

Additional file 3. Sequence verification of non-reference solo LTR or provirus alleles

>S_Yadava-1_18q21.1_W2_env_partial_sequence

GGAAATCTCNCTGCACAACCCCTACTACACCCCAATTTCAGCAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCC
AACAGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCTAGGCCAACTAAGAAT
CCCTAAGCCTAGCTGGGAAGGTGACCGCATCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCGGCCAACCA
GGTAATAAAGAGAGCTCACTAAAATGCTAATCAGGCCAAAAACAGCAGGTAAAAAATAGCCAATCATCTTTTCGCCTG
AGACCACAGTGGGCGGGACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCAATGGCTACCCTCTTTGGGTCC
CCTCCCTTTGTATGGGAGCTCTGTTTTCGCTCTATTAATCTTGCAACTGCGCTCTCTTCTGGTGCATGTTTGCTAT
GGCTCGGCTCTAGCTGAGCTTTTGTTCACTGTCCACCCTGCTGTTTGCTGCCGCCGACAGACCTGCCACTGACTTCC
ATCCCTTCGGATTCCGGCAGGGTGTCTCTGTGCCCCTGATCCAGCAAGGCACCCATTGCTGCTCCTGATCAGGCTAA
AGTCTTGCCATTGTTTCTGCATGGCTAAGTGCCAGGTTTCATCTAATTGAGCTGAACACTAGTCACTGGGTTTACA
GTTCTCTTCCGTGACCCATGGCTTCTAATAGAGCTATAACACTCACTGCATGGCCCAAGATTCCATTCTTGGAAATC
CATGAGGCCAAGAACCCCNCTCAGAGAACACGAGGCTTGCCACCATTTTGAAGTGGCCTGCCGCCATCTTGGGAGC
TCTGGGANCAAGGACTCCCCCGGTAACACTACCATCATCTCTCATCTGATAAGTGAAATAGCCTA

>S_Yadava-2_18q21.1_W2_env_partial_sequence

AATCTCNNTGCACAACCCCTACTACACCCCAATTTCAGCAGGAAGCAGTTAGAGCAGTTGTCAGCCAACCTCCCCAAC
AGCACTTGGGTTTTCTGTTGAGAGGGGGGACTGAGAGACAGGACTAGCTGGATTTCTAGGCCAACTAAGAATCCC
TAAGCCTAGCTGGGAAGGTGACCGCATCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCGGCCAACCCAGGT
AATAAAGAGAGCTCACTAAAATGCTAATCAGGCCAAAAACAGCAGGTAAAAAATAGCCAATCATCTTTTCGCCTGAGA
CCACAGTGGGCGGGACAATGATCAGGATATAAACCCAGGCATTCAAGCCAGCAATGGCTACCCTCTTTGGGTCCCCT
CCCTTTGTATGGGAGCTCTGTTTTCGCTCTATTAATCTTGCAACTGCGCTCTCTTCTGGTGCATGTTTGCTATGGC
TCGGCTCTAGCTGAGCTTTTGTTCACTGTCCACCCTGCTGTTTGCTGCCGCCGACAGACCTGCCACTGACTTCCATC
CCTTCGGATTCCGGCAGGGTGTCTCTGTGCCCCTGATCCAGCAAGGCACCCATTGCTGCTCCTGATCAGGCTAAAGT
CTTGCCATTGTTTCTGCATGGCTAAGTGCCAGGTTTCATCTAATTGAGCTGAACACTAGTCACTGGGTTTACAGTT
CTCTTCCGTGACCCATGGCTTCTAATAGAGCTATAACACTCACTGCATGGCCCAAGATTCCATTCTTGGAAATCCAT
GAGGCCAAGAACCCAGGTGAGAGAACACGAGGCTTGCCACCATTTTGAAGTGGCCTGCCGCCATCTTGGGAGCTC
TGGGAGCAAGGACTCCCCCGGTAACACTACCATCATCTCTCATCTGAATAAGTGAAATAGCCTAC

>S_Mala-3_18q21.1_W2_env_partial_sequence

GGTAGTGTTACCGGGGAGTCTTGTCTCCAGAGCTCCCAAGATGGCGGCAGGCCACTTCCAAAATGGTGGCAAGCC
TCGTGTTCTCTGACCTGGGGTTCTTGGCCTCATGGATTCCAAGGAATGGAATCTTGGGCCATGCAGTGAGTGTTATA
GCTCTATTAGAAGCCATGGGTACGGAAGAGAAGTGTGAACCCAGTGACTAGTGTTTCACTCAATTAGGATGAACCT
GGGCACTTAGCCATGCAGGAACAATGGCAAGACTTTAGCCTGATCAGGAGCAGCAATGGGTGCCTTGTGGATCAGG
GGCACAGAGGACACCCTGCCGAATCCGAAGGGATGGAAGTCAGTGGCAGGTCTGCGGCGGCAGCAAAACAGCAGTGGT
GGACAGTGAACAAAAGCTCAGCTAGAGCCGAGCCATAGCAAACATGCACCAGAAGAGAGCGCAGTTGCAAGATTTAA
TAGAGCGAAAACAGAGCTCCCATACAAAGGGAGGGGACCCAAAGAGGGTAGCCATTGCTGGCTTGAATGCCTGGGTT
TATATCCTGATCATTGTCCC GCCACTGTGGTCTCAGGCGAAAGATGATTGGCTATTTTTTTTACCTGCTGTTTTTGC
CTGATTAGCATTTTTAGTGAGCTCTCTTTATTACCTGGTTGGCCGGTGTGAGCTAAGTTGCAAGCCCCATGTTTAAAG
GTGGATGCGGTACCTTCCCAGCTAGGCTTANGGATTCTTAGTTGGCCTANGAAATCCAGCTAGTCTCTCTCTCAG
TCCCCCTCTCAACAGGAAAACCCAAGTGTGTTGGGGAGGTTGGCTGACAACCTGCTCTAACTGCTTCTGCTGAAT
TGGGGTGTAGTAGGGGTTGTGCAGTTGAGATTTCTTGGGAGGGTGCNTTGTATGTCAT

>S_Kapu-1_18q21.1_W2_gag_partial_sequence

GANAGCAAGCGNAAGAGGTCCAATATTACTCACTGCTTTGGAGATCTCTTCGTGGTTACCAAAATGTCACCGGGGGT
TCCTTGCTCCCAGAGCTCCCAAGATGGCAGCAGGCCACTTCCAAGATGGTGGCAAGCCTTGGCCTCACGGATTCCAA
AGAATGGAATCTTGCACCATGCGATGAGTGTTATAGCTCTATTAGAAGCCGTGGGTACCGGAAGAGAACCATTGGAAC
CCAGTGACTAGTGTTTCACTCAATTAGGACAAATCTGGGCATTAGCCATGCAGGAACAATGGCAAGACTTTAGCCC
GATCAGGAGCAGCAATGCGTGCCTTGTAGTACAGGACACAGAGCACCCCTGCCGGATCCGAAGGGATGGAAGTTC
AGCAGCGGGTCTGCGGTGGTGGCAAACAGCAGTGGTGGACAGTGAGCGAAAGCTCAGCTTGAGCCAAGCCATAACAA
ACATGCACCAGAAGAGAGCGCAGTTGCAAGATTTAATAGAGCGAAAACAGAGCTCCCATACAAAGGGAGGGGACCCA
AAGAGGGTAGCCATTGCCAGCTTGAAGGCCTGGGTTTATATCCCAGTATTGTCCTTACTGTGCTCTCAGGCAA
TAGGTGATTGGCTATTTCTTTTACCTCCTGTTTTTGCCTAATTAGCATTTTTAGTGAGCTCTCTTTATTACCTGGTTGG
TCGGTGTGAGCTAAGTTGCAAGCCCCATGTTTAAAGGTGGATGTGGTACCTTCCCAGCTAGGCTTAGGGATTCTTA
GTCAGCCTAGGAAATCCAGCTTGTCTGTCTCTCAGTAGCATGAGCAATATTGGTGACAGAGAGGAAGAGCTGAATT
CAGTAGACCTGATGATGGTTGGATAG

>S_Yadava-1_18q21.1_W2_gag_partial_sequence

GCTCTTCCTCTCTGTACCAATATTGCTCATGCTACTGAGAGACAGGACAAGCTGGATTTCTAGGCTGACTAAGAA
TCCCTAAGCCTAGCTGGGAAGGTGACCACATCCACCTTTAAACATGGGGCTTGCAACTTAGCTCACACCGACCAACC
AGGTAATAAAGAGAGCTCACTAAAATGCTAATTAGGCAAAAACAGGAGGTAAAGAAATAGCCAATCACCTATTGCCT
GAGAGCACAGTGAGGGGGACAATGATCGGGATATAAACCCAGGCCTTCAAGCTGGCAATGGCTACCCTCTTTGGGTC
CCCTCCCTTTGTATGGGAGCTCTGTTTTCGCTCTATTAAATCTTGCAACTGCGCTCTCTTCTGGTGCATGTTTGTTA
TGGCTTGGCTCAAGCTGAGCTTTGCTCACTGTCCACCACTGCTGTTTGCCACCACCGCAGACCCGCTGCTGACTTC
CATCCCTTCGGATCCGGCAGGGTGTCTCTGTGCTCCTGATCTAGCAAGGCACGCATTGCTGCTCCTGATCGGGCTA
AAGTCTTGCCATTGTTCTGCATGGCTAAGTGCCAGATTTGCTCCTAATTGAGCTGAACACTAGTCACTGGGTTCCA
TGGTTCTCTTCCGTGACCCAGGCTTCTAATAGAGCTATAACACTCATCGCATGGTGCAAGATTCCATTCTTTGGAA
TCCGTGAGGCCAAGGCTTGCCACCATCTTGAAGTGGCCTGCTGCCATCTTGGGAGCTCTGGGAGCAAGGAACCCCC
GGTGACATTTTGGTAACCACGAAGAGATCTCCAAAGCAGTGAGTAATATTGGACCTCTTTCGCTTGTATTCTGTCC
TATCCTTCCTTAGAATTG

>S_Mala-1_4q22.1_H8_env_partial_sequence

TGAGAAACATCGCCCCATTATCTCTCCATACCACCCCAACACTTCAACACTATTTTATTTTTCTTATTAATATAAG
AAGACAGGAATGTCAGGCCTCTGAGCCCAAGCCTGCATATATACATCCAGATGGCCTGAAGCAAGTGAAGAATCACA
AAAKAAGTGAAAATGGCCGGTTCCTGCCTTAACTGATGACATTACCTTGTGAAATTCCTTCTCCTGGCTCAGAAGCT
CCCCATTGAGCACCTTGTGACCCCTGCCCTGCCGGCCAGAGAACCCCTTTGACTGTAATTTTCCATTATCTACC
CAAATCCTGTAAAACAGCCCCACCCCTATCTCCCTTYGCTGACTCTCTTTTCTGACTCAGCCRCCTGCACCCAGGT
GATTA AAAAGCTTTATTGCTCACACAAAGCCTGTTTGGTGGTCTCTTACACGGATGCATGTGAAAATTATACCCCA
AGATTTTGCATTATTTTTATCTTGTCTAACTATATAATGAAAACAAATTGAAAAGACCCAATATATTTAAAAGCTACT
TTATAGGAGTTTTAAAAGAATTRTATCTATCAAGAATCTCATTATAATTGTAGATACTATGGAGGATTATGGAATA
GAGATACACCAAATATTTCCATAATAATAATGTTTTATCTGTTTTTA

>S_Madiga-2_4q22.1_H8_env_partial_sequence

CCACTCTAGGTTCCACACCACCCCTAATCCCGCTGGAACAGCCCTGAGAAACATCGCCCCATTATCTCTCCATA
CCACCCCAACACTTCAACACTATTTTATTTTTCTTATTAATATAAGAAGACAGGAATGTCAGGCCTCTGAGCCCAA
GCCTGCATATATACATCCAGATGGCCTGAAGCAAGTGAAGAATCACAAAAGAAGTGA AAAATGGCCGGTTCCTGCCTT
AACTGATGACATTACCTTGTGAAATTCCTTCTCCTGGCTCAGAAGCTCCCCATTGAGCACCTTGTGACCCCTGCC
CTGCCCGCCAGAGAACCCCTTTGACTGTAATTTTCCATTATCTACCCAAATCCTGTAAAACAGCCCCACCCCTATC
TCCCTTCGCTGACTCTCTTTTCTGACTCAGCCCGCCTGCACCCAGGTGATTA AAAAGCTTTATTGCTCACACAAAGC
CTGTTTGGTGGTCTCTTACACGGATGCATGTGAAAATTATACCCCAAGATTTTGCATTATTTTTATCTTGTCTAAC
TATATAATGAAAACAAATTGAAAAGACCCAATATATTTAAAAGCTACTTTATAGGAGTTTTAAAAGAATTGTATCTA
TCAAGAATCTCATTATAATTGTAGATACTATGGAGGATTATGGAATAGAGATACACCAAATATTTCCATAATAATA
TGTTTTATCTGTTTTTATTTTTAT

>S_Brahmin-1_5p15.31_H2_env_partial_sequence

TCTAGGTTCCCATGCCACCCCTAATCCCGCTCAAAGCAGCCTTGAGAAACATTGCTCATTATCTCTCCATACCACCC
CCAAAAATTTTACCATCCCAACACTTCAGCACTATTTTCATTTTTCTTATTAATATAAGAAGACAGGAATATCAGG
CCTCTGAGCCCAAGCTAAGCCATCATATCCCCTGTGACCTGCACGTACACATCCAGATGGCCGGTTCCTGCCTTAA
TGATGACATTCCACCACAAAAGAAGTGA AAAATGGCCTGTTCTGCCTTAACTGATGACATTGTCTTGTGAAATTCCT
TCTCCTGGCTCATCCTGGCTCAAAGCTCCCCTACTGAGCACCTTGTGACCCCACTCTGCCTGCCAGAGAACAACC
CCCTTTGACTATAATTTTCTTTTATCTACTCACATCCTATAAAAATGGCCCCACCCCTATCTCCCTTCGCTGACTCTT
TTCGGACTCAGCCTGCCTGCACCCAGGTGATTA AAAAGCTTTATTGCTCACACAAAACCTGTTTGTGGTCTCTTAC
ACGGACGCGCATGAAAAGTTACAACCTTGA AAAATATGATACACACATTAGAAAAAAGCATGCCAATTAATGCAATTT
ATAGAAACACAACAAAAGAACTCAGAAAAATGATGAAAAAAGATACGGTTACATACAGCAAGATGGTGTAGCTCC
ACTCGTGAAACACAATGGACTTTAAAACAACAAGCC

>S_Madiga-2_5p15.31_H2_env_partial_sequence

CCCCTCTAGGTTCCCATGCCACCCCTAATCCCGCTCAAAGCAGCCTTGAGAAACATTGCTCATTATCTCTCCATAC
CACCCCAAAAAATTTTACCATCCCAACACTTCAGCACTATTTTCATTTTTCTTATTAATATAAGAAGACAGGAATA
TCAGGCCTCTGAGCCCAAGCTAAGCCATCATATCCCCTGTGACCTGCACGTACACATCCAGATGGCCGGTTCCTGCC
TTAACTGATGACATTCCACCACAAAAGAAGTGA AAAATGGCCTGTTCTGCCTTAACTGATGACATTGTCTTGTGAAA
TTCCTTCTCCTGGCTCATCCTGGCTCAAAGCTCCCCTACTGAGCACCTTGTGACCCCACTCTGCCTGCCAGAGAA
CAACCCCTTTGACTATAATTTTCTTTTATCTACTCACATCCTATAAAAATGGCCCCACCCCTATCTCCCTTCGCTGA

CTCTTTTCGGACTCAGCCTGCCTGCACCCAGGTGATTAAAAGCTTTATTGCTCACACAAAACCTGTTTGTATGGTCTC
TTCACAYGGACGCGCATGAAAGTTACAACCTGAAAATATGATACACACATTAGAAAAAAGCATGCCAATTAATGC
AATTTATAGAAACACAACAAAAGAACTCAGAAAAATGATGAAAAAGATACGGTTCACATACAGCAAGATGGTGT
GCTCCACTCGTGAAACACAATGGACTTTAAAACAACAAGC

>S_Madiga-2_2q34_H4_soloLTR

TAGGTGAAAGAGCTCACATAGCAGAGTATTGTGATATTAGCCATGATGTACTAACCATCTACATATGTTAGTTGATT
GTAATAATAATAAAAATAATTAATTTTTAAAGATATAAACTATACTTATTGTGTCAGGTCTCTGAGCCCAAGCCAAGCCAT
CGCATCCCCTCTGACTTGCAGGTATATGCCAGATGGCCTGAAGTAACTGAAGAATCACAAAAGAAGTAAAAATGCC
CTGCCCGCCTTAACTGATGACATTCACCAAAAAAGAAGTGAATATGGCCGGTCCCTTGCCTTAAAGTATGACATTA
CCTTGTGAAAGTCTTTTTCTGCTCATCCTGGCTCAAAAACCTCCCCACTGAGCACCTTGCACCCCCACTCCTGC
CCACCAGAGAACAACCCCTTTGACTGTAATTTTTCTTTACCTGACCAAATCTTATAAAACGGCCCCACCCCTATC
TCCCTTCTCTGACTCTTTTTTCGGACTCAGCCCGCTTGCACCCAGGTGAAATAAACAGCCATGTTGCTCACACAAAG
CCTGTTTGGTCTCTTTCACACGGACGTGCATGAAACTTATTAATTAATAAGATAAAATAAGTGAAGTAACTAGATT
CATTTTTAACCTGAATTAATACAAATAAGGTGAAATGGTTGGCCTACTGATTGTCTATAAGGAGCAAACCTATAAAAT
ATCAAGTATTGTCACTGCTTGAATATGCAAATGAAAGCTGTGAATAATAATCACGTGGTCAGGGAGAGAAATAAGTT
TAAAGATATTAATAAATCTAGAGCACTTTCACATTATGTCAATGTAGTTAAAAGGATGAAAAGAAAATGTTTATCTT
CCAGGA

>S_Luhya-2_3p14.3_H1_soloLTR

TTCTCTTTGGGAATGTGAGGCCTCTGAGCCCAAGCCAAGCCATCGCATCCCCTATGACATGCACGTACACGCCCAGA
TGGCCTGAAGTAACTGAAGAATCACAAAAGAAGTGAATATGCCCTGCCCCACCTTAACTGATGACATTCCACCACAA
AAGAAGTGAATATGGCCAGTCCCTTGCCTTAACTGATGACATTACCTTGTGAAAGTCCTTTTCTGGCTCATCCTGGC
TCAAAAAGCACCCCCACTGAGCACCTTGCAGCCCCCGCTCCTACCCGCCAGAGAACAACCCCTTTGACTGTAAT
TTTTCTTTACCTACCCAAATCCTATAAAACGGCCCCACCCCTTATCTCCCTTCGCTGACTCTCTTTTTCCGACTCAGCC
CGCCTGCACCCAGGTGAAATAAACAGCCTTGTGCTCACACAAAGCCTGTTTGGTGGTCTCTTTCACACAGACGCGCA
TGAAAGGGAAGACATACAAAACAAGGTAATAAGTAACTACGTTATATGTTTGATAATGGTGTATGTTAAGGGTGG
GGAAAGAAGAAAGCAAAGAAGGATAAGAAATGGGAGGGGGCAATTCTAGAAAC

>S_Luhya-2_5p13.3_K2_soloLTR

TCCATCATATAAACAGAACTGTGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAAGAAGTAGAC
ATAGGAGACTCCATTTTTGTTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTGACCTTACCCCCAACCCCGT
GCTCTCTGAAACATGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCRGTGCAAGATGTGCTTTGTTAAACAGA
TGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCCTCCCTCATCTCAAGTACCCAGGGACACAAAAACTGCGGA
AGGCYGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGTTTCATAGTCTGAAATATGGCCTCGTGGGAAG
GGAAAGACCTGACTGTCCCCAGCCCGACACCCGTAAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGCAT
GCCTCTTGAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATRGAATGTCTCGKTATAAAAC
CCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACYGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACT
GCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGAT
GCAAAGACCTTTGTTACGTTGTTGTCTGCTGACCCTCTCCACATTGTCTTGTGACCCTGACACATCCCCCTCTTT
GAGAAACACCCACGAATGATCAATAAATACTAAKGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGG
TTCCCCGGATCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCTTAAAGTCTCTCATTCCACCTT
ACGAGAAACACCCACAGGTGTGGAGGGGCAACCCACCCCTACACAGAACCAATGACAAAAACCACATGATTATCTCA
ATAGATGCAGAAAAGGCCTTCGACAAAATTCAATAGCCATTAATGCTAAAAACTCTCAATAAACTAGGTATCGATGG
AACATATCTCAAAAATAATAAGCGTTATTTATGACAAAACYCACAGCTAATATCATACTGAATGTGCAAAAACCTGGAAG
CATTCCCTTTGAAAACCTGGCAAAAGACAAGTATGCCCTCTCTACCCTATTCAACATAATGTTGTTAGTTCTGGCAA
GGCAATGAGGCCAGAGAAAAGAAAATAAAGGGTATTTAATTAGGAAAAGAGGAAGTCAAATTTGCCCTGTTTGCAGAT
GACATGATTGTATATTTAGAAAACCCCATTTGTCTCAGCCCAAAAATCTCCTTAAAGCTGATAAGCAACTTCAGCAAAAT
CTCAGGATACAAAATCAATGTGCAAAAATCACAACTTCTTATAACATCAATAACAGACAAAACAGAGAGCCAAATCA
TGAGTGAACCTCCATTACAATTGCTACAAAAGTGAATAAAGTACCTAGGAATCCAACCTTCCAAGGGATGTGAAGGAC
CTCTTCAAGGAGAACTACAAACCACTGCTCAACAAAATAAAAAGAGGACACAAACAAATGGAAGAATATTCCATGCTC
ATGGATAGGAAGAATCAATATTGTGAAAATGGCTATACAGGTCAAGGTAATTTATAGATTCAATGCCATCCCCATCA
AGCTACCAATGACTTTCTTTCACAGAATTGGGAAAACCTACTTTAAAGTTCATATGGAACCAAAAAA