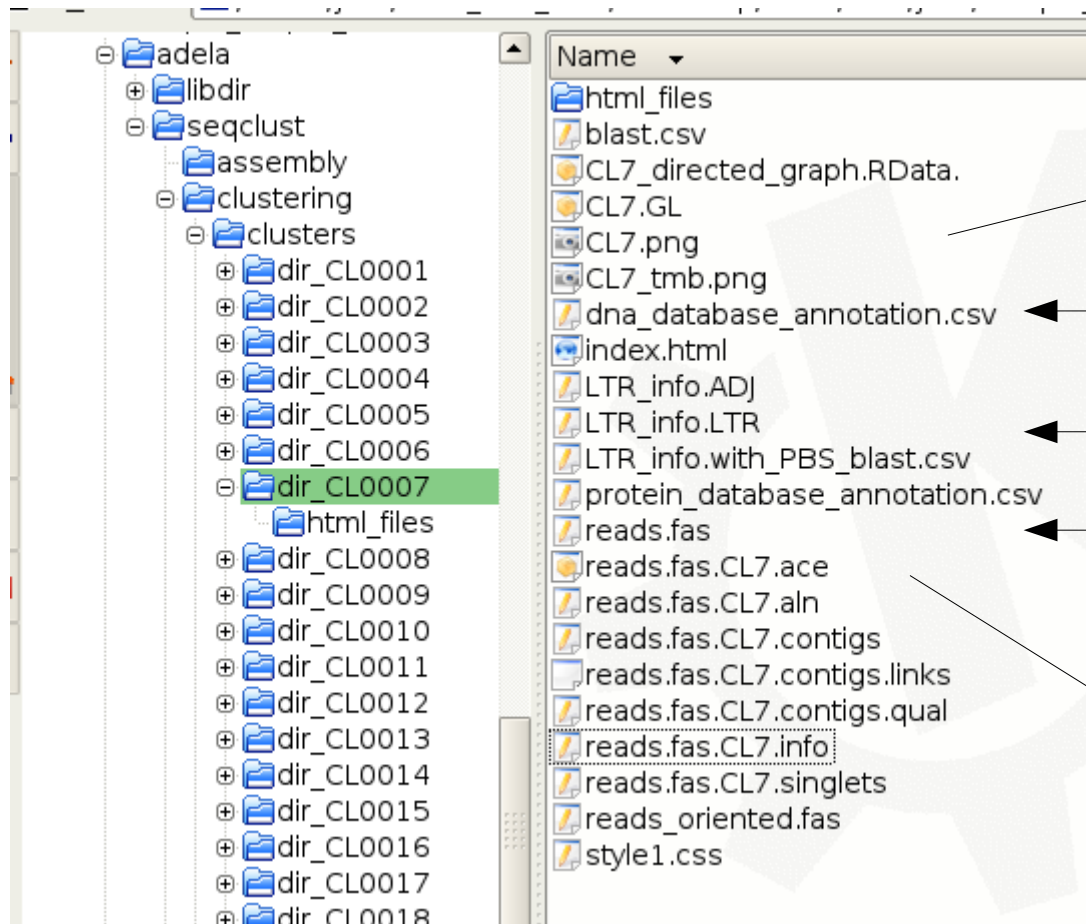
CL	contig	pos.	site	site_depth	out_ma	maske	maske	region_in	region_out	blast to tRNA	%	length		site from	to	tRNA from	to	E-val	
19	400	364	TGCGACA	106.6	30.4	0.0362	0.2975	GCGAGGAA	GATGGCGA	At-chr2.tRNA28-Arg	100	18	0	0	3	20	23	6	7E-007

(window size 7)



RepeatExplorer output files

Cluster directories

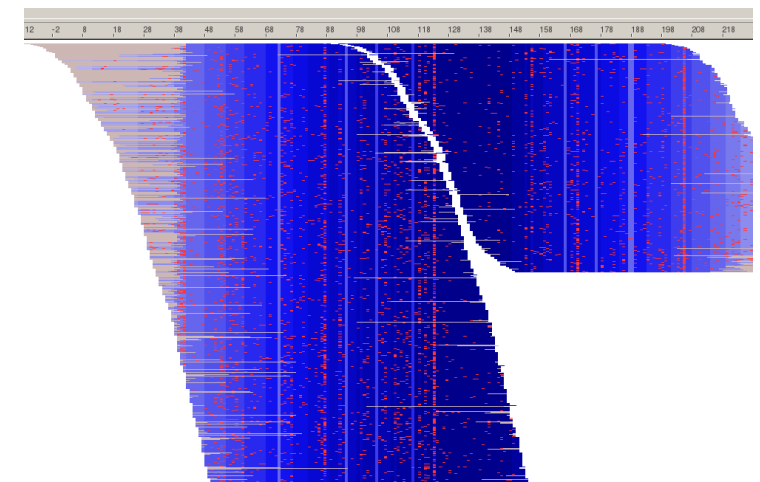


Cluster graph

Similarity searches

LTR-detection

Read sequences



assembly