Genomic divergence outside genes

Evolution of Genome Size Workshop
Plant and Animal Genome 2011

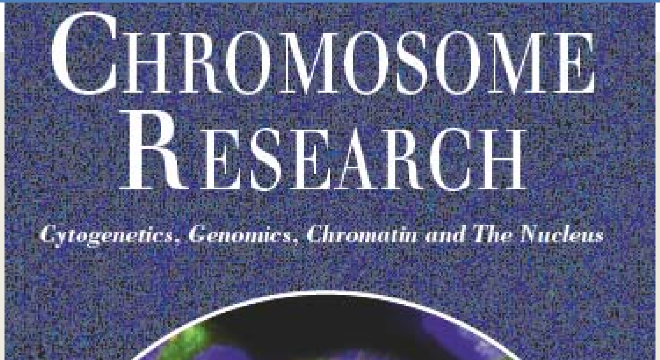
Genome sequencing (and common sense!) shows the gene sequences are similar (from bacteria to animals to plants)

In eukaryotes, genome organization is conserved: discrete chromosomes, centromeres/telomeres, genes exons/introns, transposable elements, rDNA repeats, satellite DNA
(but there are some exceptions – this is not the only organization possible)

Chromosome number (plants: 2n=4 to 1440) and genome size (plants: 2,350 range from 1C=63 Mb to 150,000 Mb) are far from conserved

Pat Heslop-Harrison, Faisal Nouroz, Trude Schwarzacher

and Farah Badakshi, Ana Claudia Guerra, Guto Kuhn, John Bailey, Thomas Schmidt



CHROMOSOME RESEARCH

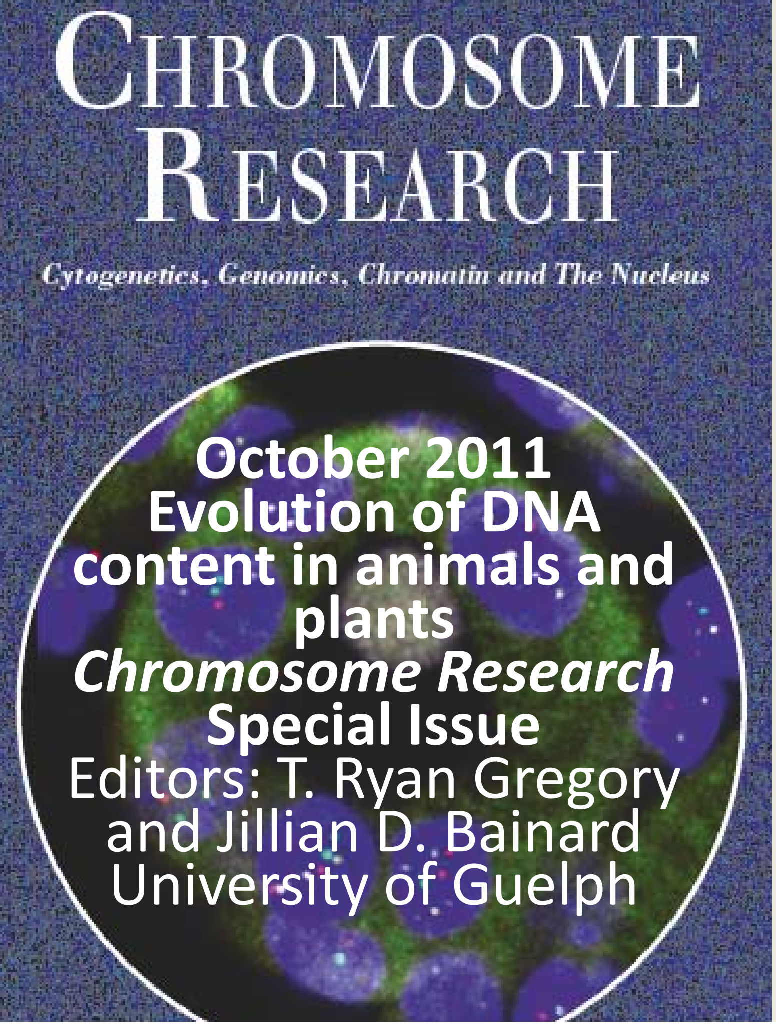
Cytogenetics, Genomics, Chromatin and The Nucleus

ANNALS OF BOTANY

AoBBlog.com

University of Leicester

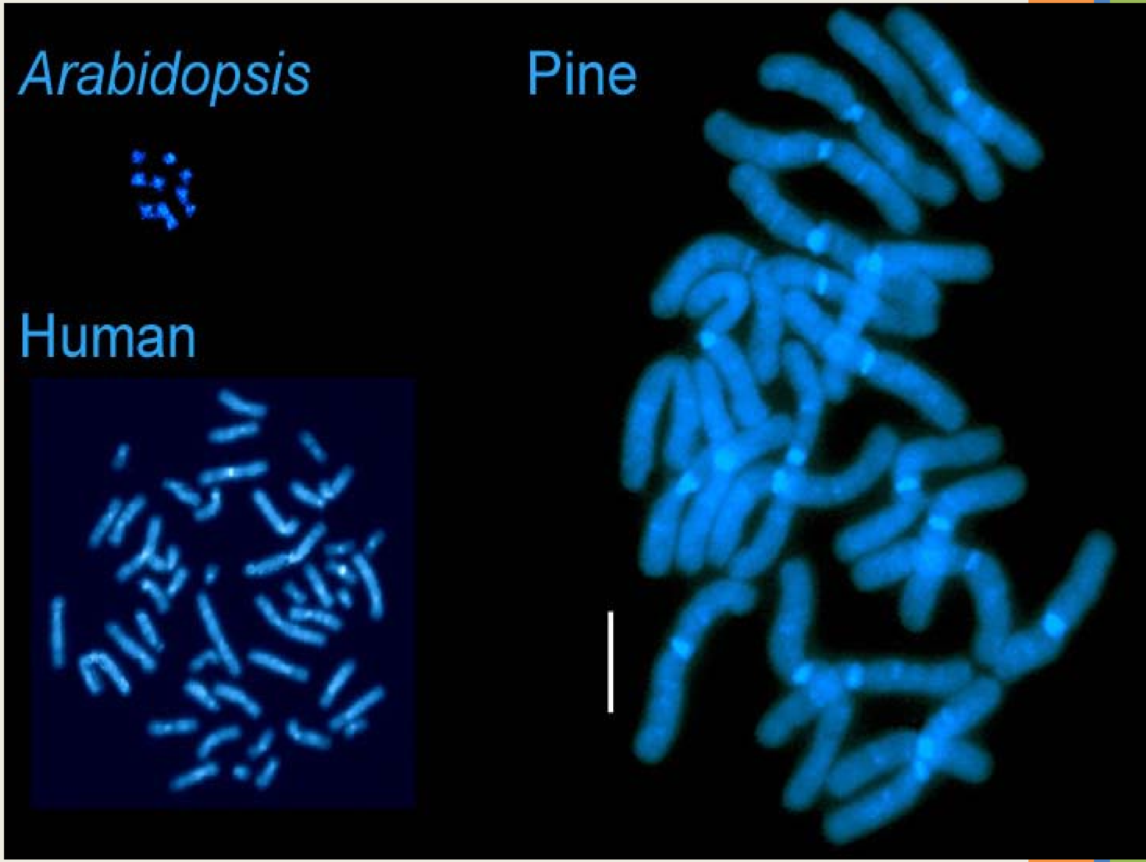
phh4@le.ac.uk
www.molcyt.com



CHROMOSOME RESEARCH

Cytogenetics, Genomics, Chromatin and The Nucleus

October 2011
Evolution of DNA content in animals and plants
Chromosome Research **Special Issue**
Editors: T. Ryan Gregory and Jillian D. Bainard
University of Guelph



Annals of Botany Page 1 of 124
doi:10.1093/aob/mcq258, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

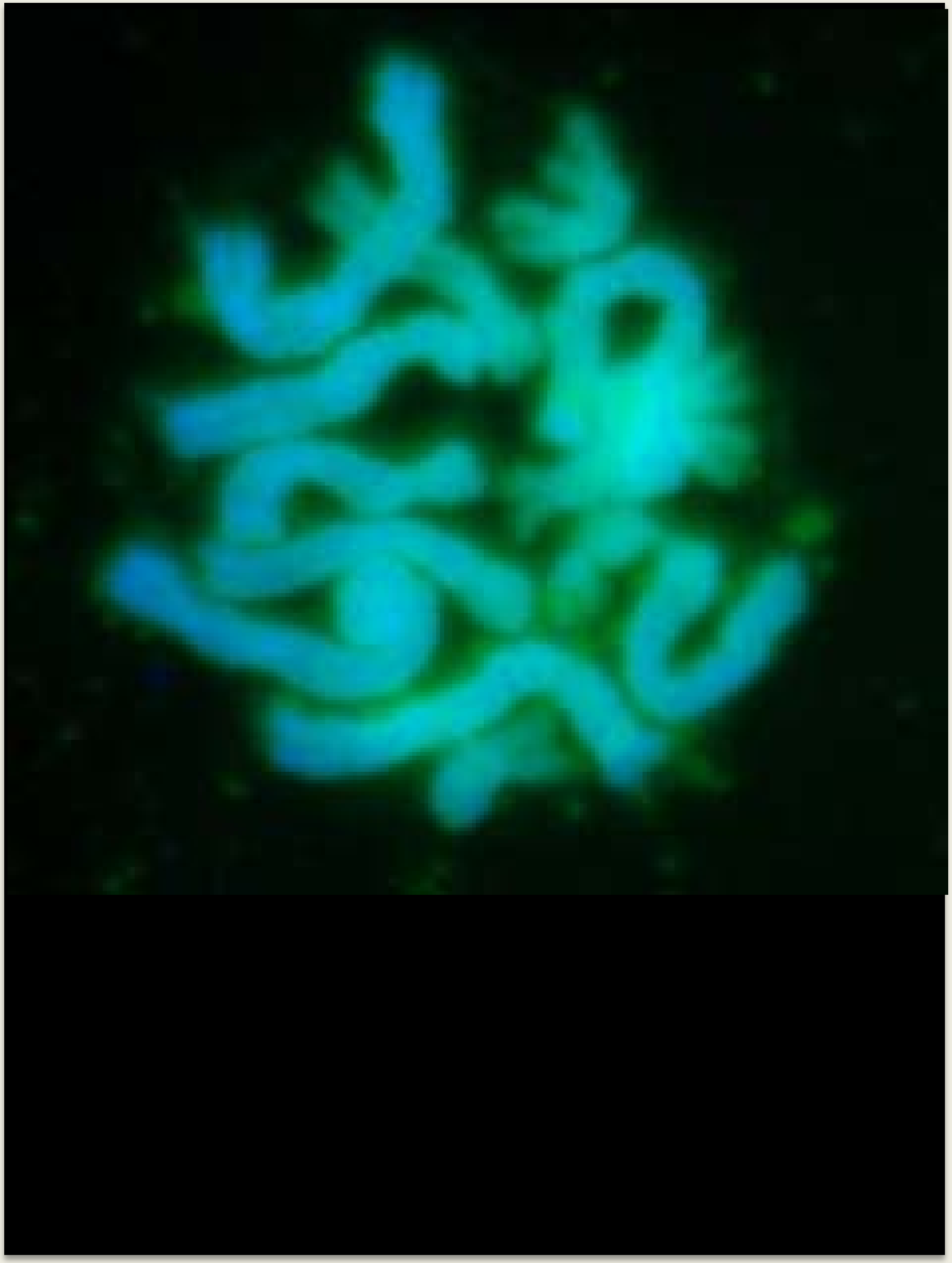
Nuclear DNA amounts in angiosperms: targets, trends and tomorrow

M. D. Bennett* and I. J. Leitch

How do the genomes of related species differ?

- Chromosome number
- Ploidy
- Transposable elements
- Microsatellites
- Satellite DNA sequences
- Gene sequences
- Chromosome rearrangements

Orgaard, HH et al.



How do the genomes of related species differ?

-Chromosome number

-Ploidy

-Transposable elements

-Microsatellites

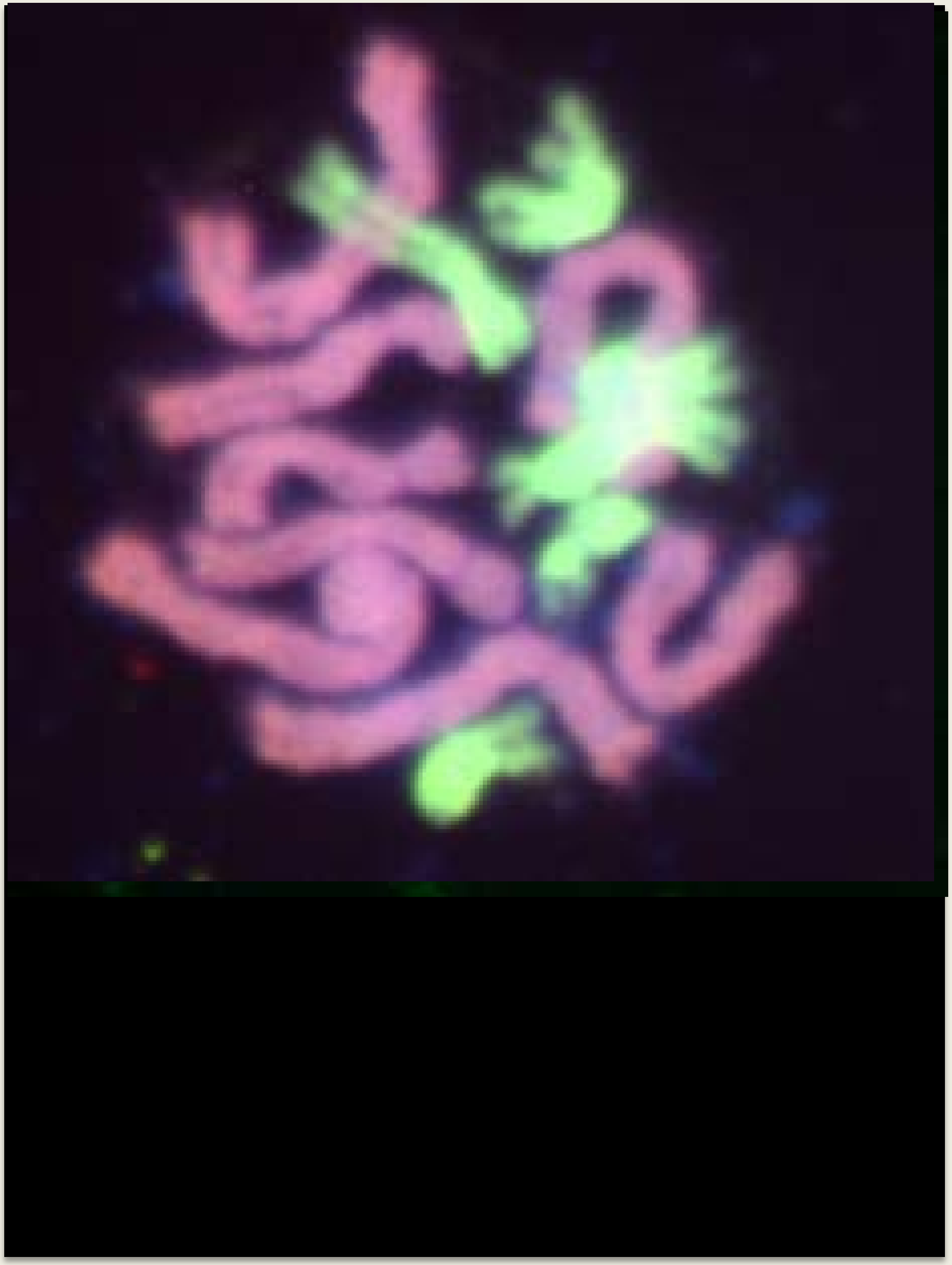
- Satellite DNA sequences

-Gene sequences

-Chromosome rearrangements

-Mechanisms

-Why? and consequences

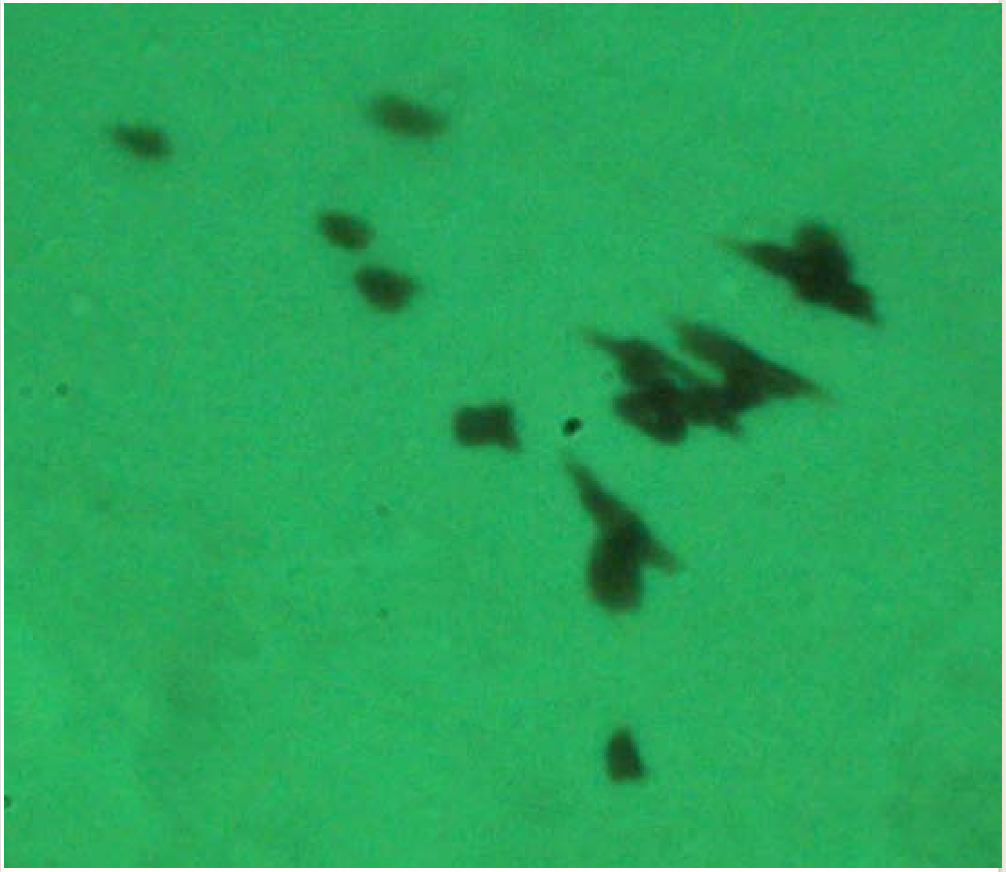


How do the genomes of related species differ?

- Chromosome number
- Ploidy
- Transposable elements
- Microsatellites
- Satellite DNA sequences
- Gene sequences
- Chromosome rearrangements

- Mechanisms

- Why? and consequences



How do the genomes of related species differ?

-Chromosome number

-Ploidy

-Transposable elements

-Microsatellites

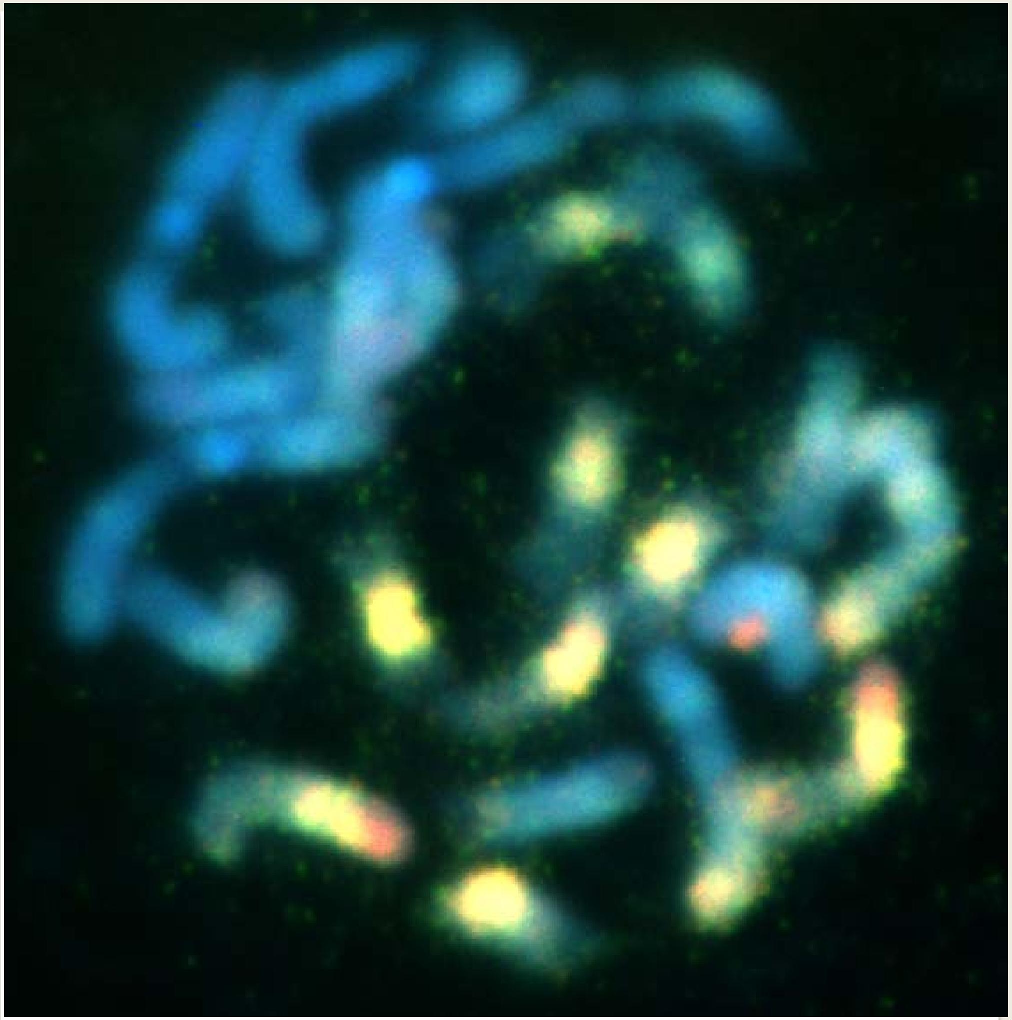
- Satellite DNA sequences

-Gene sequences

-Chromosome rearrangements

-Mechanisms

-Why? and consequences



How do the genomes of related species differ?

-Chromosome number

-Ploidy

-Transposable elements

-Microsatellites

- Satellite DNA sequences

-Gene sequences

-Chromosome rearrangements

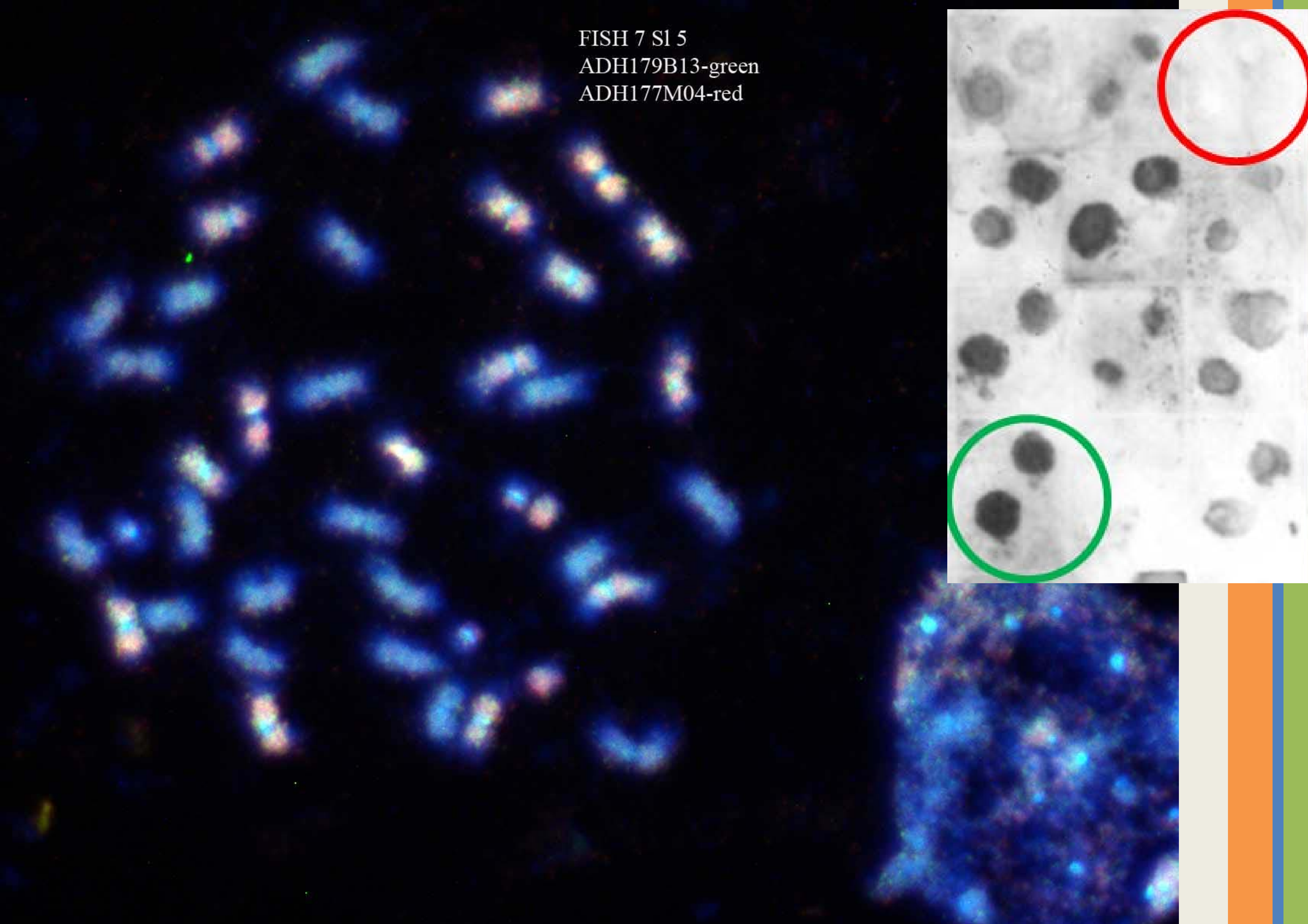
-Mechanisms

-Why? and consequences

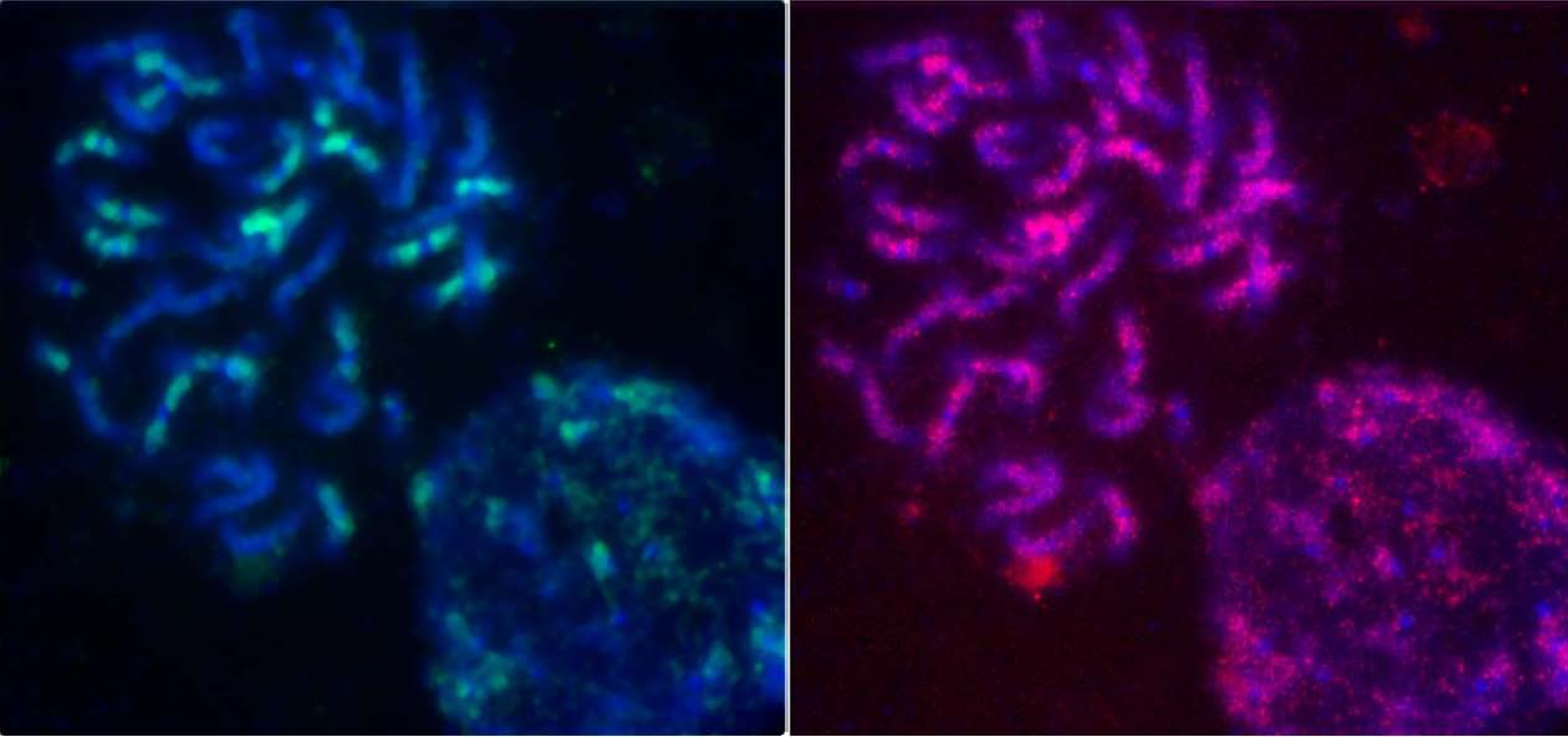
Crocus Bank
www.crocusbank.org
Action CROCUSBANK
(AGRI GEN RES 018)
is supported by the
European Commission,
Directorate General for
Agriculture & Rural
Development

Peanut, *Arachis hypogaea*, $2n=4x=40$
AABB genome constitution

Ana Claudia G. Araujo, EMBRAPA, Brazil

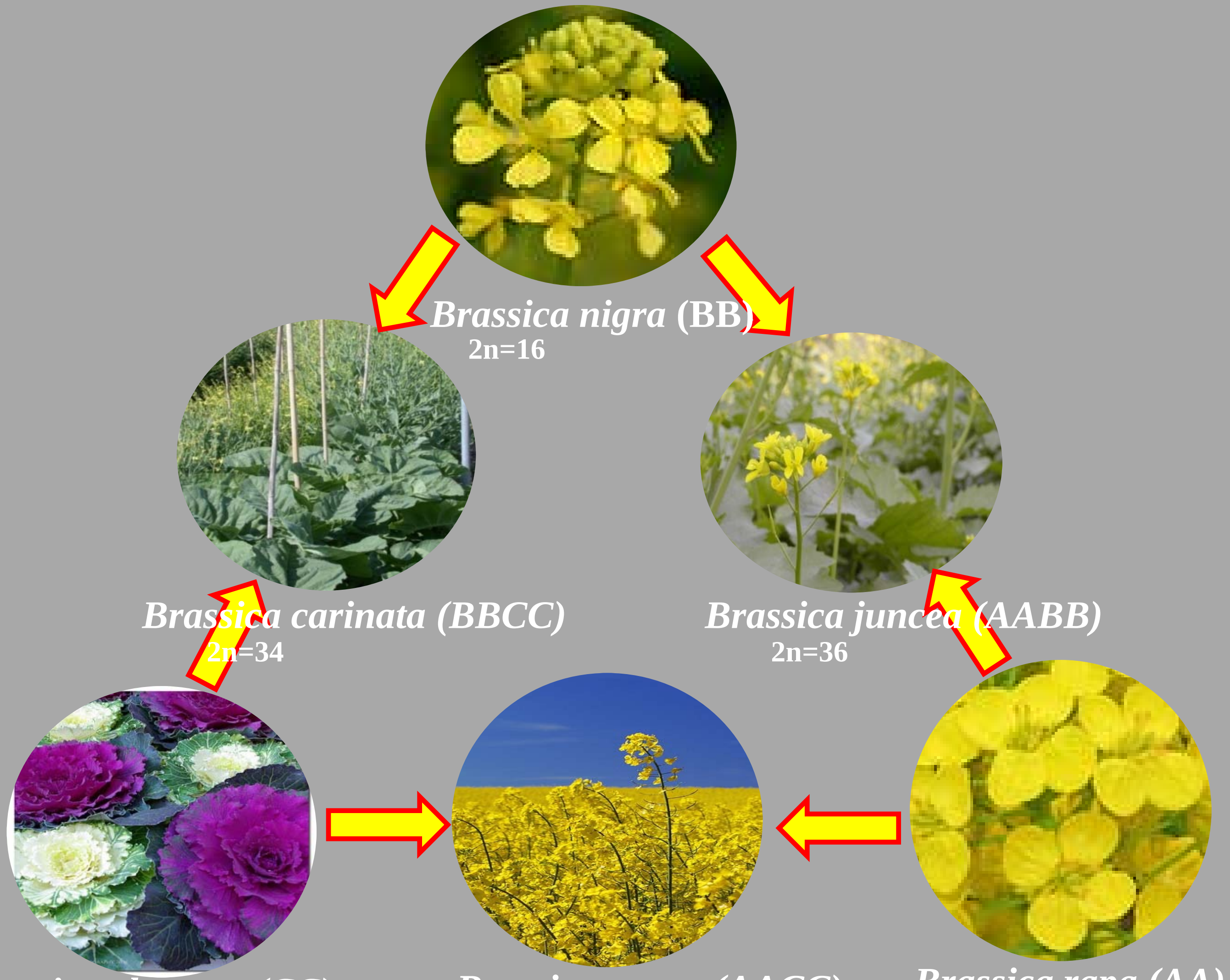


BAC in situ – individual BACs have various repetitive sequences with characteristic distribution patterns



Many of the repetitive sequences are retrotransposons and DNA transposons
Some are microsatellite motifs
Some are satellites – including the most rapidly evolving sequences

Arachis in situ 6 Slide 2 Meta 7
ADH167F07 red ADH129F24 green
167F07 on both genomes;
129F24 on all A genome chrs excluding centromeres



Brassica nigra (BB)
2n=16

Brassica carinata (BBCC)
2n=34

Brassica juncea (AABB)
2n=36

Brassica oleracea (CC)
2n=18

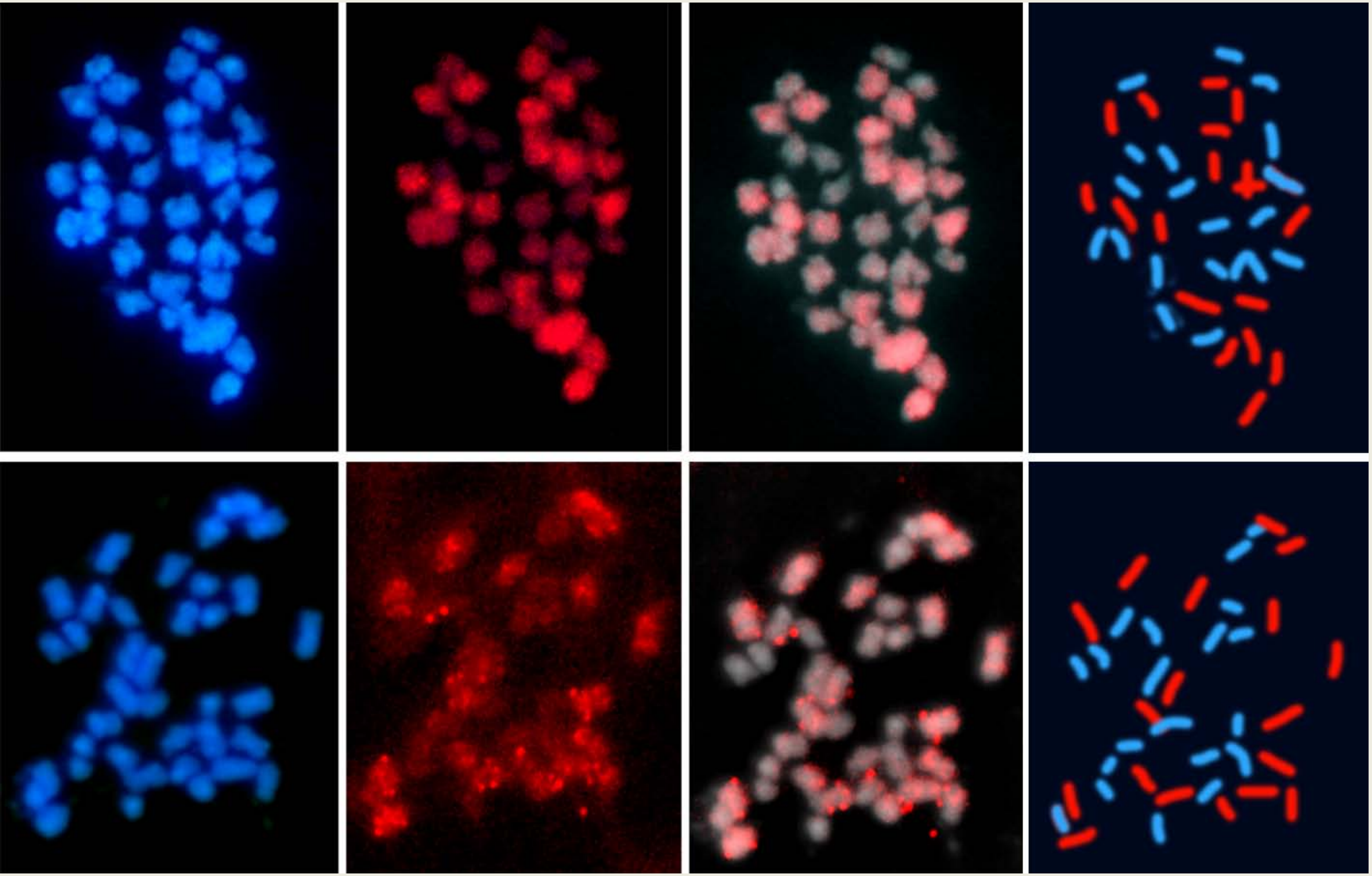
Brassica napus (AACC)
2n=38

Brassica rapa (AA)
2n=20

Genome Specificity of a CACTA (*En/Spm*) Transposon

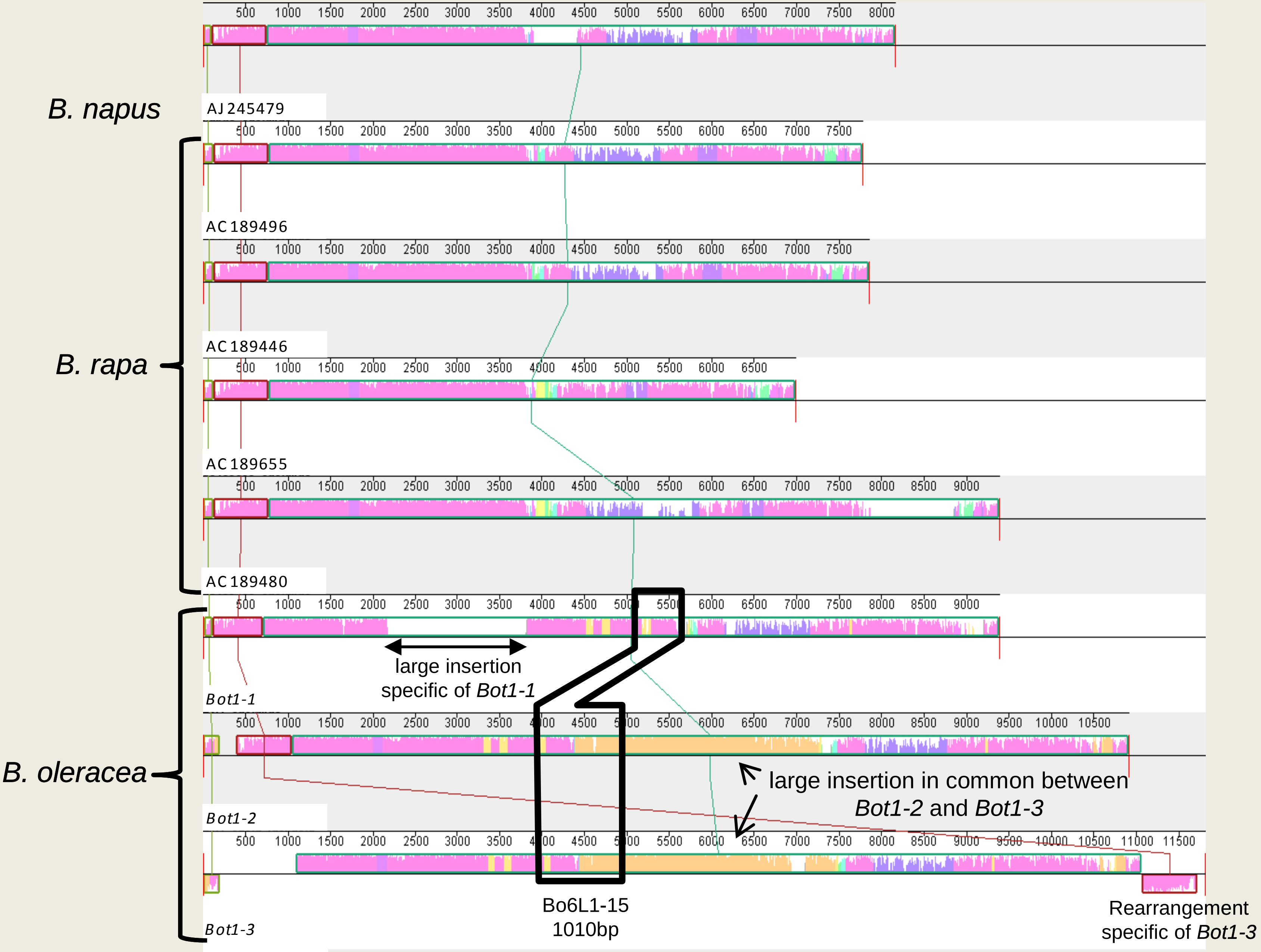
B. napus **(AACC, 2n=4x=38) – hybridized with C-genome CACTA element red**

B. oleracea (CC, 2n=2x=18) **B. rapa (AA, 2n=2x=20)**



Alix et al. The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. Plant Journal

Genome Specificity of a CACTA (En/Spm) Transposon



B. napus

B. rapa

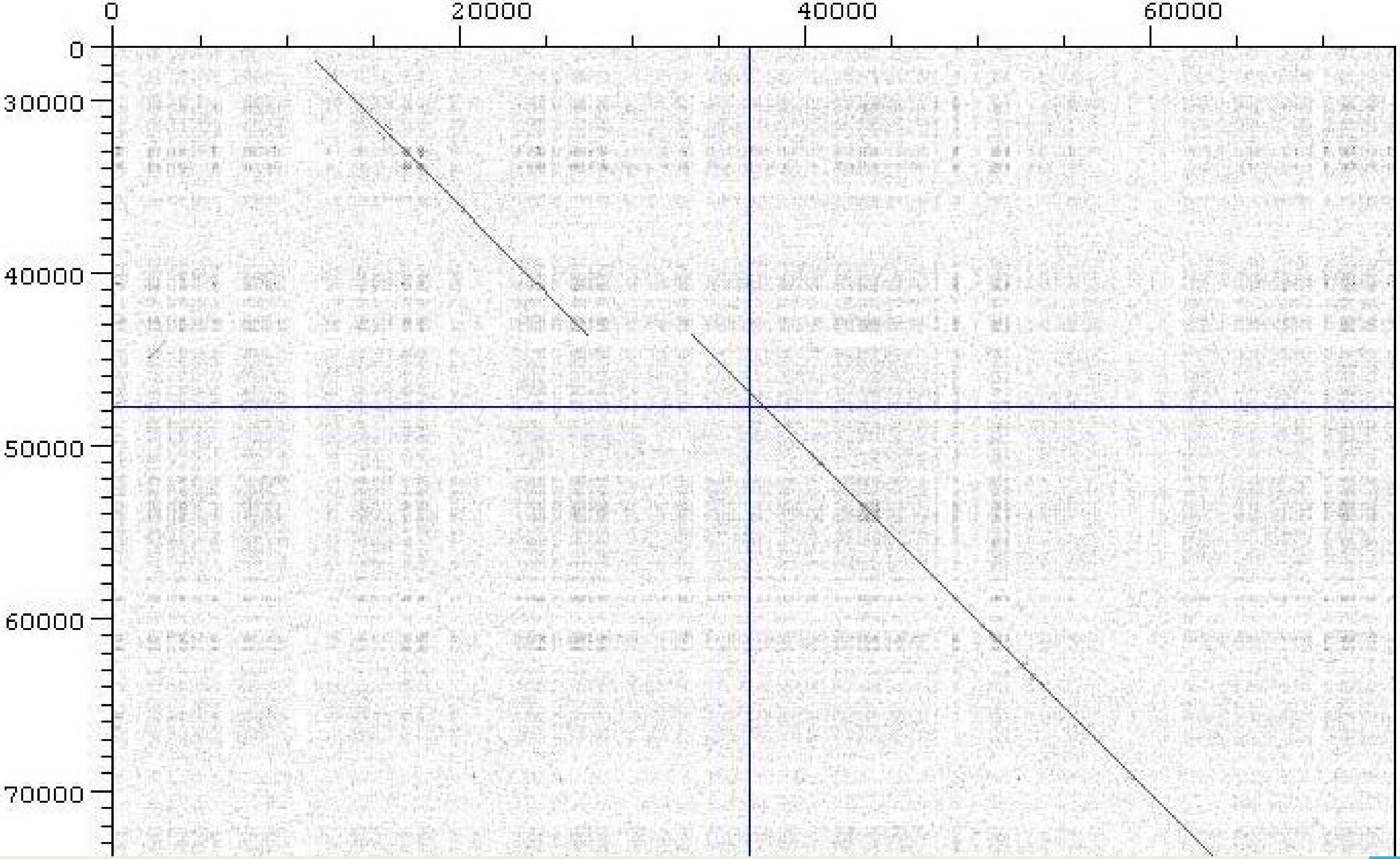
B. oleracea

large insertion specific of *Bot1-1*

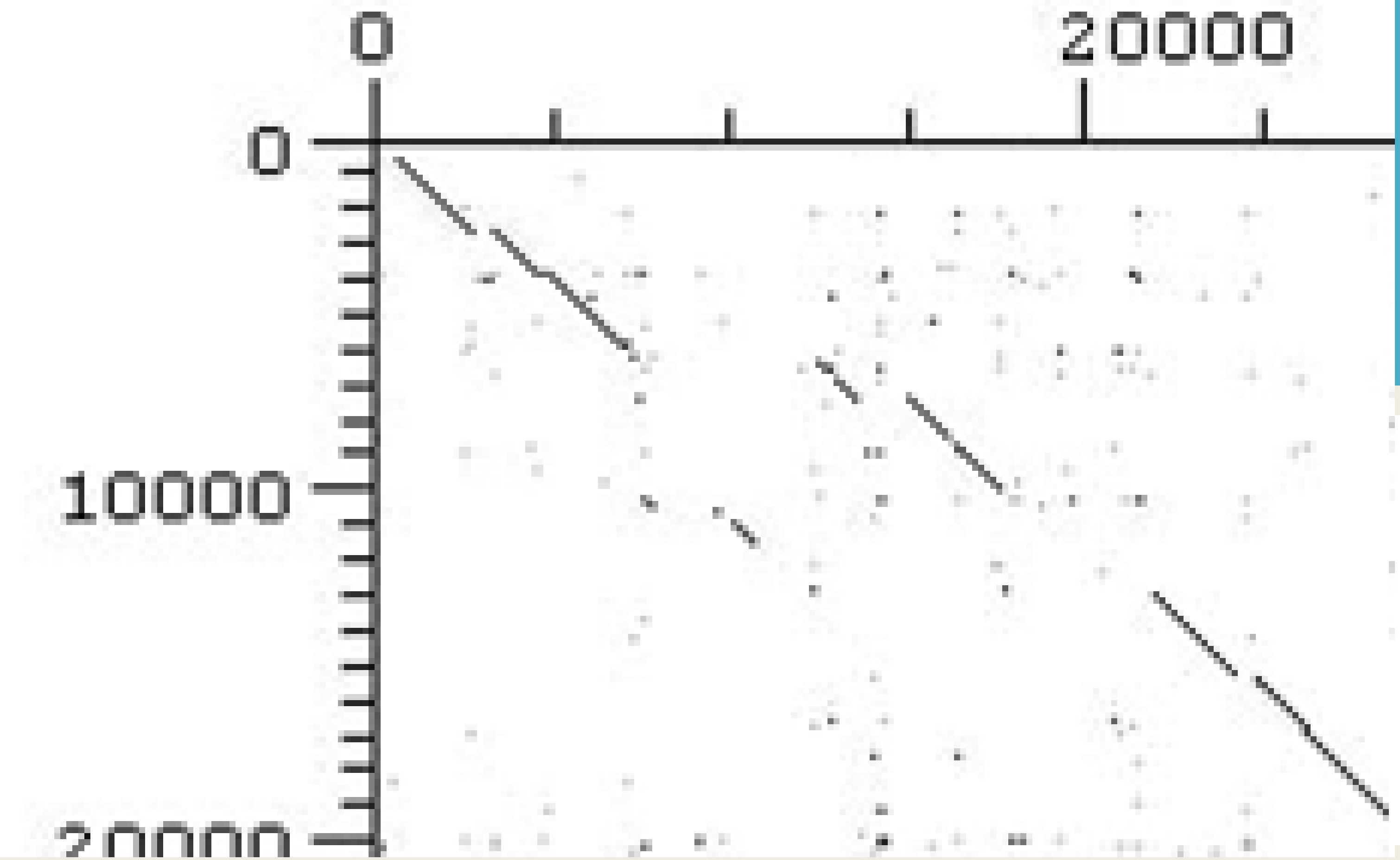
large insertion in common between *Bot1-2* and *Bot1-3*

Bo6L1-15
1010bp

Rearrangement specific of *Bot1-3*

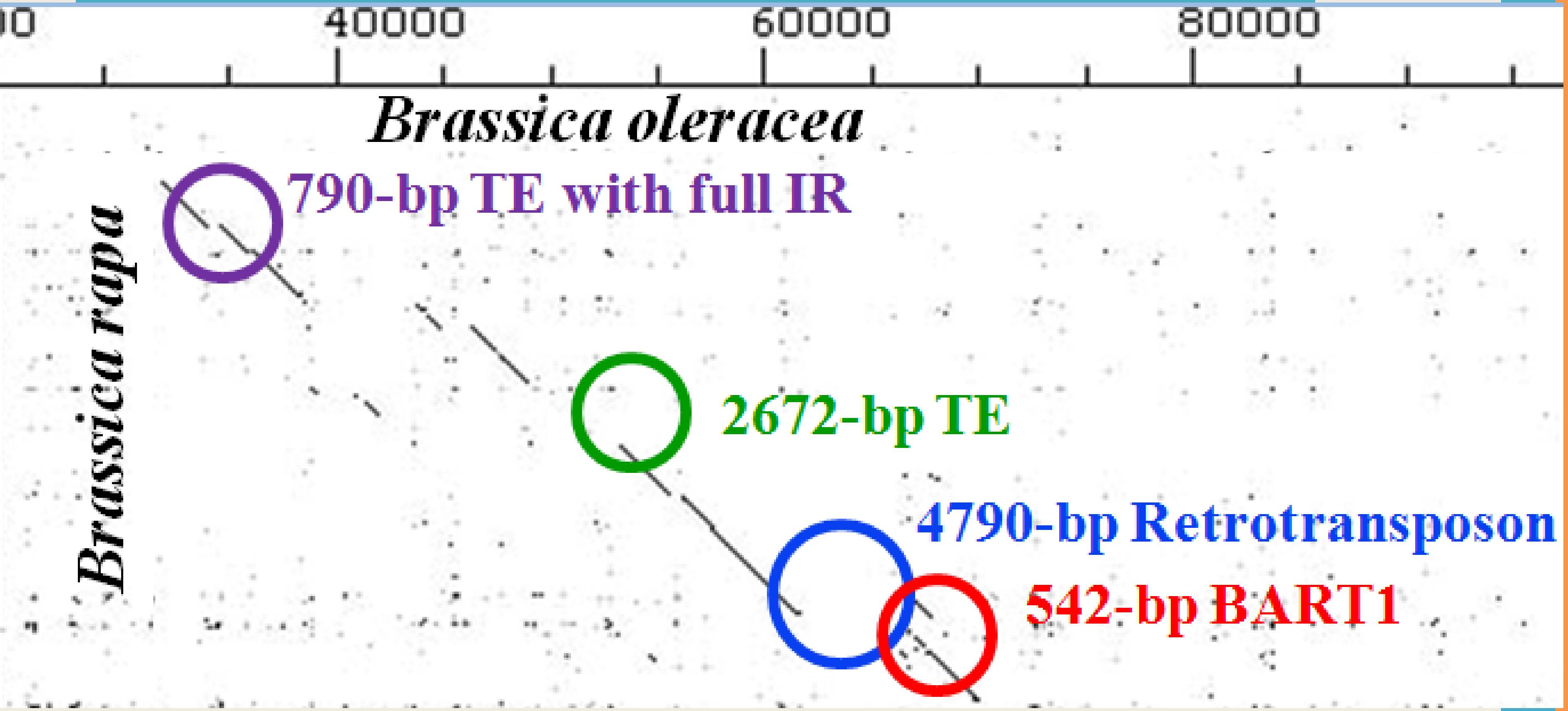


Sequence level at scale of 100s to 10000s of bp – analysis by dotplot



Sequence level at scale of 100s to 1000s of bp – analysis by dotplot

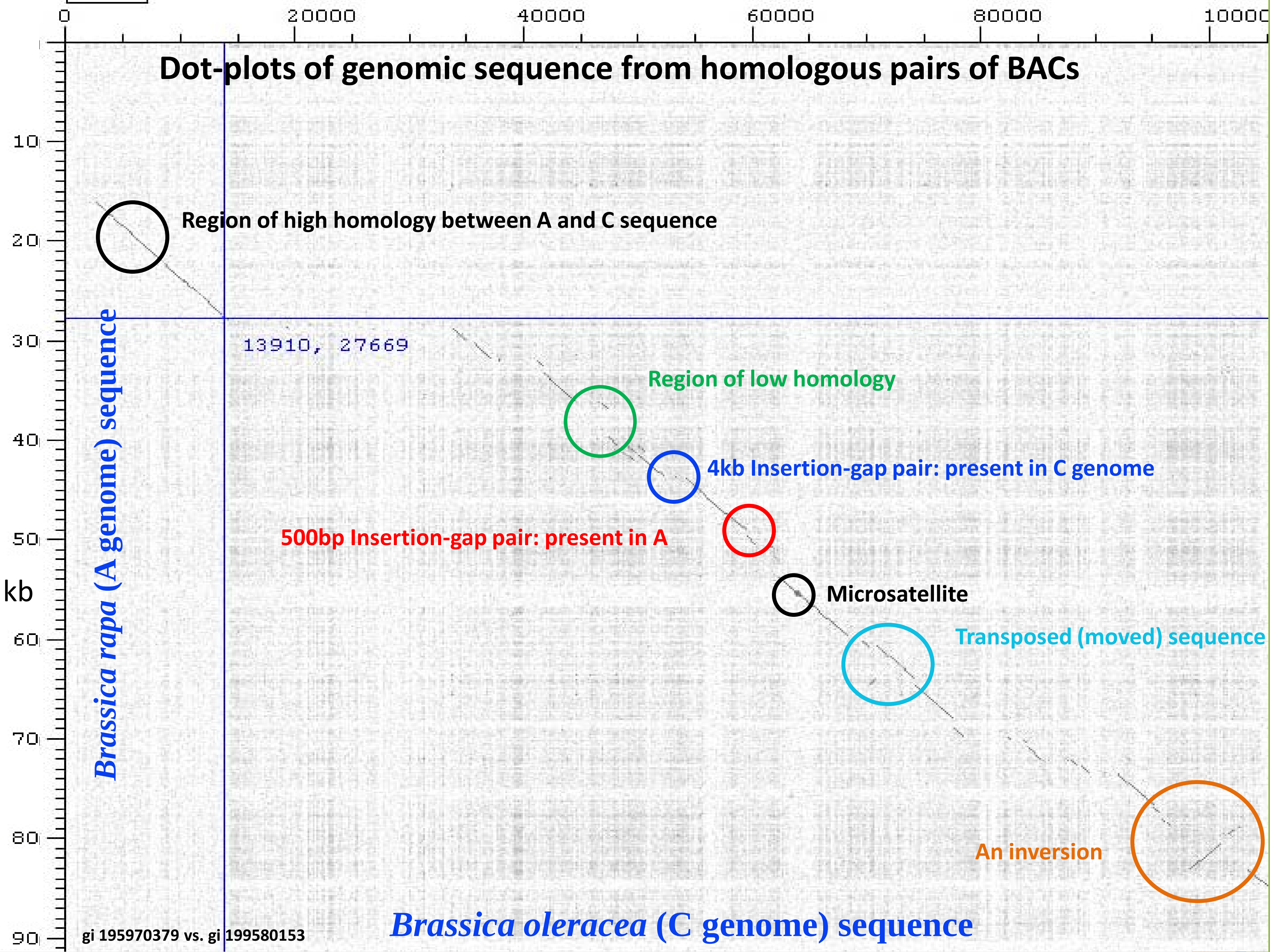
…



Sequence level at scale of 100s to 1000s of bp – analysis by dotplot

…

Dot-plots of genomic sequence from homologous pairs of BACs



Region of high homology between A and C sequence

Region of low homology

4kb Insertion-gap pair: present in C genome

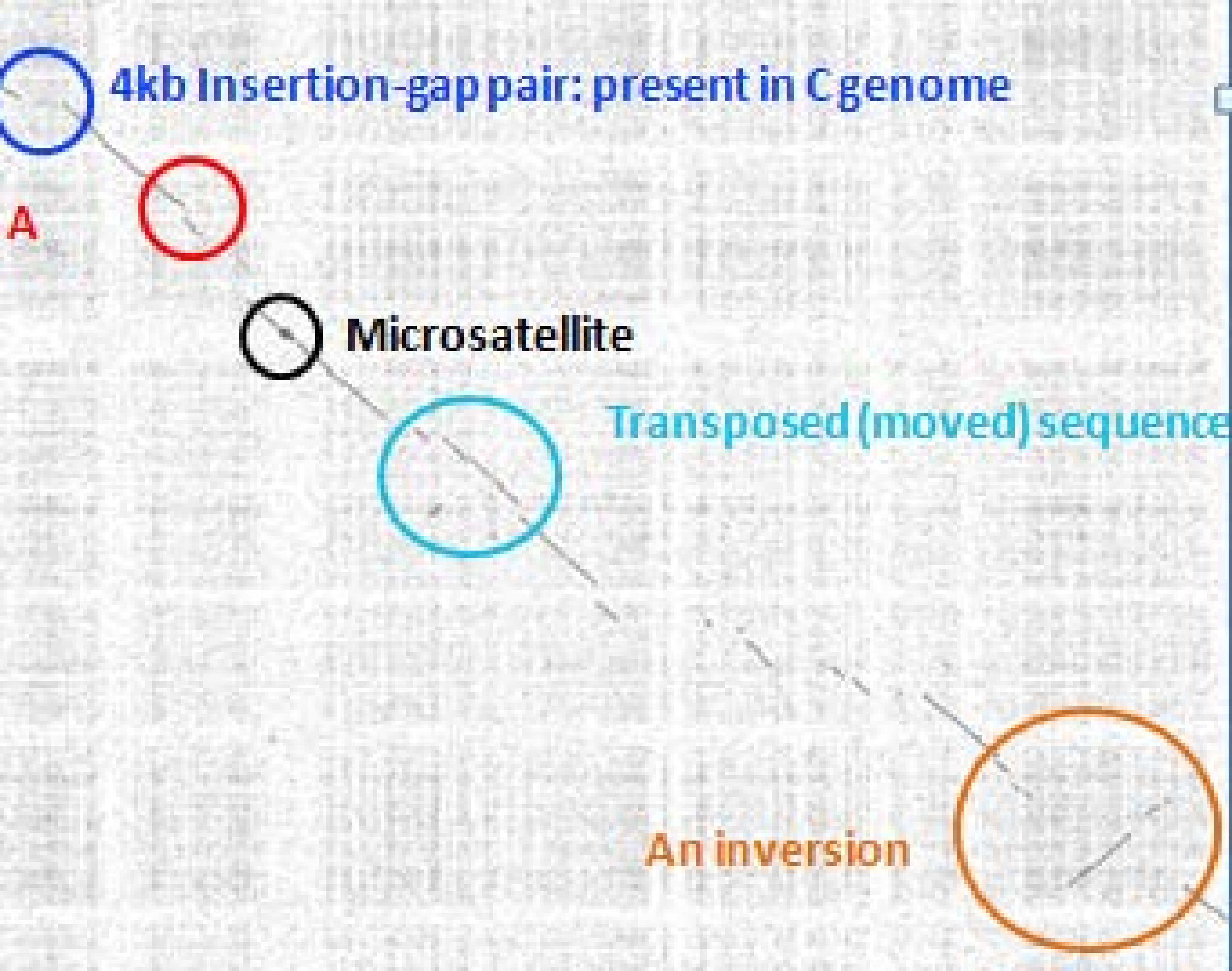
500bp Insertion-gap pair: present in A

Microsatellite

Transposed (moved) sequence

An inversion

gi 195970379 vs. gi 199580153



4kb Insertion-gap pair: present in C genome

Microsatellite

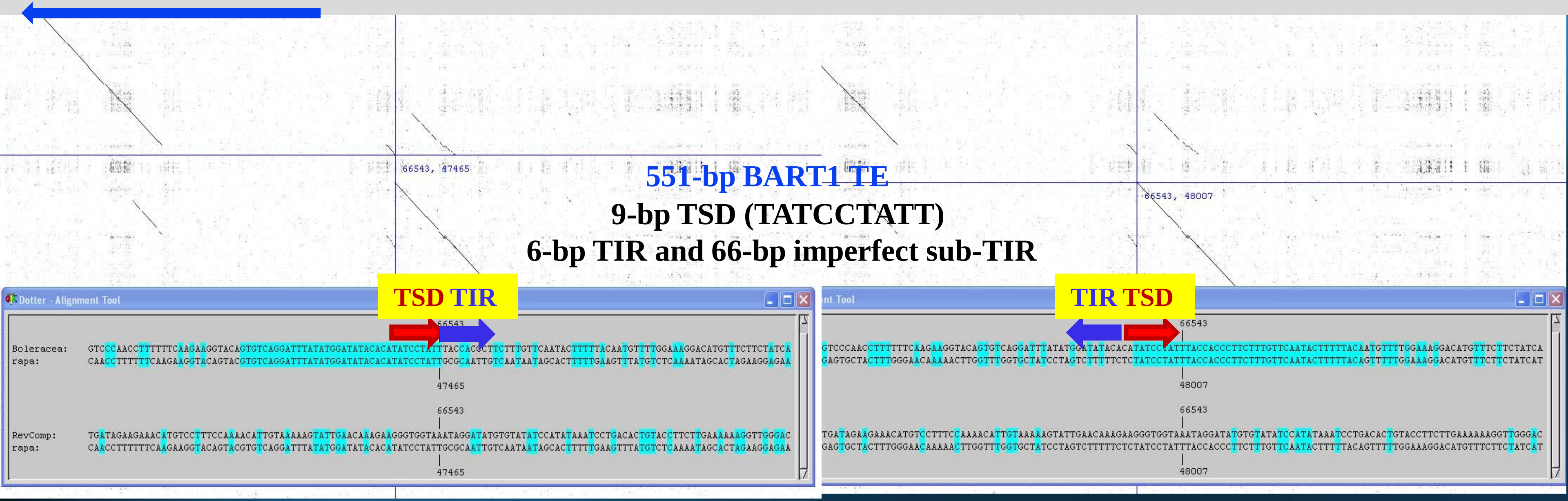
Transposed (moved) sequence

An inversion

AAGTGAATGGATGCTCGCATTAGTTACTATGAGCCGATTCTCGCTCTTGCGAAAGCTAAAGAGGAAAAGGCCTTCGCATTGCAGAAG
AGCTGGCTGCCAGCGAGCAAGAGGTTTTCAATATTGGCTTGTGGAAAATTTGTTGCCACTTTTGCTTTACTAAGGAATGAAATAATAC
TTGTTTTTTTTTTTCATGGTTAATATTAGAAGATATAATTTCCTTTGAAGTTAGATTACGTTTCTTTATGTCGACGAAGTGAAGAAATATT
GTCTTGTTTATGGTTCCTTCTAGTCCCAACCTTTTTTCAAGAAGGTACAGTACGTGTCAGGATTTATATGGATATACACA

TATCCTATTGCGCAATTGTCAATAATAGCACTTTTTGAAGTTTATGTCTCAAAATAGCACTAGAAGGAGAAAGTCACAAAAATGATATT
CATTAAAGGGTAAAATATCTCTTATATCCTTGGTTTAAAATTAAATAAACAAACAAAAAATAAATAAAAAATAAATAAAAAAAAATGAAAAAA
AAGAAATTTTTTTTATAGTTTCAGATTATATGTTTTCAGATTCGATTTTTTTTTTATTTTTTTTATTTTTTTTCGAAATTTTTTTTTTTATTTTTTTTTCA
AATTTTCTTTTTATAATTTAAAAATACTTTTTGAAACTGTTTTTTTAATTTTTATTTTTTTATTTTAGTATTTATTTTTTATAAAATTTTAAACCCT
AATTCCTAAACCCCCACCCCTTAACTCTAAACCCTAAGGTTTGGATTAATTAACCCAATGGATATAAGTGTATATTTACCTCTTTAATGA
AACCTATTTTTGTGACTTTGAATCTTGAGTGCTACTTTGGGAACAAAAACTTGGTTTGGTGCTATCCTAGTCTTTTTCTC**TATCCTATT**

TACCACCCTTCTTTGTTCAATACTTTTTACAGTTTTTGGAAAGGACATGTTTCTTCTATCATCACTTAATGGTTATATATGTATGAGAAG
TTTGAAAGAGATTACACTGTTTTGGAATATTAAAAAAAAAAAGATATTACAAGATCTGATTTTGTTTGTATTTTAAAATTCTACCAAATC
TCTCCTCAAAATCTTGGTCAAAGTCCAAAAATCCAAATATCTCAGTTAAATTCCACCAAATATGAAATCCTAAAACTTTTCCAAAATA
GTTCAATAAGCCCTTAGTGTTTGGTG



551-bp BART1 TE

9-bp TSD (TATCCTATT)

6-bp TIR and 66-bp imperfect sub-TIR

```
                                                                     66543
                                                                       |
Boleracea:   GTCCCAACCTTTTTTCAAGAAGGTACAGTGTCAGGATTTATATGGATATACACATATCCTATTTACCACCCTTCTTTGTTCAATACTTTTTACAATGTTTTGGAAAGGACATGTTTCTTCTATCA
rapa:        CAACCTTTTTTCAAGAAGGTACAGTACGTGTCAGGATTTATATGGATATACACATATCCTATTGCGCAATTGTCAATAATAGCACTTTTTGAAGTTTATGTCTCAAAATAGCACTAGAAGGAGAA
                                                                       |
                                                                     47465

                                                                     66543
                                                                       |
RevComp:     TGATAGAAGAAACATGTCCTTTCCAAAACATTGTAAAAGTATTGAACAAAGAAGGGTGGTAAATAGGATATGTGTATATCCATATAAATCCTGACACTGTACCTTCTTGAAAAAAGGTTGGGAC
rapa:        CAACCTTTTTTCAAGAAGGTACAGTACGTGTCAGGATTTATATGGATATACACATATCCTATTGCGCAATTGTCAATAATAGCACTTTTTGAAGTTTATGTCTCAAAATAGCACTAGAAGGAGAA
                                                                       |
                                                                     47465
```

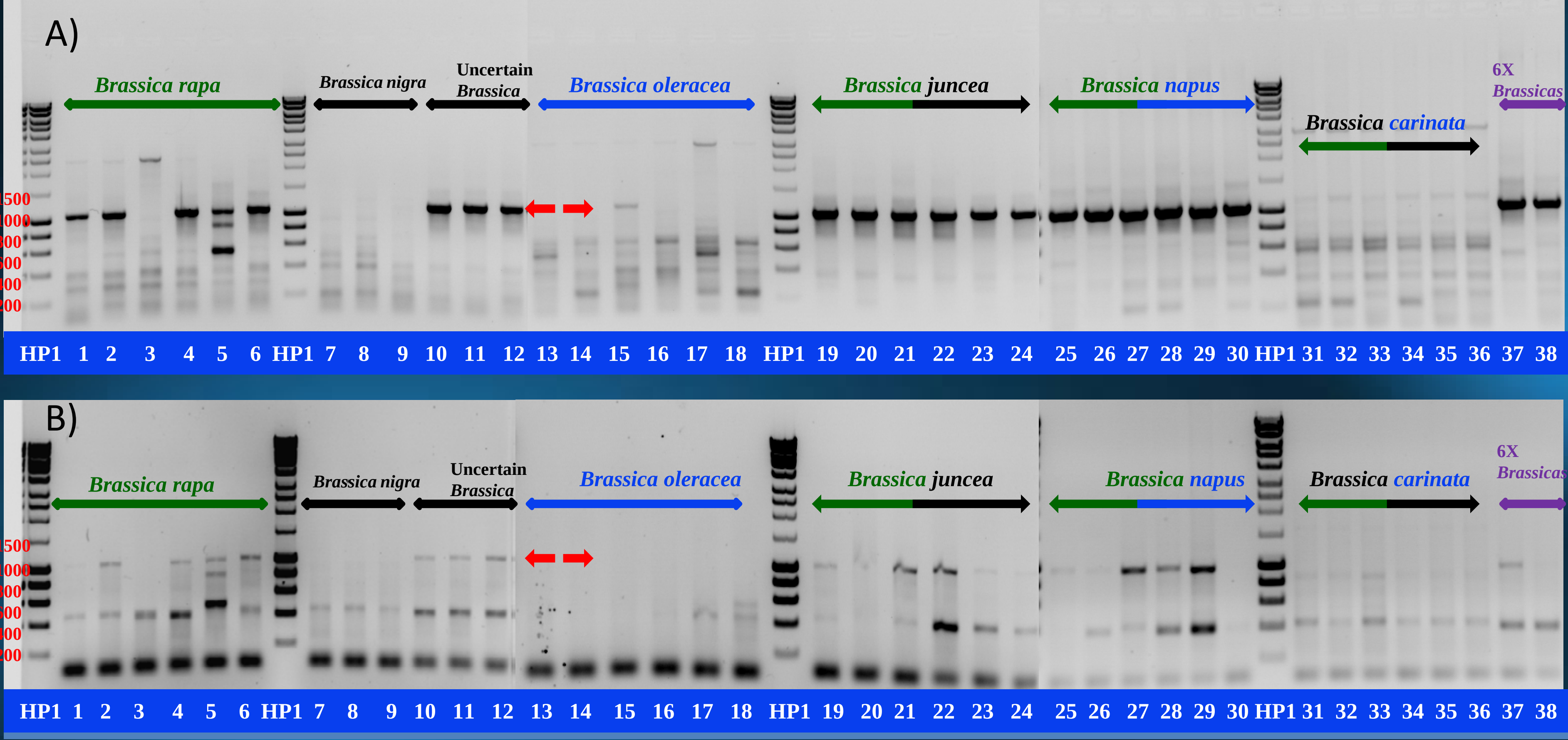
```
                                                                     66543
                                                                       |
GTCCCAACCTTTTTTCAAGAAGGTACAGTGTCAGGATTTATATGGATATACACATATCCTATTTACCACCCTTCTTTGTTCAATACTTTTTACAATGTTTTGGAAAGGACATGTTTCTTCTATCA
GAGTGCTACTTTGGGAACAAAAACTTGGTTTGGTGCTATCCTAGTCTTTTTCTCTATCCTATTTACCACCCTTCTTTGTTCAATACTTTTTACAGTTTTTGGAAAGGACATGTTTCTTCTATCAT
                                                                       |
                                                                     48007

                                                                     66543
                                                                       |
TGATAGAAGAAACATGTCCTTTCCAAAACATTGTAAAAGTATTGAACAAAGAAGGGTGGTAAATAGGATATGTGTATATCCATATAAATCCTGACACTGTACCTTCTTGAAAAAAGGTTGGGAC
GAGTGCTACTTTGGGAACAAAAACTTGGTTTGGTGCTATCCTAGTCTTTTTCTCTATCCTATTTACCACCCTTCTTTGTTCAATACTTTTTACAGTTTTTGGAAAGGACATGTTTCTTCTATCAT
                                                                       |
                                                                     48007
```

***Brassica rapa* with inserted 542bp sequence not present in *B. oleracea.* 9bp TSD (red letters and arrow) and TIR (blue). Flanking primers used in PCR (next slide) as blue arrows on sequence**

Faisal Nouroz 2011

Insertion polymorphism in *Brassica* genomes shown by PCR with flanking primers



Amplification with two primer sets (top and bottom). *B. rapa (AA), B. juncea (AABB)* and *B. napus (AACC)* include the longer fragment with insertion. B and C genomes have only the shorter, lower, fragment without insertion.

2011/01/15

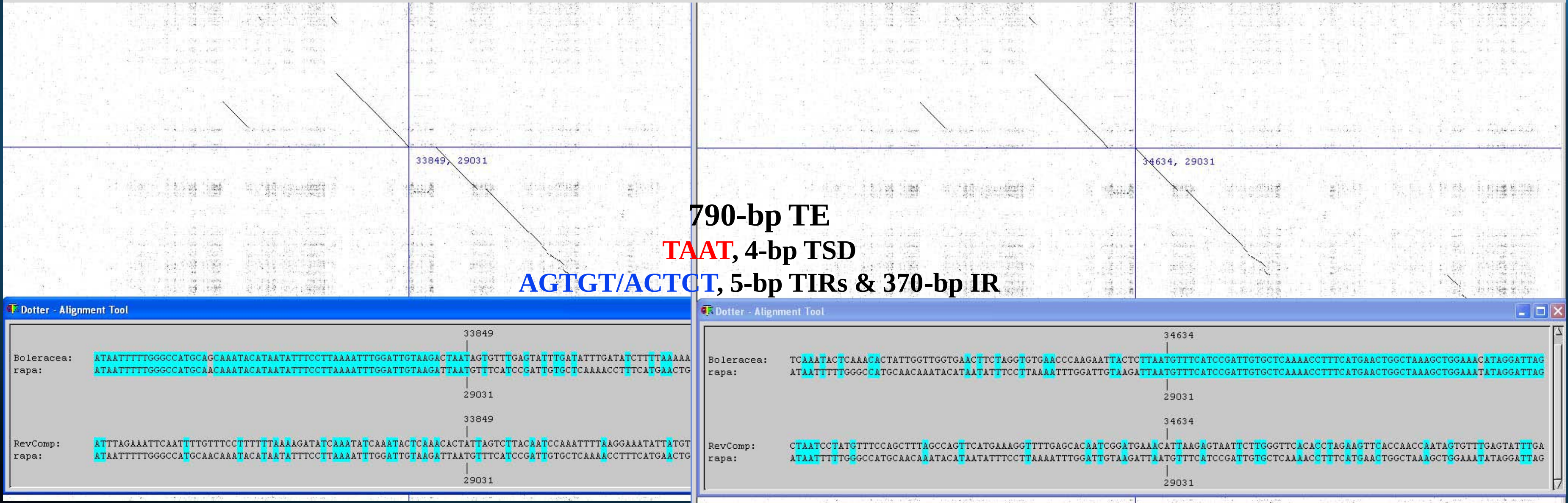
Schematic representation of insertion in Brassica rapa and other Brassica genomes. Green, red, blue and black boxes showing DNA motifs.

GACACTCTTCCCAATCGTTCATTCCTGACGTCATTAGGCAACCACCTCTGTTTTTCCCCACCACAAACAGTGAATACATCTCTCCTATCTCTC
TCAGAATCGTCAGTGTTTGCTCTCCGTTGCTTACTCGCTTCTCTATGAATCCAACTTGCCCCGTCGTTACAAATCTGCCAAAAATAAACCAAA
ACCAGTCCGGTCAATGAAAAAAATGCCAATGTTTCAGGTCTAGAAATTATCCACAACCCTAGTACTAAGATCTGAAATTTATGAGGGAGATAA
ACATTTTTAGGTTAATTGTAAGAAAAAATATT**TATAATTTTTGGGCCATGCA**GCAAATACATAATATTTCCTTAAAATTTGGATTGTAAGAC

TAAT*AGTGT*TTGAGTATTTGATATTTGATATCTTTTAAAAAGGAAACAAAATTGAATTTCTAAATAAGATTATATTTTTAAAATAAAACAAT
AAAAATACATAAAAATAGTTACAAAAAAAAATATATATATTGTTAAACCGTTAGCAAATTAAATACTAAATCCTATACCCTAAATCCTAAACT
CCAAACCCTAAATGATAAACCTTAAATCTTGGATAAACCGTAAACCATTGGAAAATTTTAAAACCTAATCATACATTAAAAACTAAAATTTAA
TAACACTAAACCCTAAACCCTAATCACTAAACCCTAAACCCTTAGATAAATCATGAACCCTTGGATAAATCATAAACTCTAAATCAAAAATAT

TTAAAATTAAACCCTAAAATATATAATTTATCCAAGGGCTCAGAGTTTACCCAAGGGTTTAGGGTTTAGTGATTAGGATTTAGGGTTTAGTGT
TATTAAAATTTAGTTTTTAATGTATGATCTAAGGTTTAAGAGTTTCCAATGGTTTAGAGTTTATCCAAAGTTTAAGGTTTAACGTTTAGGGTT
TAGGATTTAGGATTTAAGGTATAGGGTTTAGTATTTTGCTGAAGATTTAACAATATTAATTAATTTATTTTTTGTAACTATTTTTATATATTT
TTATTATTTTATTTTTAAAATATAATATAATTTGGATATTCAATTTTATTTTCTTTTTTAAAAAATATCAAATATCAAATACTCAAACACTAT

GGTTGGTGAACTTCTAGGTGTGAACCCAAGAATT***ACTCT*TAAT**GTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTGGCTAAAGCTGGAA
ACATAGGATTAGTAAGAAGTAGAATCTTGTAAAGTACCTGTTATAGTATTCCTCTAAGAAAGTTCGATCAG**TTTCGTCGTTTGTCTGATCG**TT
ACCAACAATCTCCATCAAAACATCGTTGTTTTCTTTGGTCACCGCGTCTCCGACAAGATTCTCTGTCTCCGAGCCATAAGCGACAAACTGTAT
GATAGTGAGGTGAATCTGAGAGTTATTGATAAGCCACTGGCACAAGGACAGAGCCTCTCGATCATCAGGACCACCAAAGAACAATGCAGCGAC
GTGTTGTACCGACTCAAACCCGTGAAGCTGGTGGAACCCGGTTATGTTTCTATCCACATAGATACCGATCG



790-bp TE
TAAT, 4-bp TSD
AGTGT/ACTCT, 5-bp TIRs & 370-bp IR

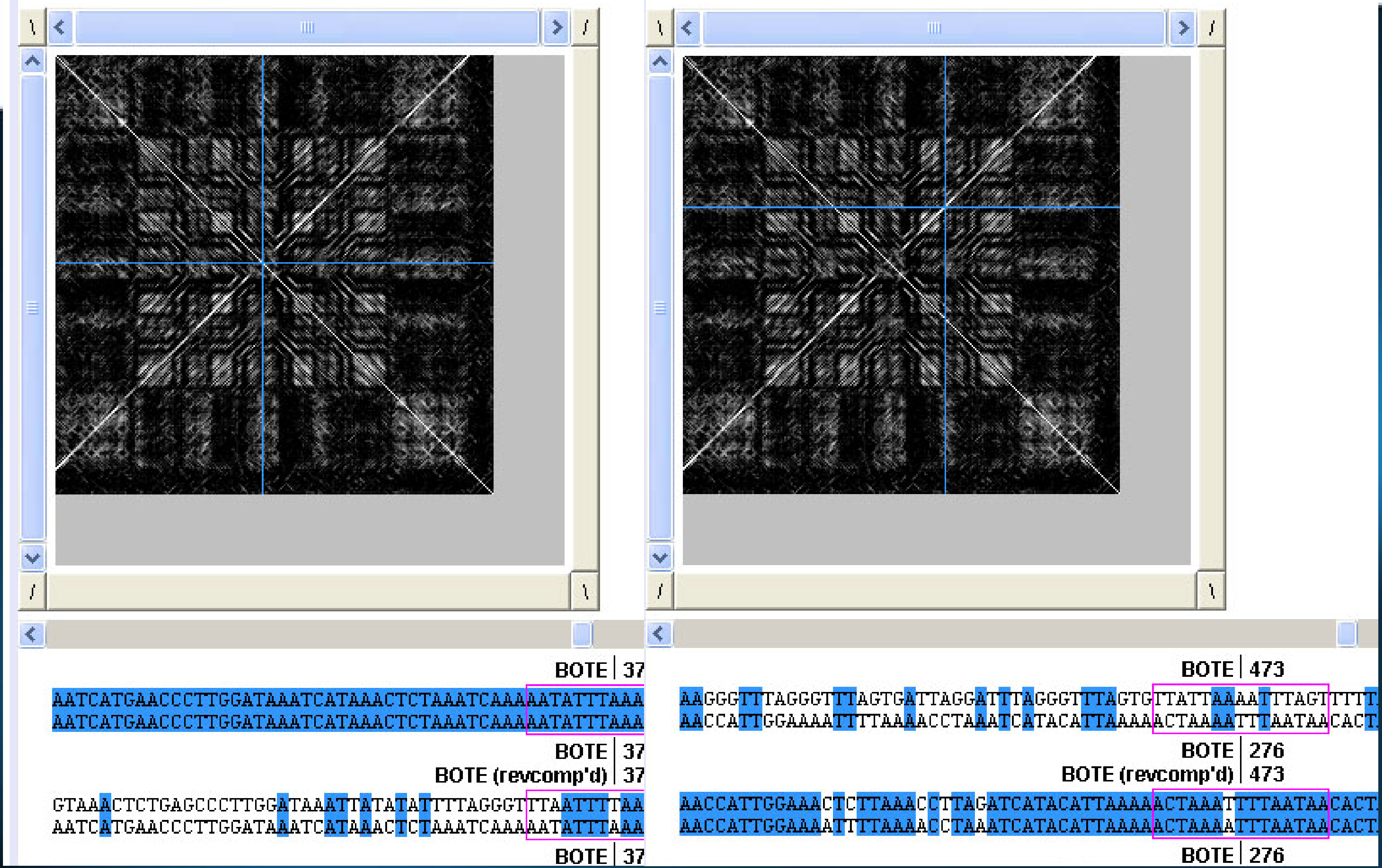
```
                                                    33849
                                                      |
Boleracea:  ATAATTTTTGGGCCATGCAGCAAATACATAATATTTCCTTAAAATTTGGATTGTAAGACTAATAGTGTTTGAGTATTTGATATTTGATATCTTTTAAAAA
rapa:       ATAATTTTTGGGCCATGCAACAAATACATAATATTTCCTTAAAATTTGGATTGTAAGATTAATGTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTG
                                                      |
                                                    29031

                                                    33849
                                                      |
RevComp:    ATTTAGAAATTCAATTTTGTTTCCTTTTTTAAAAGATATCAAATATCAAATACTCAAACACTATTAGTCTTACAATCCAAATTTTAAGGAAATATTATGT
rapa:       ATAATTTTTGGGCCATGCAACAAATACATAATATTTCCTTAAAATTTGGATTGTAAGATTAATGTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTG
                                                      |
                                                    29031
```

```
                                                                           34634
                                                                             |
Boleracea:  TCAAATACTCAAACACTATTGGTTGGTGAACTTCTAGGTGTGAACCCAAGAATTACTCTTAATGTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTGGCTAAAGCTGGAAACATAGGATTAG
rapa:       ATAATTTTTGGGCCATGCAACAAATACATAATATTTCCTTAAAATTTGGATTGTAAGATTAATGTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTGGCTAAAGCTGGAAATATAGGATTAG
                                                                             |
                                                                           29031

                                                                           34634
                                                                             |
RevComp:    CTAATCCTATGTTTCCAGCTTTAGCCAGTTCATGAAAGGTTTTGAGCACAATCGGATGAAACATTAAGAGTAATTCTTGGGTTCACACCTAGAAGTTCACCAACCAATAGTGTTTGAGTATTTGA
rapa:       ATAATTTTTGGGCCATGCAACAAATACATAATATTTCCTTAAAATTTGGATTGTAAGATTAATGTTTCATCCGATTGTGCTCAAAACCTTTCATGAACTGGCTAAAGCTGGAAATATAGGATTAG
                                                                             |
                                                                           29031
```

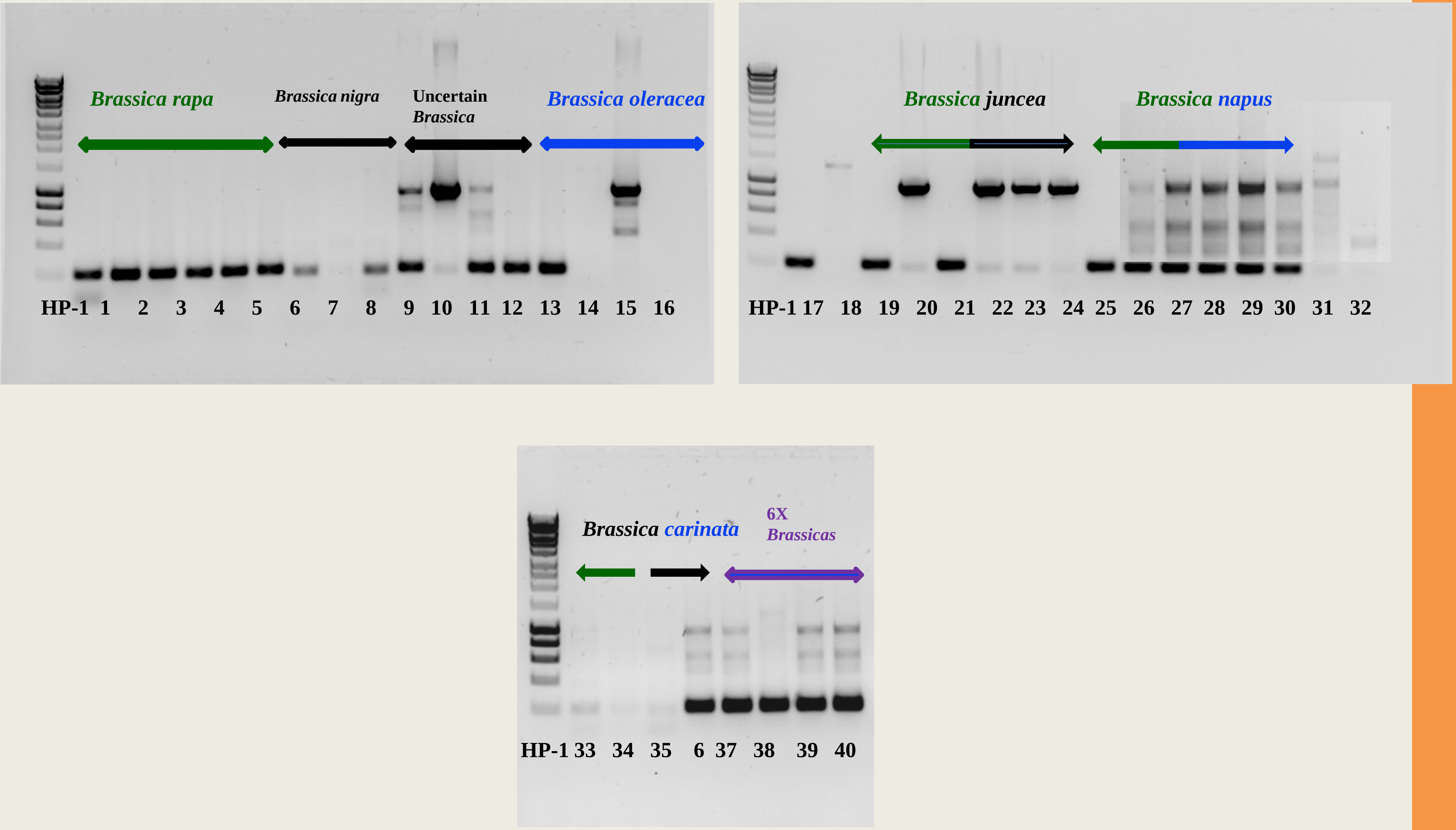
Insertion sequence present in *B. oleracea,* missing from *B. rapa*. TSD (red); green and blue boxes shows remarkable internal structure with 370-bp inverted repeat near-filling the insert.



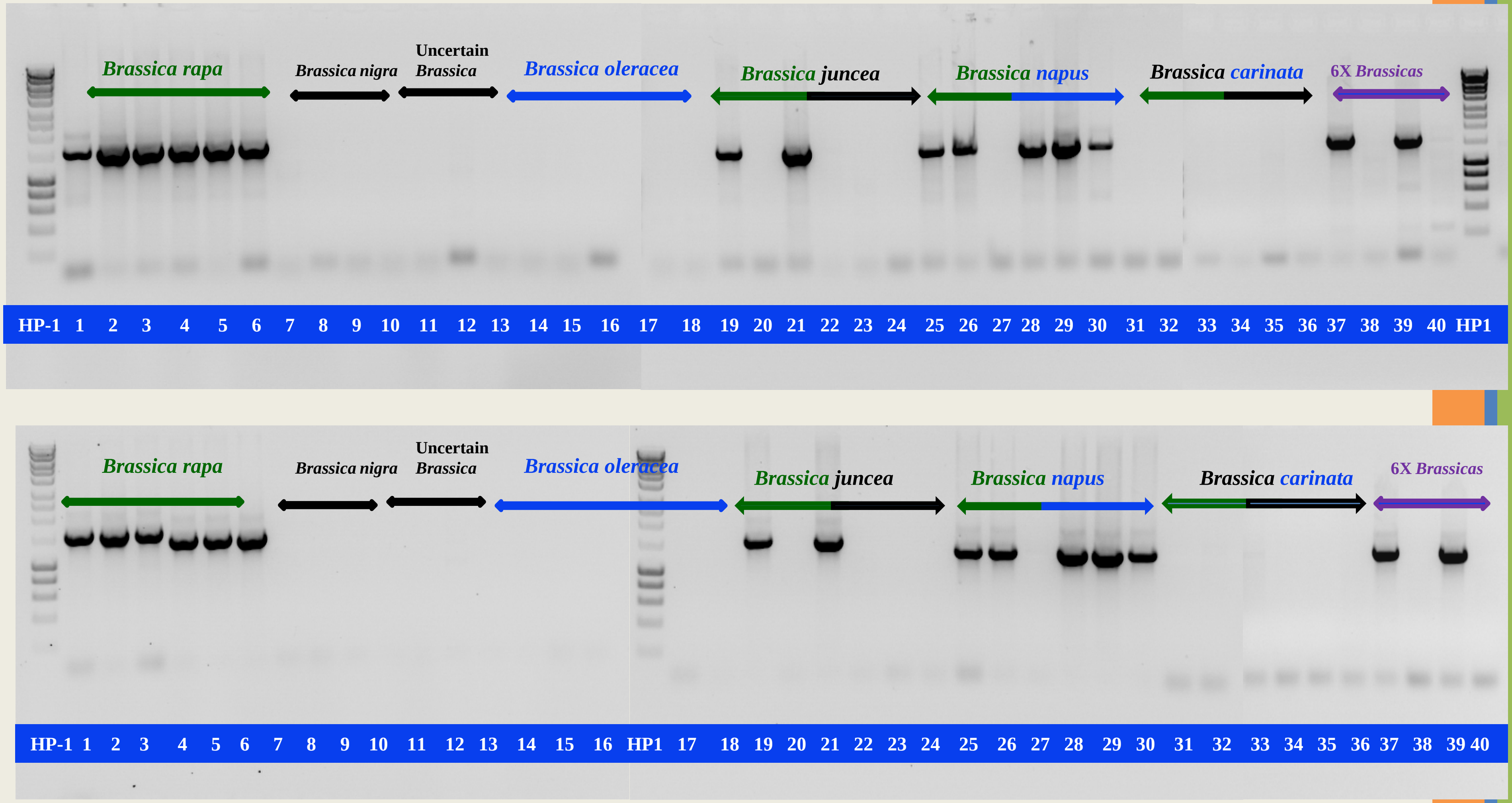
```
BOTE | 37
AATCATGAACCCTTGGATAAATCATAAACTCTAAATCAAAAATATTTAAA
AATCATGAACCCTTGGATAAATCATAAACTCTAAATCAAAAATATTTAAA
BOTE | 37
BOTE (revcomp'd) | 37
GTAAACTCTGAGCCCTTGGATAAATTATATATTTTAGGGTTTAATTTTAA
AATCATGAACCCTTGGATAAATCATAAACTCTAAATCAAAAATATTTAAA
BOTE | 37
```

```
BOTE | 473
AAGGGTTTAGGGTTTAGTGATTAGGATTTAGGGTTTAGTGTTATTAAAATTTAGTTTTT
AACCATTGGAAAATTTTAAAACCTAAATCATACATTAAAAACTAAAATTTAATAACACT
BOTE | 276
BOTE (revcomp'd) | 473
AACCATTGGAAACTCTTAAACCTTAGATCATACATTAAAAACTAAATTTTAATAACACT
AACCATTGGAAAATTTTAAAACCTAAATCATACATTAAAAACTAAAATTTAATAACACT
BOTE | 276
```

Dotplot of 790bp insertion element showing inverted repeat structure.

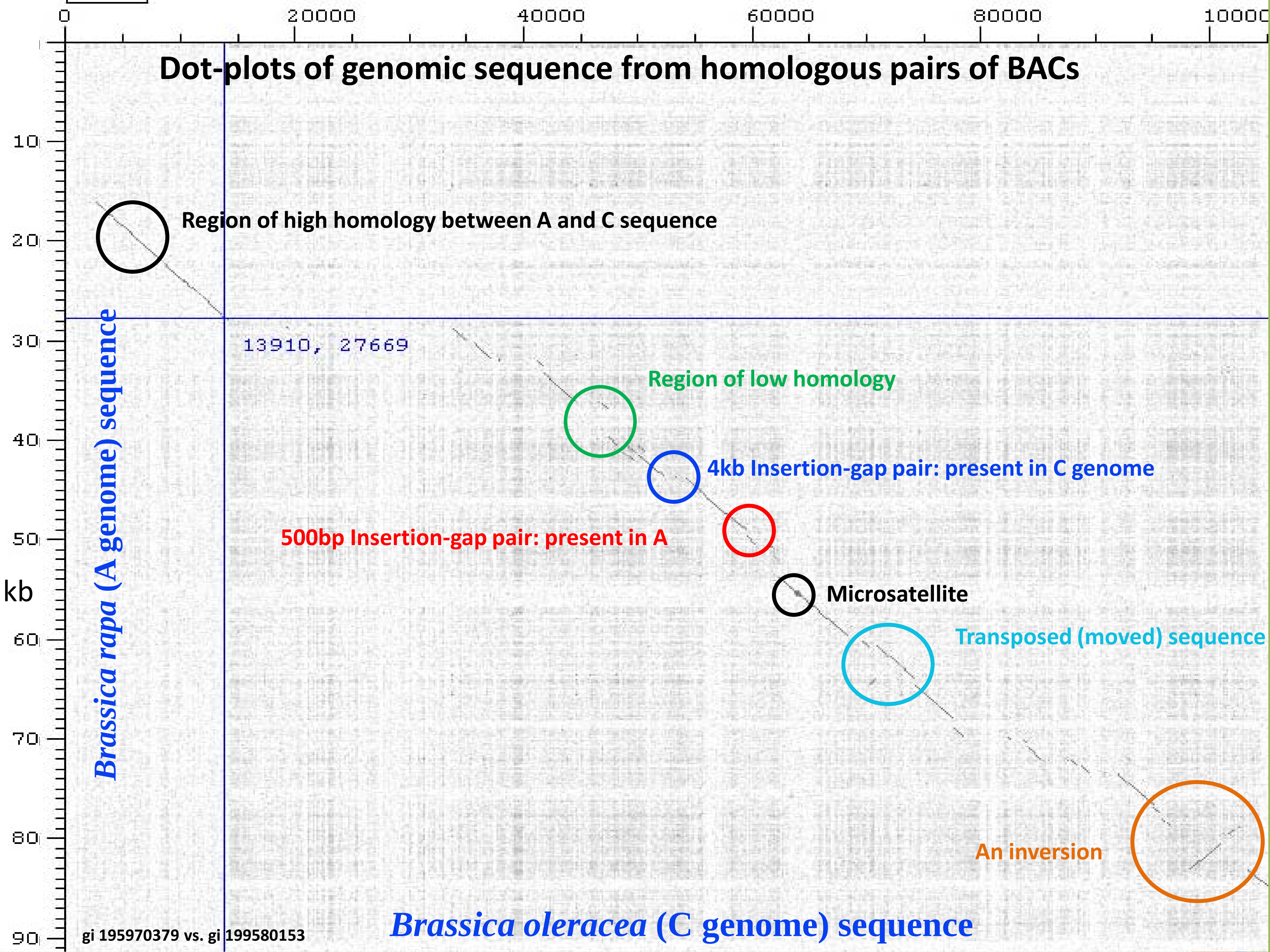


Amplification of 790-bp TE



1st and last part of 2672 bp TE (seems mariner) in *Brassica rapa*

Dot-plots of genomic sequence from homologous pairs of BACs



Region of high homology between A and C sequence

Region of low homology

4kb Insertion-gap pair: present in C genome

500bp Insertion-gap pair: present in A

Microsatellite

Transposed (moved) sequence

An inversion

gi 195970379 vs. gi 199580153

Genomic divergence outside genes

How do the genomes of related species differ?

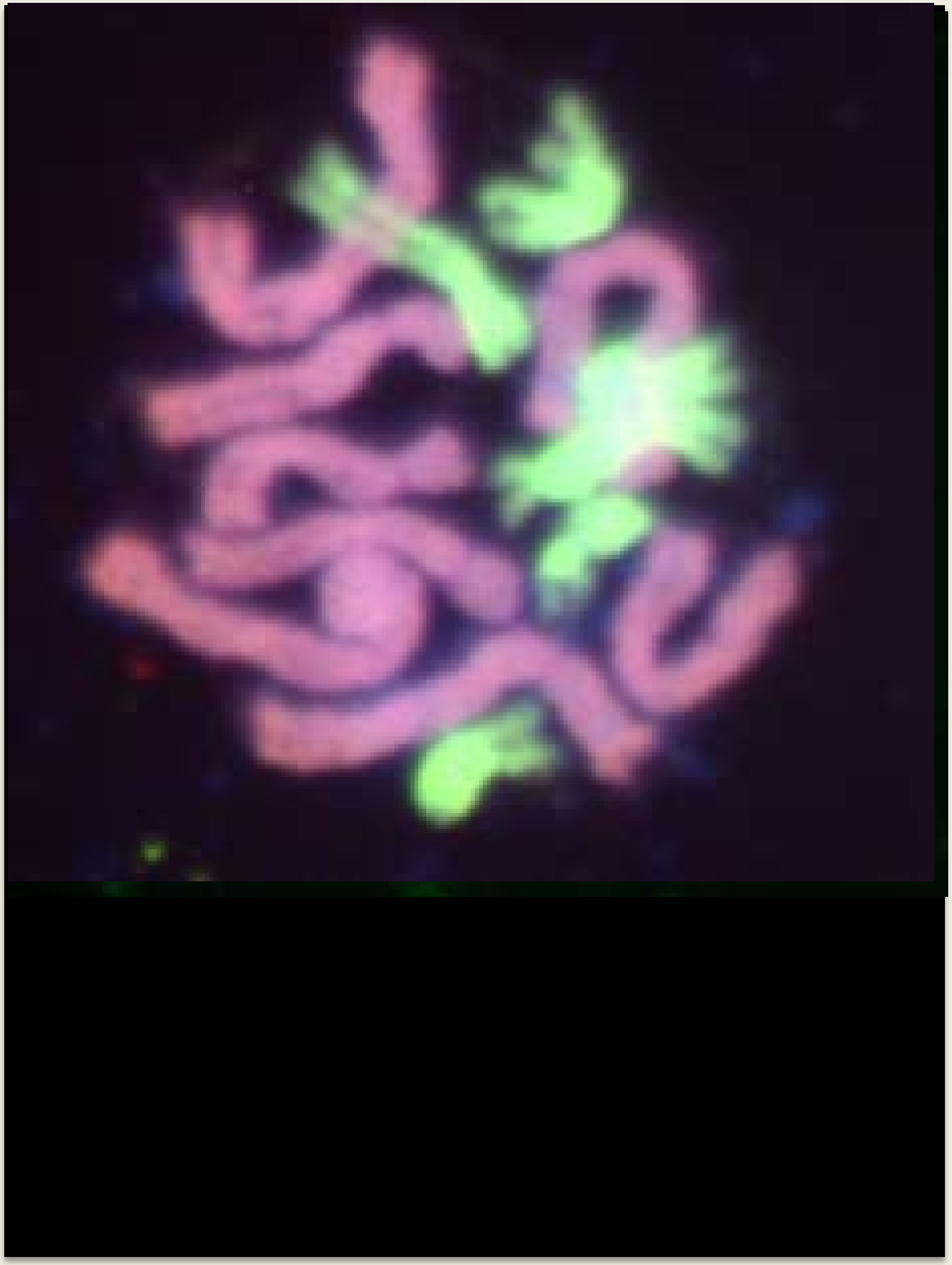
- -Chromosome number
- -Ploidy
- -Transposable elements
- -Microsatellites
- -Satellite DNA sequences
- -Gene sequences
- -Chromosome rearrangements

- -Mechanisms
- -TSD target site duplications various lengths
- -Sometimes TIR terminal inverted repeats
- -Very variable and many uncharacterized ‘families’/mechanisms

- -Why? and consequences

Pat Heslop-Harrison, Faisal Nouroz, Trude Schwarzacher
and Farah Badakshi, Ana Claudia Guerra, Guto Kuhn, John Bailey

phh4@le.ac.uk www.molcyt.com

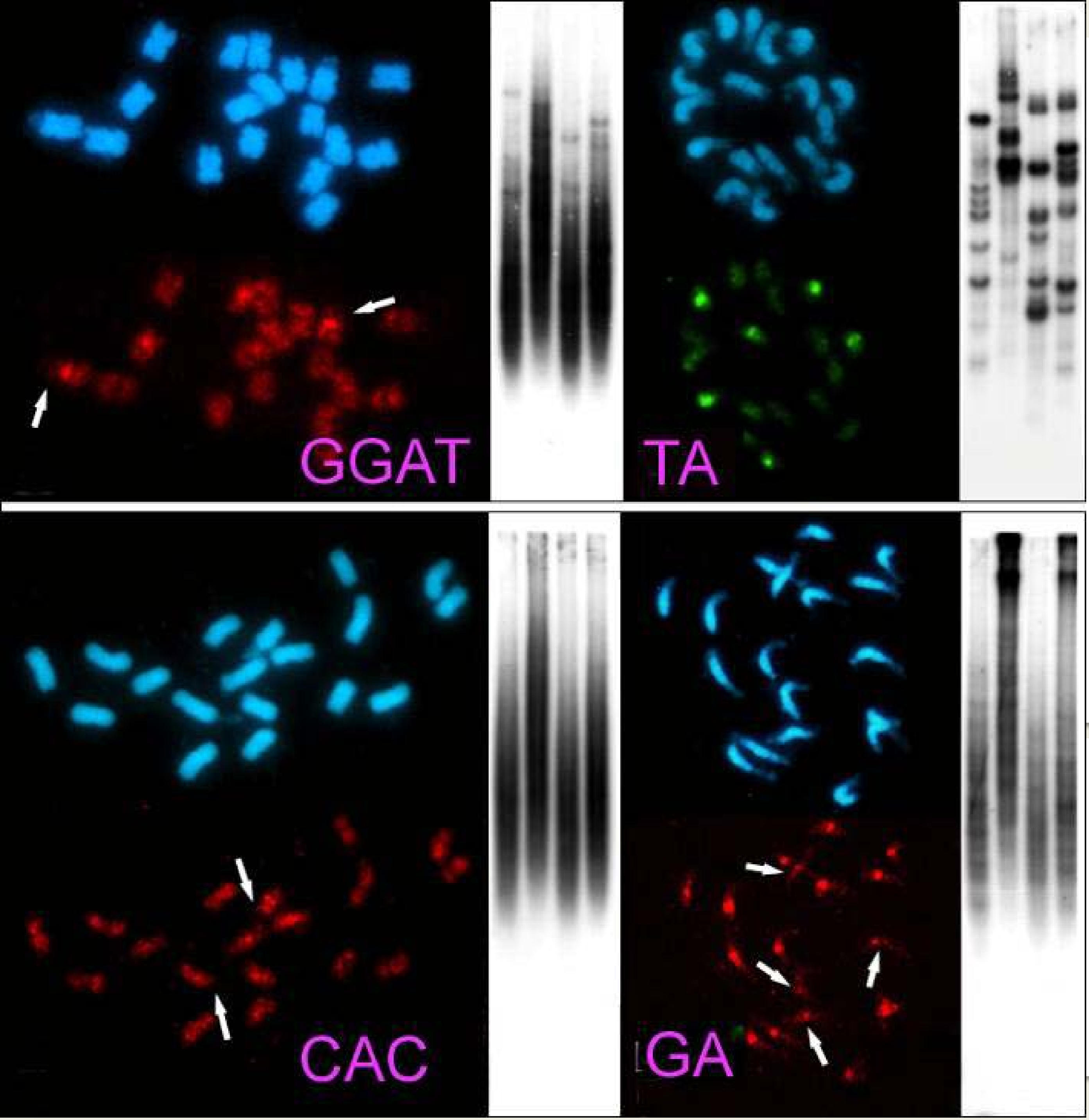


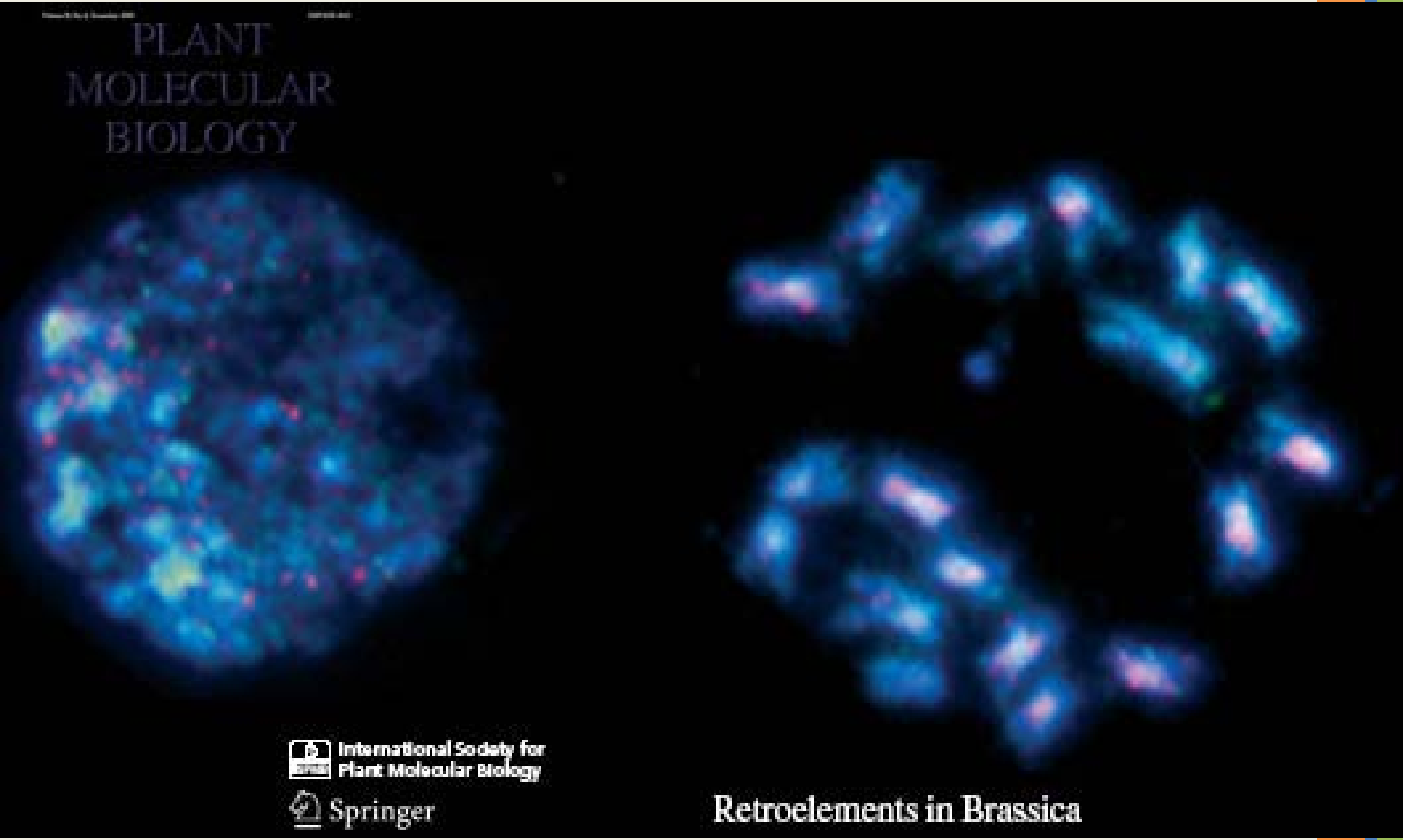
How do the genomes of related species differ?

- Chromosome number
- Ploidy
- Transposable elements
- Microsatellites
- Satellite DNA sequences
- Gene sequences
- Chromosome rearrangements

- Mechanisms

- Why? and consequences


GGAT
TA
CAC
GA

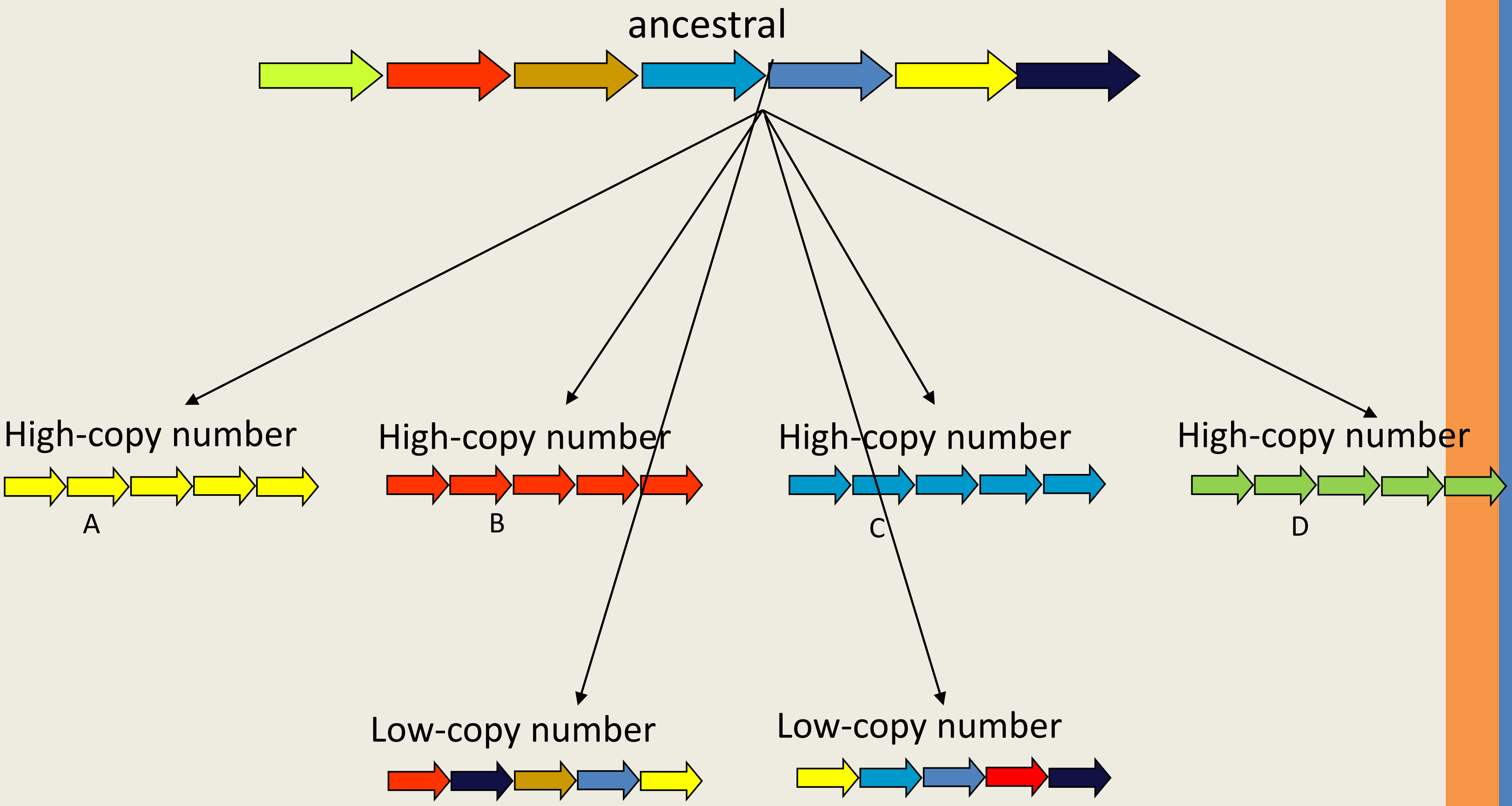


PLANT
MOLECULAR
BIOLOGY

International Society for
Plant Molecular Biology

Springer

Retroelements in Brassica



ancestral

High-copy number

A

High-copy number

B

High-copy number

C

High-copy number

D

Low-copy number

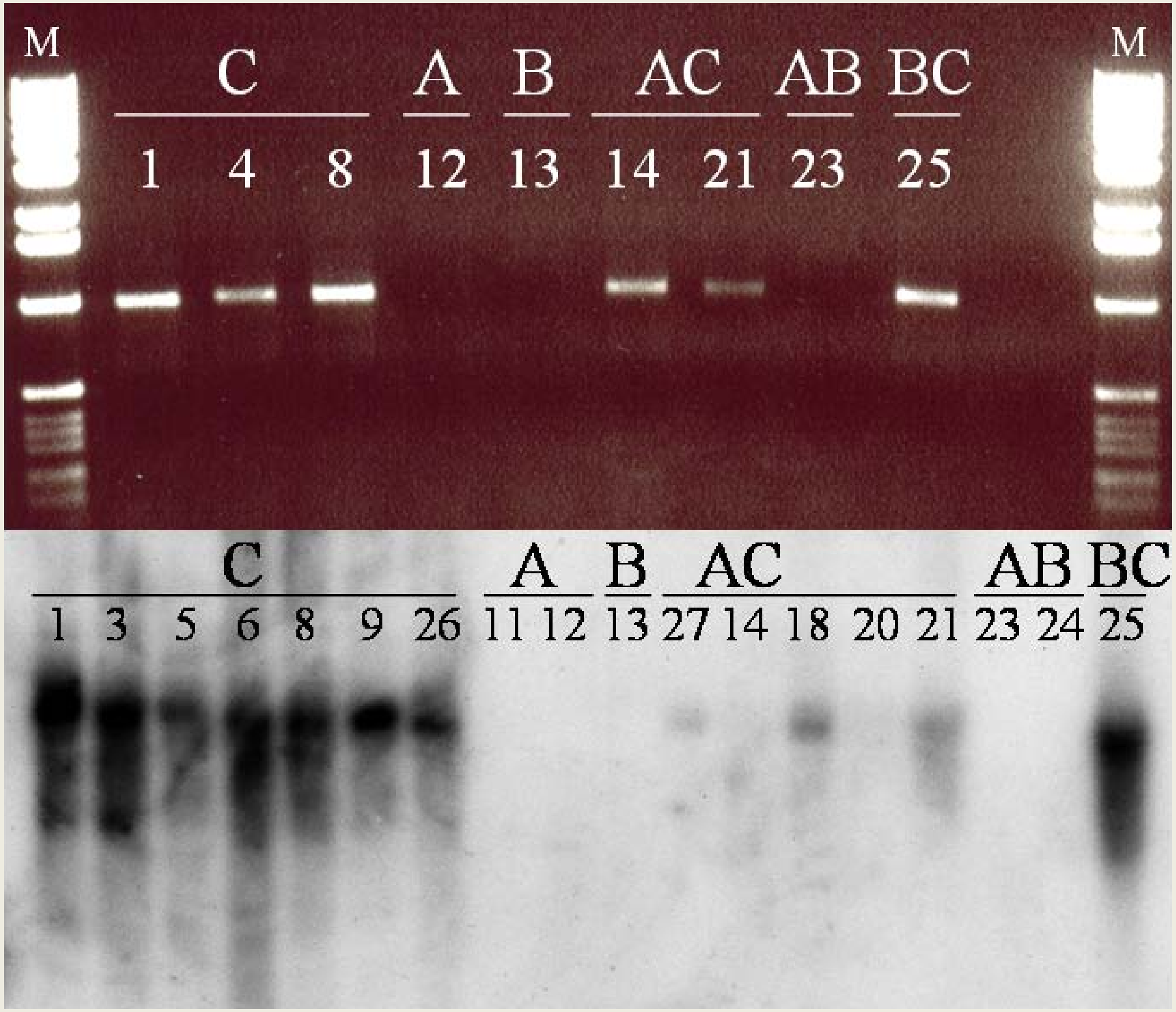
Low-copy number

High copy spp: homogenized, amplification from a limited number of master copies

Low copy spp: much variation

Genome Specificity of a CACTA (*En/Spm*) Transposon

B. napus (AACC, 2n=4x=38) B. oleracea (CC, 2n=2x=18) B. rapa (AA, 2n=2x=20)



Genome Specificity of a CACTA (En/Spm) Transposon

- *Bot1* has encountered several rounds of amplification in the C (*B. oleracea)* genome
only, playing a major role in the recent *B. rapa*
and *B. oleracea* genome divergence
- *Bot1* carries a host S-locus associated *SLL3* gene copy; is the transposon associated with *SLL3* proliferation?

→ Transposons are a driver of genome and genome evolution

Alix et al. The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. Plant Journal December 2008