
Article

Development of retrotransposon-based molecular markers for characterization of *Persea americana* (avocado) horticultural races.

Mario González Carracedo, Samuel Bello Alonso, Rahil Brito Cabrera, David Jiménez Arias and José Antonio Pérez Pérez.

Supplementary material

Figure S1. Description of iPBS technique and pipeline for potential LTR identification

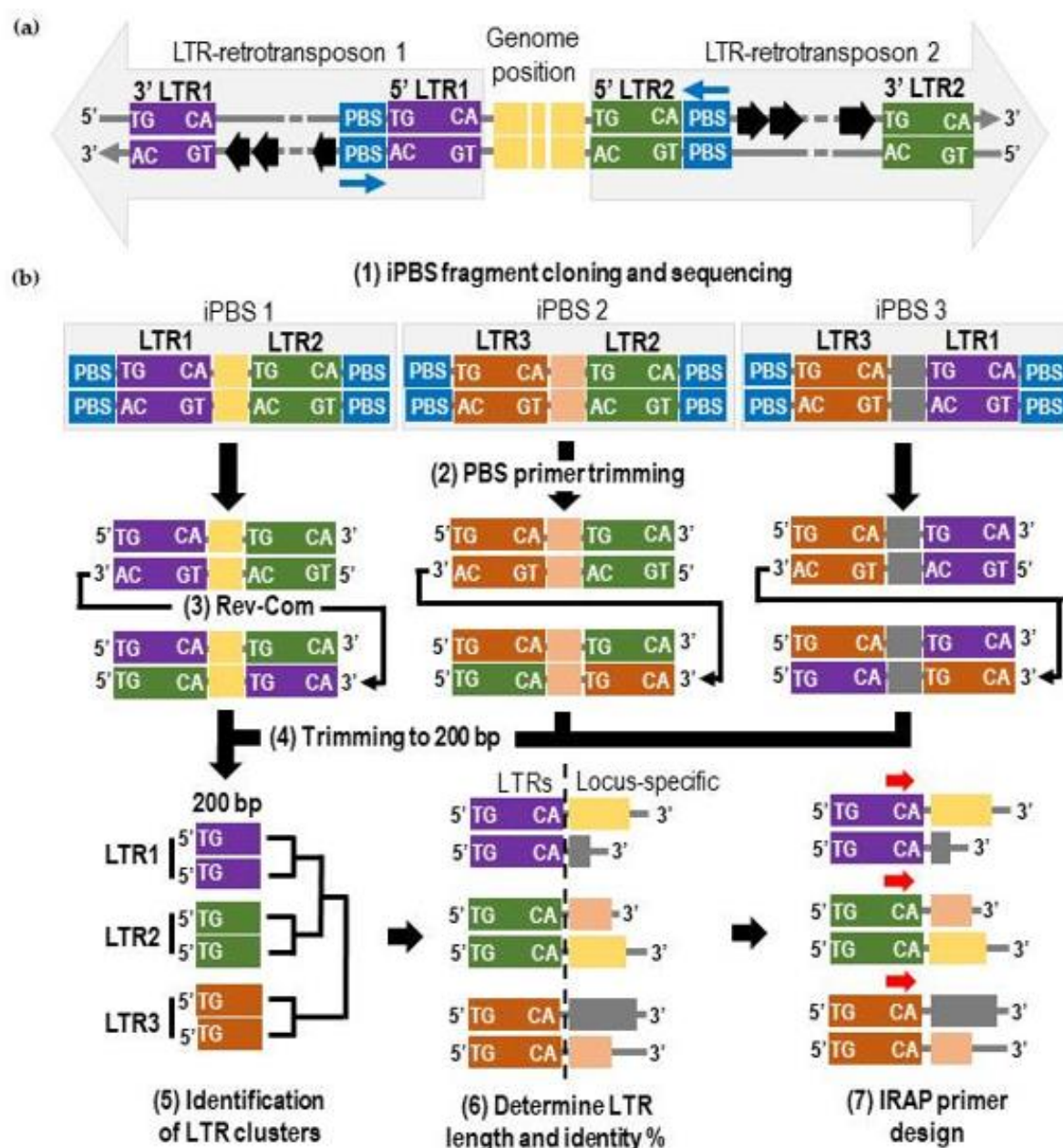


Figure S1 - Description of iPBS technique and pipeline for identification of potential LTRs. (a) Description of the iPBS technique. Two inverted LTR-retrotransposons (grey arrows) are shown, flanking a certain genome position (yellow boxes). Each LTR-retrotransposon contains two direct copies of Long Terminal Repeats (LTRs; purple or green boxes), placed at both 5' and 3' ends. Each LTR typically shows a canonical 5'-TG...CA-3' sequence. Up to 5 residues downstream of each 5' LTR, a region known as Primer Binding Site (PBS; blue boxes) is shown, which is highly conserved

since is recognized by host tRNAs during the LTR-retrotransposon replication. The PBS region is downstream followed by one or several ORFs, which encode transposon elements (black arrows), and the 3'LTR. The iPBS technique makes use of conserved PBS regions for universal primers design (blue arrows), allowing a single-primer PCR amplification of two inverted 5' LTRs flanking a locus-specific sequence. (b) Pipeline for identification of potential LTRs. Several iPBS fragments (three different iPBS fragments, placed at their respective genome locations, are shown) were purified from agarose gels, cloned into *E. coli* plasmids and sequenced. The bioinformatic pipeline for LTR identification includes the following steps: (1) A unique consensus sequence was obtained for each iPBS fragment, which essentially contains two inverted 5'LTRs (that could be identical or different, since they come from two inverted LTR-retrotransposons) flanking a locus-specific sequence. (2) Each iPBS sequence was trimmed from both ends, exactly to remove the PBS primer sequence. (3) Reverse-complementary sequences were generated, in order to address the identification of the two LTRs contained in each iPBS fragments. (4) Direct and reverse-complementary sequences were trimmed to 200 bp from their 5' ends, to exclude locus-specific sequences from next steps, as much as possible. (5) Sequences multiple aligned was used to construct a UPGMA tree, allowing the identification of LTR clusters. (6) Separately for each cluster, respective complete iPBS sequences were aligned, allowing the identification of a region with high identity, which should be flanked by 5'-TG and CA-3' dinucleotides (potential LTR), and followed by a second region of low identity (locus-specific sequence). (8) Potential LTR sequences, or locus-specific sequences, were then used for primer design, in order to develop new IRAP (Inter-Retrotransposon Amplification Polymorphism).

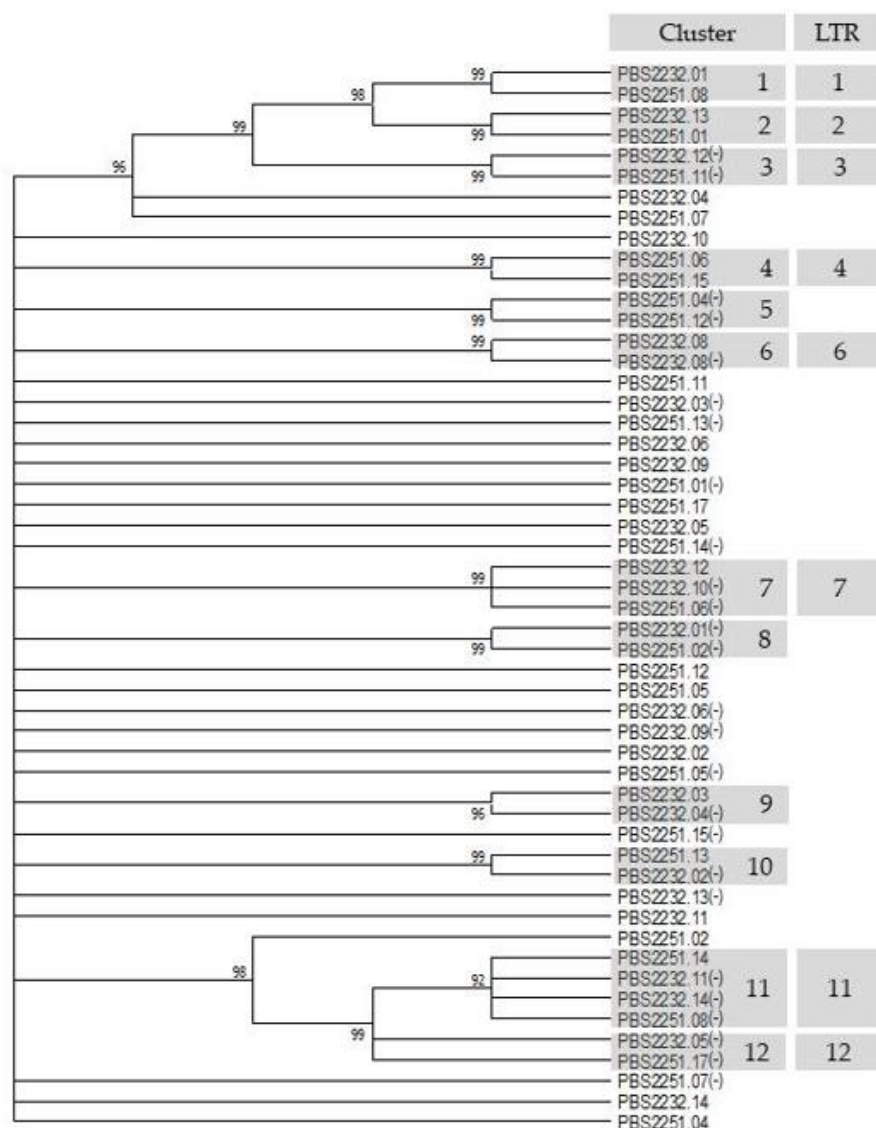
Figure S2. Identification of LTR clusters and potential LTRs

Figure S2- Identification of LTR clusters and potential LTRs. Direct and reverse (-) iPBS sequences were trimmed to 200 bp from their 5' ends, and aligned with CLUSTALW algorithm in MEGAX software. The multiple alignment was then used to construct an UPGMA tree, using the the Maximum Composite Likelihood method to compute distances. The analysis involved 52 nucleotide sequences, and the pairwise deletion option was enabled to eliminate ambiguous positions by sequence pairs (251 positions were included in the final dataset). A consensus tree generated from 10000 bootstrap replicates, with branches reproduced in less than 80% of replicates collapsed, is shown. A total of 12 LTR clusters. For each cluster, the respective complete iPBS sequences were aligned at Clustal EMBOSS Needle server to confirm the presence of a conserved LTR flanked by 5'-TG...CA-3' dinucleotides, and followed by a less-conserved locus-specific sequence. A total of eight potential LTRs which showed these features were found (Dataset S1).

Dataset S1. Sequences of potential LTRs

LTR Cluster 1

>PBS2232.01_LTR (278 bp)

TGTCAAAATCACGATGAGAAAGAGTTTTGAATACTCGTGAGAATGTAATATTTCTGTCTCTGCTATGTAGATTGTAC
ATCATTATATAGGAATTACATGCCTTTAATGTAGGCTCTAAGAAAGGAAATACTATCTAATTCAACAATAATTTA
CAACAATTTCTATTCTAAGAAAGGAAAGCCACAATACTCTCAAGATTGACCAAATCCACGTGAGATATGATAGTA
ATGGGCTTTTGGGCTTTTGGGCTTTTGGCCTGTTTAACACCTGTTTAACA

>PBS2251.08_LTR (280 bp)

TGTCACAATCACGATGAGGAAGAGTTTTGAATACTCGTGAGAATGTAATATTTCTGTATATCTGCTATGTAGATTGT
ACATCATTATATAGGAATTACATGCCTGTAATGTAGGCTCTAAGAAAGGAAATACTATCTAATTCAACAATAATT
TACAACAATTTCTATTCTAAGAAAGGGAAGCCACAATACTGTCTCAAGATTGACCAAATCCACGTGAGATATGATAG
TAATGGGCTTTTGGGCTTTTGGGCTTTTGGCCTGTTTAACACCTGTTTAACA

LTR Cluster 2

>PBS2232.13_LTR (297 bp)

TGACAAACTCACGATTAAGAAAGAGTTTTCCAATGCTTTTTAGAATGTATTATTGTATATCTTTCTATGGAGATTATA
CATCATTATATAAGTATTTACATGCCTTTAATCTAGGCTCTAAGAAAGGAAACCTCTCTAATTACATCAATCAGTGT
TCTAAGTTAAGAAGGAACTAATTACAGCAATTGGTATTCTAAGTAAAGAAGGCTCACTGCTATCTCCCATTCAT
CCCACGTGAGATATAACAATGGGCTTTTGGCCCTCTGATTACTGATGGGCTTTGGGCTATTTAACA

>PBS2251.01_LTR (303 bp)

TGACAAACTCACGATTGAAAAGAGTTTTGGAATACTCTTGAGAATGTATTATTTGTGTATCTTTCTATGGAGATTAA
TACATCAATATATAGGTATTTACATGCCTTTAATCCAAGCTCTAAGAAAGGAAAACCATCTAATTGTATCAATCAC
TAATCTAAGTTAAGAAGCACTAATTACAACAATTGTTATTCTAAGTAAGGAAGGCCATAGCTGCCTCAATATTG
ACCAAATCAATTCCACGTGAGATATATCATTAGTGTGATTACTGATTGGCTTTGGGCTTTGGGCTGTTTAACA

LTR Cluster 3

>PBS2232.12(-)_LTR (313 bp)

TGTCATTCTCACAATTGCAATCATAACAGAGACTGAATGAGACTTTTGAATATTGGATTGTATTCATCGTCCTTTCT
CTTGGAGAGGGATCTCAATATATAGGATTATACATGCCTTAATCAAGGCTCTTACAAAAGGAACTAGCTAATTAC
ATTAATCACTAATCTAAAATAGGAAAACAATCAAAACAGAATATGTACAGCTATCTAATTGTTGACCAAATCAAA
TCCCTCGTGAGATATAGCTAGCTGGACTGGATTGATGGGCTTTGGGTCCCTTCATGGGCCAACTAACACATGAGCC
ATCTTAACA

>PBS2251.11(-)_LTR (307 bp)

TGTCATTCTCGCAACTCTAAACCACACAGAACGCTGAGATTTCTTGATCAATTGATTGTATTGTATTGAATGAATT
CTCTTAGAGAATGAGCACTAATATATAGGAATACACATGCCTTAATCAAGGCTTTTACAGAGGAACTAGCTAAT
CACTAATCTAAAATAGGAAAACAATCAAAACAGAATATGTACAGCTATCTAATTGTTGACCAAATCAAAATCCAC
GTGAGATATGGCTAGCTGGACAGGATTGATGGGCTTTGGGTCCCTTGATGGGCCAACGTATCATATGAGCCAACTT
AACA

LTR Cluster 4

>PBS2251.06_LTR (234 bp)

TGACATGCATATAACGATAAATTCATAAAATCTAAAATTACGAAAAAGCCCTTGGCTCTACCTAACCGAGTCAAGT
CGGGTCCAAATTACCAATCTAAAACCCCAAGGTCTAACCAGTAGCCTAATCCAACTAATAAGGTCAAAACAA
CTTAACCTAACAAACAAAAAAGATTAAATTAATAAAACACACTTAATTAGTGTTAAACGCAATTAATTGGATTG
GGGTGTAACA

>PBS2251.15_LTR (234 bp)

TGACGTGCATATGATGCAAAAATATAAAATCTATAATTACGAAAACGCTCCTCGACTCTACCCAACCTGAGTCAAA
CGGTTCAACTCACCCCTTCGAGACTCCAATAATTCCAACCTACCTATAAGGGTCTAATCAACATTAATTACTCTA
AGTAGCGACTAATCACAATTAATCAACTAAAGCCCGATTAAATTAGTATTAATTAGGCTAAATTTGGGGCTGGGGT
GTAACA

LTR Cluster 6

>PBS2232.08_LTR (225 bp)

TGATGCAATCCTATGGTTAGGATCACAAGTACATGAGCACAAATTAGGGTTCCCATGGGAACAATCACACAATGAA
GACACTCTCCAATGATAAATCCACAAAGGGAAAAATTTTATTAAGTCCAAAGTTGAGATTATTAATCTCTATAGTGC
CATATATATAATAAAAAGAACTTAGGGGCTAAGAAATCAATCCTAATAAGAATCGGAAACCAAATAATAATCCA

>PBS2232.08(-)_LTR (224 bp)

TGATGCAACCTAGAGTTAGGATGCAAGTGCCGGAACACAATTAGGGTTCCGAAGAACAATCTCTCGAAGAAGAA
ACAATCTCCAACGAAAAACCCACACGGGAAATTTCTATTAAGTTCATAATCTCTGAGATTACCAATCTACAATGT
TATATTTATAAGAAAAGAAGCTTAAGGACTAAGAAAGCAATCTTAATCAAAGTGGAAACCAAATAATAATCCA

LTR Cluster 7

>PBS2232.12_LTR (146 bp)

TGTAGGATCTTACACCCCATTAAGCTACCACTTAACCCAAAACCAATTGGTGATAGGAAACGGCACAACAATTGT
TTAATGGTGAGGGCGAGGCTCGACCACAAATATGACCGCTCAAAAACGAGGGTGGACGTTGACCTCAACA

>PBS2232.10(-)_LTR (146 bp)

TGTAGGATCTTACACCACTTGAAGATACCATTTATCCAAAACCAATTGGTGATAGGAAACGGCACAACAATAGT
TTTTTGGTGAGAGCGAGGCTCGACCACGATTATGACCACTCAGAAACGAGAGTGGACGTTGACCTCAACA

>PBS2251.06(-)_LTR (146 bp)

TGTAGGATCTTACACCCCTTTGAGCTACCACTTATCCAAAACCAATTGGTAATTGGAAACGGTACAACAACAGTT
TTTTTGTGAGGGCGAGGACTTGACCACAATTATGACCGCTCAGGAGCAATGGTGGACATTGACTTCAACA

LTR Cluster 11

>PBS2251.14_LTR (181 bp)

TGTCACGCCCCACCTACACCCATGCATGATACTGTCCGCTTTGGCCACGGGCCTCACGGTTTTGTTTCCGATGGG
TCAGAACCCATCAAAAAGGCCTCACGCAAGGTAAGGGTGTAGGCTCTACATATAAACCCATTCACTTTCTCTCTAT
AGTCGATGTGGGACTCCGGGGCGTCACA

>PBS2232.11(-)_LTR (181 bp)

TGTAACGCCCCACCTACACCCATGCATGATACTGTCCGCTTTGGCCACGGGCCTCACGGTTTTGTTTCTGATGGG
CCAGAACCCATCAAAAAGGCCTCATGCAAGGTAAGGGTGTAGGCTCTACATATAAACCCATTCACTTTCTCTCTAT
AGTCGATGTGGGACTCCGGGGCGTCACA

>PBS2232.14(-)_LTR (181 bp)

TGTAACGCCCCACCTACACGCATGCATGATATTGTCCGCTTTGGCCACGGGCCTCACGGTTTTGTTTTGATGGG
CCAGAACCCATCAAAAAGGCCTCATGTAAGGTAAGGGTGTAAAGCTCTACATATAAACCCATTCACTTTCTCTCTAT
AGTCGATGTGGGACTCCGGGGCGTCACA

>PBS2251.08(-)_LTR (181 bp)

TGTAACGCCCCACCTACACCCATGCATGATATTGTCCGCTTTGGTCCACAGGCCTCACGGTTTTGTTTTAATGGG
CCAGAACCCATCAAAAAGGCCTCATGCAAGGTAAGGGTGTAAAGCTCTACATATAAACCCATTCACTTTCTCTCTAT
AGTCGATGTGGGACTCCGGGGCGTCACA

LTR Cluster 12

>PBS2232.05(-)_LTR (179 bp)

TGTCACGCCCTACCATACTCATATGTATGATATTATCCGCTTTGGCCACGGGCCTCATGGTTTTGTTTCTGATGGG
CAGAACCCATCAAAAAGGCCTTATGCAAGGTAAGGGTGTAAAGCTCTATATATAAACACATTCACTTTCTCTCTATA
GTATATGTGGGACTCCGGGGCGTCA

>PBS2251.17(-)_LTR (176 bp)

TGTAACGCCCCACCATACACGCATGCATGATATTGTCCGCTTTGGCCAGGGGCCTCACAGTTTTGTTTTTATGGA
TCTGATCTCAAAAAGGTCTCATGCAATGTAAGGGTGTAAAGCTCTACATATAATCCTATTCACTTTCTCTCTATGGT
GATGTGGGACTCCGGGGCGTCA