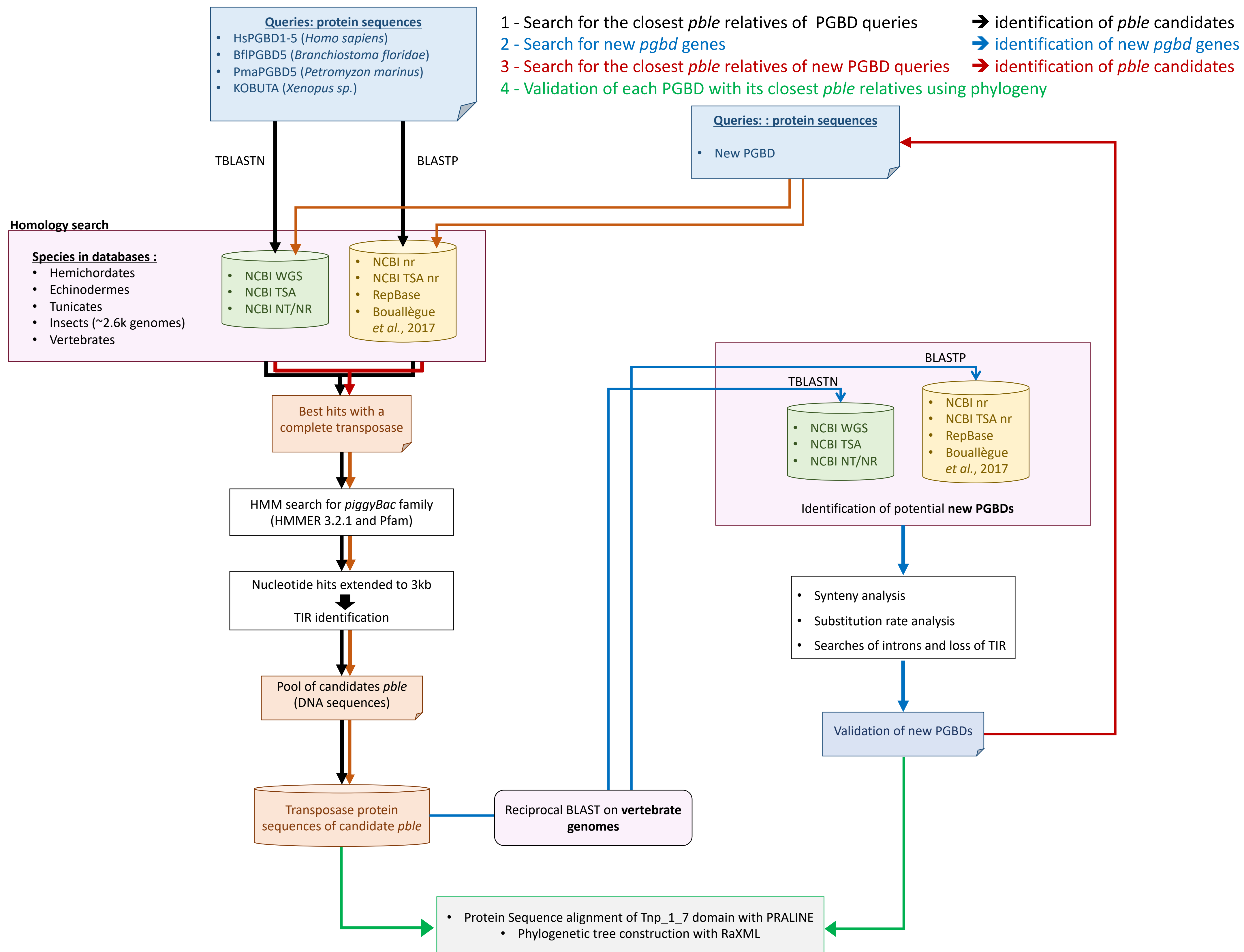


Supplementary Data 1: Graphic representation of the data mining process to inventory *pgbd* genes and their closest *pble* relatives.



Supplementary data 2. : Localization of PGBD3 and PGBD4 ORFs and respectively MER85 and MER75.

Fasta sequences of localization of PGBD3 and PGBD4 genes. PGBD3 sequence is reversed (complementary strand). ORF of genes is underlined in green and TIR of the element are in bold. For PGBD3, the first bold sequence corresponds to position 1 to 43 of MER85 and the second to the positions 39 to 140. For PGBD4, the first sequence is for positions 68 to 341 of MER75 and the second one is the positions 379 to 514. TSD were not identified for MER75.

>NC_000010.11:49515105-49524281 Homo sapiens chromosome 10, GRCh38.p13 Primary Assembly (PGBD3)

ACATCTTTAA**CCCATTTATGTC**TGAGGTT**CAATTTTTTTGAATTTTTGCATGA**AATTCATATATGTTACC
TAGAAAAATTCAGGAACAAAAATTTGAAGAGAAAAAATTTATTATGTTCCATTGTTATGTGACGTT
CCAAAATTTGGAACACTAGGCAAAACCCTGCTACACTGTACATAATGTATAAAACTATGTTTCTTATCTC
AGGAGGTGGTGACACCTG**CTATTCAGTGTGATATTCACCGAACACTTCACATGTAAGGCAACATCACAT**
TTTTACATCGAAAAGTTGTGTTCTTATGACATTCAGCGCATCGCGTTTGCTTTCCCTGTTTGACTATCA
CATGATTTATGCCATCATAACGTGAGTCAATGTTACGCTTCTGAGGTCTTCTTTTTGGCCAGGTTCTGG
AGGATGACCATGGGTCTCCAGATAATGGCATAACCACACGTCGACGAACTCCAGAAAATCCACTGGTTTC
TCATCATATGTTTTATGCAATTGCCAAGCATTGTAAGACCAGTTCGAAACAGAACAAAAGAGGGCTTG
AATACCATTTCTTCCACGGATTGATGCCGATACTTATCAATGTTTTCATCAGCTCTGTCTACGCCTCC
CATGAACTGGTTATACACTTTGATCATGTTGGCTGCTGAACTTGTATCTTCTTTTTCAGTTTCTGGGAG
TAACGACTGACAAGACACAGGGGATGGATAACCAGCACCAGATGAGGCAACAGTGACAACACTGTTATCAT
TCCATCTGCAGACAATATTGCCCTTTGCCATCAATTCGATAATCAAATGTGCCTCTTCTTTTTTCTTTAA
AGCTACATCTGATTCAGTGGAACTCTGTCAATGTGATCCTTCTCACTGTACCTGTTGCCTGATGTCCC
ATTGAACTGAGCTTATCAAGAAGTGCAATACTGGTGAAGAAAGTTATTGAATACAAAATGGTATTGTCCAG
GGTGTGCCTCTGTAAGTGCCTCACTAAACTGAAGGACAAGTGACGCACCGACACCATATTCCTCATGTTT
AGTATTTGGGTTTTTACCCTGATACGGCTGAAACCAGCAAATGTAGCCAGACAGGTGGCACCACACCAA
AACTTATAGCCAAACCAGATGGCTTTCCCCGAATAAATGTTTGCACCCGTGACGACCAAAAATAAGGAA
CCATGAATTCATCAAAGCTGAAATATGTTTCATTTGGAACAAATTCATGCATCTCTCATTAAGTTTGCT
TATGAGAGGTCGCAATTTGAAAAATTTGTCCACTGGATCCAAATTTGCATTGTGCAACATGCAAAATTA
GAAAAATAGTTTCAAACCGTCACTCTCATGGCAGCACTAACCAGTACATTATGCACATCTGTTCTTT
GTTCCCAAAACATACGCCTTCTAGGAACTGAGACATAACCCTCAGAAAAATAATCCAGAAAAACATTT
GAATTCAGAGCTAGTCAAGCCAAGATGTACACCTTACTGCAAGCATATAAGTTGGAGTACTTGACAATG
AGTTCAATGACCTCGTCATCAAGAAAAGTTCAAGAATTTCTGTGGGAGTTCTCATTACGGTGAAGAAAT
CGTTTGGTGGTGTGTAACCTTACCTGCTACGGGTTGTACAGTTAGGTCCGCTTTTTTCCATTTGCAAAG
AATTTTTGTCATTTTCTCCTCCTTGATGGTGGAGGTTGCTGCACAGTAAACGTAGATGGAACCTTCATCA
GGAGAGTCATCTTAGGTGCGTATGAGGGATCATCTGAGTCAGACTCAGCATCAGAGCCATCTTGAATAA
GATACGCAGCTGTGTGCAACAAAGAACCTGGCAGATTATTTATTGTTCCACCTTCTTCATCTCCTGATTC
CTCATCAGAAACAGGTGCTGTAGCATTTTCAAGGTGGTTGTATCACTATAGCACTTGCTTCTATGCTGTCA
TCTGTCTCTAAAAGGTCAGTTATTTTCATGTAACCTTAGTGTTCGAGGCCATCTTGGGACTAAAACGGAAAA
AAGGTATTTATAGTCCTAGTGGTAGTGGTTTTGCCTAGTGTCCAAAAATGGAACATGAAAAATGTCAAA
CCCATAACTAATGCAATAATAATACTTTCTGTATTTTCTTTTTACCGCACAATAGCATAACTAGTAGAGA
AGTTATTTCTTAAAAATTACTAGTATAGGATGTAGTGCAGTAAAGTTATATGTACCTGTCTTTGGATACT
CAAAAATGTTTCTGAAAACCTCACATCACTTTTTGACAAGCTCCAGCATGCCAGTATCTGTAGCAGTAAC
TAACCTCTGTGACCAAGTGATAACATGTACCCTCAGGAGGCAGAC**CATGTTGTAAAAATCAGACTTTGGCA**
ATGACCTTTGACCAAGTAGGCATGAATAAATCCACATGCTTAGCGTTCCAATAATGGAACACTAGGCATA
AATAGGTTTTTAAAAGTGTAACTTTTTCTTTCAACTTATAGATTCCATAAAGAAAAATGAACACAGTCTTCT
TTCTTTTTTTTTTTTTCTCTCAGGCTCCAGTGATCCACCCACCTCAGCCTCCTGGGTAGCTGGGACTACAG
GCGCACATCACCACACCAGCTAACTTTTTCTATTTTTGGGGTTTACCATGTTGCCAGGCTGGTCTCG
AACTCTTGAGCTCAAGCAATCCTCCTGCCTTGGCCTCCTAAAGTGTGGGATTACAGGTATGAGCCATTG
TGCCCGCCAGGCTTATTTCTTGAGAAAAATGAACTCTCAGAGATGTTAAGAACATTCTTAAAAATAATTG
TATTTAGGTTTCTGGATAAAATCTCAAACCTTTACTAAATTCAGATTTTGGGAGAGTTTTGTTTTTCTC
TTTAATTTGGGCTTTAAAAAAAAGGCACAATCAAATCAACCTCCACATTGTGAGTTGGCCTTTGCCGAAG
CACAGCTAAAATGCTAAATCCTCATCGGGTGCAAACCAGGATCCAACCTGAAGGACAGACAGGACAATC
CCAGTGCAAACAGGTGCTAGACAGAAGGGAGAAGTTCCAGAAGACAATAAAATGGAGACAGGCTGTGAG
GTGGTGGTTCTCAGGGCACTGCCTTTAGAAGCATTTCAAGGATAGGAATGGGACATTGGGTTGGAGTTGG
CCCCATCAAGTGGAGGGAGAAACCAGGCACCTGAAACAATGCTTGTCCGAGTCTGCTGCAGCAAACCAAG
CTGCGAGATCACTCTCATGTGTCCACACCACCTTCAGACCCCTTCTCCTTCTCCAAGCCCTTCCACTGA
CTGCTGCCCTGAGCAAATGACTTCATCTCTCTGGGCTCAAAGTCTTATCTCCTGGGTGGCTGTAAGGATT
AAACGAATTAACCTATATAACAAGAAGAACCACAAGAACCCTAAGTCGTACGCATTCAAACATGTGAGC
CATCATTTGTTCTTGTGCACCCTCCCGCCCCGCCCCACCTCCACTTGATGCCACCTTTGCTCAGGCC
TAAAGGCTGCAACGTTCCAAATCCACAAGACTCTTTCATAACAGTCACTTACTTTGGCTTGACCCCTAG
CATCTGAAATGCTGACAGTCCCTCCTCAGAACAGTGGATGTGCACAACCCCAACCTCCCAAGCCAGGGT
TCAATGGGGCCCTGGCCACAGAGGCCAGTGTTCAGGGGAAGGACAATGAAGAGTATTATCTTCGCAGGA
GATTATGTCTCACAGACTATTAGAAGACATTCTGTAAGACTTACACAATGAATGGAATCTTTATCACCCC
TTTACATCTCAGAGTTACTCATGTAATAAAGGTAAAAAATAGGGCACAAGAAAAACCTTTAGGATCAAAA
ATTTATTAATAAACAATATATCAACAGGCATCAAGTCTACAGATTACAGGTTACACCAGACCATGAAGTA

AATTCTGTCCCCATCCACACCATACTTGCCAGGTCTTCTAGACTCCTGAGCCATCTCCCTATATCCTCAT
CCAAATTCCCAGATTACAGGCTTAGGTTTTTGGTTTTGGTTTTTCAGTCCAATGGAGGTGGGGGCAGCTATG
TGTTGATTTTTTGGCACCACCCCTGTGGTCATACCTAAATATTCACCTTCTACTCAATCCCACAAAGGGAA
GGAAACATGTATGTAGAGGGCCCCAGCCGACTCAAATCTCAGCTCTACTTCACAGTAAAACCTGTGGGCAA
GCCAATTAACCTGGAGGGCGTGATTATAACCTAGTAAAGATGTGTATTCACAGAACCAAAAGACTGGAAT
GACATGTAGAAAAATAAAAAACATAAAAAGAGGTCAATTTTTTTCATCTTAATATTGTTGTACATACTTTTACAAT
AAAAATAGGGACTCATCAGAAGTGCTGAAAGGTGGCGAGGGAGCCTTCAAACCATGCAGAGGAACCAACT
CAAATCCACTTCAGGCTATGCATTTAGCAGCTTAGCCCAATGCACCAGGTCATCTGAGGAGACTCAGGCA
CCGCTACTGAAGAGTCCGTTTTCTGAACTCCCCACTATGTTTTCCCTGATCGGGGTGGGCATGGGTCTGGGG
ATTTATCCCAGAGGCAATTATTTGCTGAGTGTGGGCCATAGTAAAGAAAAAAGCATTACATTCTCAAGCC
TCAACATCCTTACCAACAATGCTTCCCAACTGCCTCAAAGCTCTCCTAAATGAGAACATAGTTCTTTCTG
AGCAAGTCCCTGTGGACCATGAAGAATGTCACCAAGCTCCCTCAGAGTCAGCGGGAGCTCAGCCAAAGC
ACAAGTGCAGTGGCCAGCTCCTCCACTCTGCACCTGCTGCCTCAGACTCCCCACGCTGAGCCCAGGCCC
CTACCCTCTGAAGGTGTTTTCCCATGTGATTTCTGACACACACACCCCAAGAACCAGATGATCTATGACA
TACAGCATTTAGCTAAGGCTCCCTGCCAGGCCCTCCCTGTGTACATACTGTGCAGGGGCACATAAAA
GAGGATTTTTGAGAAACTGATATACAAACTGCATCAGCCAAAAGCTGCAGCCCCAAAAAAGAGAAATACA
TAAGACTGACTTATTTCCATCCCTAGCTCCCAAACCTGACAGTGATCAGAGCCATCTGGGAACATTTTTT
AAAAGTCCCTGCCCCGGTGACCCCACGCTACACTTCCTAAGTCAGCGTCTCCAGTGTAGGGGCTCCCAGTG
ACACTGCATAATCAGTCTGTGGGTGGGGAAGCTTTGCTCTCGTCTGTGCTCGTTGCCCTCACTGTAGTCT
GCGAGGGCAGCAACTTAAATAGTAACTGCCGATACATGAAAATAGAGTAAGAAGTGGGGGAATGAGTTT
AAGTCAAATGATAAACACTAACCCTAGTCTGCCAGCAAGAAGAGACAAAAACATATCTGGAAATGCATT
TTCTCACTCAGGATGCAACCCATCTATGCCCTGCAGCCTGACCAAGCAGTGTGACCCTGTGGACAAA
GAGGCAGGATCAGTACATCTGTTTTATTATGTCAAGGCTGGCAAATACCTAGCATGAATGTGTGCTAGT
TCCCCACTGTGCCATGATAGACATCACTAAGGAACGCACACTCTTCCCAATCCACCCAAACAGGGCCT
CGGAGCCCCCTGAATAGACTACCCTCCCAAGAAATCAGAGCTGGCTGGAAAGTGAAGCCACCTGCCCT
CCCTGTCTCAGCCTCTCTGCACATTTGTGGCATCACACCAACCTCCAAGGCAATCCAAATGAGAACCCTG
AGGAAATAATGTGCCCTCAAAGCAAAGCAAATTCCTTCCCTTCTCAATGATGCACCTGGAGTTTAGGGGC
AAATGCGTAACATTTGTTAATTTGGATAAGGGTCATATTAATATATCTGAGCCACCCAACAAGACAAGGCC
CTGCACTGGATGAACTTCCCCAACAGTCTCTCCCAAATGCCTCAGTCTGAATTCATTATTTCCCTTCACAC
TTGGTTAGGTCCTATTTGTGCCAAGTCTATTAACAGACTGCAAATTTTTTGGCAGAAACCCTTGAAGCC
AGTCCAAGGATGCAGACATGCATTGTTGACATGAAAGCAAAGATAAAGCTATGGAGGCTTACCTCGAATA
TGGCAACTGGGAAGTGAATCAATAACATGGAGGTAACAAGGACTGGGGACAAGCGACCTGATCTTTATCA
CCCTGAAGGGGAAAGGTTACATGAAATGTTTTGGTCTTCTGGGCCACACAGGGCTGTCTTGGCTGCACAG
CCTGTACTAGGTAAGCAGCATGTAGGTGCTCACCGGGAACCCACATCACTCAAACACTGAAGGTGTGTTA
GTCCATAATCTTTATTTGGCACCCGGAGCTCCCAGGCCACAATGTGCAAAGCAGTAGTATATTTAAGAA
AGCACTAGACCTCACAACCAACAGCCATAAGCAAATGGAAACCGTGGTAGCACAAAGTAAAATGAAAAA
TAATGTATGTTTTACAAGACCTGTATACCCTGACCATAGAACCAGGACAGAACATTTACTTCTAAGCAGTG
GCTCATCCATCCTGTATGACATGAAAGGGTATGAAAACCAGGGAACCCACTGAGGAGTCAGCCTCCAGAT
ACAGTTCAGTGCAGAGGAGGAGGAGGAGGAGAAACACATGAGAAGGGGGAATACACGGTTTTGAAAAATGGTT
AAAACCTCAAAGACTGAGCTGAGGCTTGAGCCAAATGAAGGAAAGCAGGATCCAAAGGGGCACTGCCCA
TTCCCATTAAGAAAGCTATGAAAGGGCTTGACGGAAAAATTCCTAGTAAGTATCAGCTAAAACAAAGGCA
CCTAGTTAGGGTGAACCTAGTGCCAGTGAAGGCGGATGGCTCTCAAGCTGAGAGAATAAAGAGGATGAG
ATATCCAGAAGTCACCCACTCTCACTTCCCTAGAGTTGAAAAGAGCATTATAATTAACAGGCTATAACCCTC
AGTAAAAGGTGTTAAGTTCAGCAAAGGGTGCCACCCTGGGAACTACTCCTAAATTTCTAACCCTAAAAA
AAAACTAATTTTTCTATGTTGTAACCTCTATGACTTTATTTCTAATCATCCATCGTCAAACCCTGCTAAGA
CAAGCAACCCATAAGGTCTTTCTGTTTTCTCAAGGGAATAGCATTACCATTTCAATTAGTCTACAAAGATC
AGCAATGAGTAACAAAATATAGAGAGACTGGCAGTTCATAATATGCTTCTCAATCCTGAGCATTATTTTT
CTTTCATCCCCCAAACCAAACAGCACCAAAGGAGATGTAATAACAAATATTTCAATGTGGATATCTGCAG
TTTAAACAGCCTTAATAACTGTAGCTATAAAATGTTAAGAGCTATCGTATTTTTGGTGGCAATTAAGAGTTA
TTGATATTTCTGAAATTCATTTGAGATTTCACTGTAGCTTTAGCAAATATTGGCTCTGGAAATAATAAA
TTGTTTTACAATCTTAACATTTAAGAAAACTTATTGCCATCAAATTAACCTAAACCCACATTTAAAAT
TCAATCAAAATAAATGCTCTTGATCAAAATACTGATATATTTTAAACATATTTAATATTGACCCAATTT
AATGGAAAAATATATATTGCTCCATTTATAATCAATATGTTGAGATTAACACTTTTATTAATTCCTTGC
AGTAGTTAAGAAAAAGTCTGACACCGCTCACTGTTCAAGAGTACTGGGCAACATTCCTCAAACAACCTTCA
CAGGTAATAAACGTGCACAGAAACGAAAGCTGTGCTGAGGGATGACAAATATAAGCCAGCCCTCATCTGT
GCATTACTACATAAGGTGCTACTAGAATGTGAGTGCCGCAACTCTAAGAACAGGATTTAAGTCACAGGT
CCTCAAATTACTGTTTTAAAAGAAATGACACACTATTTTCGGAAAGGTTATGGAATGTCTGTAGTGTCTGG
ACGGGGGTTAATGAGCAGTCTTAAGTAACAAGAAAAACACCAGGTCAACCTTCACTCACAAGGCCAAT
AAATACAACAAACTAGTCTCTTCCACTTGAGAAATAAGAAAAAAGTGTGAGGATTGGGAAAAAGCCAT
CCCCTATCAAGTATTTTTAAAAGTAATAAGGTTCCCTTTTTCCCTTTGACTTCCCTTAGAGCTAAGTGAGCT
CTAAGTCAAGGAGACTTGAAATGCGCGGCAGTAGCACAGTGTTCAGTGGGCCACCATGGGGAGGCCTGCA
ATACAGATTTATTTTTCTGGCTCTCAGAAAACGTGTGGCAATATGCAATTCAGACTTCATGTTTCACATG
TTCACCTTCCCTTTTATGGTGTCTGCTTTATACAATGACAATGTCATGGAGCTAAAAGCCAAAGTCAGTC
TCTAACATAAAAAACAACAGTGAAGAAAGGTATGCTGTGCATACAAAGTACACTTTTTTGCATGATGGT

GGCTTATAATTATGCAGATGACTATTTTCTCAAGACAAAGGAAGGATTTACAAAAGCAAGCTTATTCTTT
CTATGCCAAAGCTTAAGTTTCAGCTCTTTTTTGAAGAAAAGTTCTCTAAAATTACTTTTCTGCTTCAAAA
GGGAAAACAGAAACCCCTATCCAGTATGCATTAATTCACCCTTTTCTTTTCTGATACTCATTCTAGGG
TGACTGTCGGTATAAAGGCGACACAGCTCACACTCAGCCCTCTGTATGCCTCTTGCCCCAGCCTTCCCTT
AAGTTTTTCATCTTAGACTTGCCTCCAAATCCAACCTAAAGACCTAGAAGTACATGCATTTGGGGCCAGCA
TCATTAAGGACTCAGGGGCCCACTGCTTCTAGCAGGTTAAGGCTGCAGAAATCCAACCTCTGTATAATC
GGGGGGTCTAATATATTAAGCTGGATCTAGAATGCTTACATATTAATAAACCTGTCACCTAAGATAAAG
CAAACAGTACAATGTATTTATAATCCCCACAGACCGACCTTAACCGCTGCTTATAATAATCTTCATCTCC
ATCATCTCGGTATCTTCCCACTTTCCGACCTCCTCCTCCTTCTCCTACAGAAGCAGCTTCAGCTTCT
TCCCCAGAACTTGGGAAAAAGTCATCATCAATCTCCTGCACTGGCACTTTCTTCTGCCGTTTCCCGCCCT
TGGGCAGAGGCTTCAGCTCATAGTCAGTACCATCTCCAGACAGGTCCGCCTCTGCCCCCTCCACCTCGTC
ATCTTCC

>NC_000015.10:34102083-34108686 Homo sapiens chromosome 15, GRCh38.p13 Primary Assembly (PGBD4)

GCCTATGAAGGGGACTGCCCATGAAGTGAAAGTCAAGTGTGTGTTGCTGCGGCAGCCACGGAGGCCAAGG
ACCTCACGGGAGTAAAAGATGACGAGACTGGCTTCGGGAGAAACACCATCCAGAAGAGACCTTTCAAAAA
ACTTCTAGAGACTCCCCAAGACGTATGAGATGAAAGGCTTCTTCTGTCTGTGTGCGTTTAACTCATTCT
CCTTAGCCCCGAGATTACGCGCTGTGTGCCGTGCAGCTGCAGCGTTTACGCCGAGATAACTCGTGGATTA
CAGTGCCAACCTTACTCCCA**AAGTTTGCCACGAAATATCTCGCTTCTGTTATTTTTCGCATGTTTCTGGTA**
TATTGACTTTTAAAACAAAAGACATCATTCTGTTTATAGCATTCTGTTTTTATAGTAGTGGATTTCATCT
ACAAAATATAGTAATTCTCGATCGCTGAAATGTCAAATCCTAGAAAACGTAGCATTCTATGCGTGATAG
TAATACCGTCTCGAACAGTTGTTGGCTGAAGATTCATTTGATGAATCTGATTTTTTCGAAAATAGATGAT
TCTGATAAATTTTCGGATAGTGCTTTAGAAGCCGATAAGATCAGGCCTCTGTCCATTTAGAATCTGATG
GAAAGAGCTCTACATCAAGTGACTCAGGGCGCTCCATGAAATGGTCAGCTCGTGCTATGATTCCACGTCA
AAGGTATGACTTTACCGGCACACCTGGCAGAAAAGTCGATGTCAGTGATATCACTGACCCATTGCAGTAT
TTTGAAGTGTCTTTACTGAGGAATTAGTTTCAAAAATTACTAGAGAAAACAAATGCCAAGCTGCCTTGT
TGGCTTCAAAGCCACGGGTCCGAAAGGATTTTCGCGAATGGATAAATGGAAAGACACTGACAATGACGA
GCTCAAAGTCTTTTTTGACAGTAATGTTACTGCAAGGATTGTGCAGAAACCTGAGCTGGAGATGTTTTGG
TCAACAAGGCCCTTTTTGGATACACCTTATCTCAGGCCAAATTATGACTGGTGAAAGATTTTTACTTTTTGT
TTCCGGTGCCCTGATTTTTGTCAACAATTTCTATATCTGCTGGTCAATCAAAGGCCAGATTTTATTGCA
GAAGATCAAACCTGTGTTCCGACTTTCTTGTAATAAAATTTCCACTGTATATACTCCAAACAGAAACATT
GCAGTTGATGAATCACTGATGCTGTTCAAGGGGCCATTAGCTATGAAGCAGTACCTCCCGACAAAACGAG
TACGATTTGGTCTGAAGCTATATGTACTTTGTGAAAGTCAGTCTGGTTATGTGTGGAATGCGCTTGTTCA
CACAGGGCTGGCATAAATTGAAAGATTCAGCGGATGGCCTGAAATCATCACGCATTGTCTTACCTTG
GTCATGACCTTCTTGCCCAAGGGTATTGTCTCTTCCCTCGATAACTTTAATATATCTCCCAGCTTTTCA
GAGAATTACATCAAAAATAGGACTGATGCAGTTGGGACAGCTCGTTTGAACAGAAAACAGATTCCAATGA
TCTGAAAAAAGGATTGCAAAGGGACGACTGTAGCCAGATTCTGTGGTGAACCTTATGGCACTGAAATGG
TGTGACGGCAAGGAGGTGACAATGTTGTCAACATTCACAATGATACTGTGATTGAAGTAAACAATAGAA
ATGGAAAAGAAAACAAAAGGCCACGTGTCATTTGTGGATTATAACGAGAATATGGGAGCAGTGGACTCGGC
TGATCAAATGCTTACTTCTTATCCATCTGAGCGCAAAGACACAAGTTTGGTATAAGAAATTTCTTTCAC
CATCTTCTACACATTACAGTGCTGAACCTCTACATCCTGTTCAAGAAGGATAATCTGAGCACACGATGA
GCCATATAAACTTACAGACTGGCATTGATTGAAAGAATGCTGGAAAAGCATCACAAGCCAGGGCAGCAACA
TCTTCGAGGTGCTCCTTGCTCCGATGATGTCACACCTCTTCTGCTGTCTGGAAGACATTTCCCCAAGAGC
ATACCAGCAACGTCCGGGAAACAGAATCCAACCTGGTCGCTGCAAAATTTGCTGCTCCCAATACGACAAGG
ATGGCAAGAAGATCCGGAAAGAAACGCGCTATTTTTGTGCCGAATGTGATGTTCCGCTTTGTGTTGTTCC
GTGCTTTGAAATTTACCACACGAAAAAAATTTATTAATACCCGATCATCATACACATCTGTTCCATTAGG
ATTAGAGACAAGTTCTGTTTAGAAAATAACTCCAAGATTTGGGAGGCCGACCTGGGTGGATCACCTGAGATC
AGGAATTC AAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAACAATTAGCTGGA
TGTGGTGGCACATGCCGTGAATCCAAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCATGAG
GCGGAGTTGCAGTGAGCTGAGATCACACCCTGCCTCCAGCCTGGGTGACATAGCGAGACTCCATCTC
AAAAAAGAAAATAACTCCAAGAACAGTTTGTGTTTGTGTTTTGTTTTGTTTTGTTTTGTGTTTTGAGACGGAGT
TTCACTCTTGTGCCCCAGCTGGAGTGGCAATGGCATGATATTGGCTCACTGTAACCTCCACCTCCCAGGT
TCAAGCGATTCTGTGCCCTAGCCTCCCTAGTAGTGGGATTACAGGCATGTGCCACCACCCCACTCAAT
TTTGTCAATTTTGTAGAGACTGGGTTTTATCCATGTTGGTTCAGGCTAGTCTTGACCTCCCAGCTCAGGT
GATCCGCCTATCTGGGCCCTACAAAAGTCCCTGGGATTATAGGCATGAGCCACTGCGCCTAGCACCAAGAAC****
AGTTTTTATATTTTATTTTTCACATTGAAAATCAGTCAGATTGCTTTCAGCCTCAAAGAGGGTGTGTTATGT
AAAACTAAATGAGTGCAGGCAGCGAGCTACACTTTTTTTTTTTCTAAATGGAAAATGGGAAGGAGATACT
GTCTCTGAACAAAAATAGCTGGAAAAACACATAACTGTTAACTAAGGGTTGCCTCTAGTGAAAAGGATTG
GAAGACAATTTATCTGATGCTACACTATTTCTTCTGTGTTGTTGATTGTTTTACTGTGAACATAACAAGTTG
TATTGGGGAGGAGGGCAGGGTGTGATTGCTGCCAATGTTGATTGCTGAGGCCTTAGCCAAGATTTACGGC
ATCATTGTCTCCAGTCTTTTCAATTTCTTTATGGTGGAAAGTGAAGTGACTTGCCCAAGATCTTTGTTGG
CCCTAAACATCGTGTCTCTGCTACAGATTCCTATTGAGAGCAATGCATGTTTTAGCCTTGTGCACTCTT
AGAGTCAGCATCATGGGAAAGAGTAACAGCAATTGAAAAAAGAGATTCAAGAGCCAAGTACTCAGCCCTG
GCAGGGGATGGGGACATGATGCCAGAGCCACCCCTATACTCCAGTCTCCCTGGCATCAGCTACTGCAGT
ACTGGGGTACTTGCCCTCATCTTTTTTTCTGAAGATAGTTACTTCTTCATATTTGATCATCGGAAGGAGCA
GCATTGCTCAGAAACACACACACACACACACTTGACTTTCTTGTCTGCCAAATAAAATTTCTTATTT
GCTAGGGAACCTGGTAATCTAGGCTCAAGAAGTGTGTGGAGAAGGTTATTTTCACTTTTACTTTAAA
ATATGCCTTAGTTTTTGAATAGATAAGACATTC AAGTGTGCAAAAAGTAAAAGTCAGTTCAGGACTGAGAA
AGGAAAGATCAGTGCTGATTTTTAATCAGACTGTAATGTACCAACTAGATACAGACAGGCGCCAGTGGCAAG
GCCATCAGAAAACCTTCTCCCAAGCCACACCACCTGCAAAAATGCCAATAACAACCTTCTTTTTCAGT
GATTACCACCTTTCTACCAATGACTTCGCTTTGTTTCCCTCGTGCCTCTGAACAGACTGCTGAGATACCCAA
TCAAGAGCCAATCCTTGCTTCCCTAGACTCTCCTCAATCATTCAACAGAAGCCCAAATCTTGTAAGGTT
CCTTCTGGGCCGGGCACAGTGGCTCAGCCTGTAATCCCAGCACTTTGGAAGGCCGAGATGGGCGAATC
ACAAGGTCAGGAGTTTGAGACTAGCTGGCCAACATGGTGAACCCCTGTCTCTACTAAAAATACAAAAA
AAATTAGCTGGGCATAGTGGTGGGTGCCGTGAATCCCAGCTACTTGGGAGGCTGAGACAGGAGAATCACT
TGAACCCGGGAGGCAGAGTTGCAGTGAGCCAAGATCGCACCACTGCCTCCAGCCCTGGCGACAGAGTG

AGACTCCATCTCAAAAAATAAAAAAGTTTCCTTCTGTTCCCTCTTACTGTTATATCCTTGGTTTTCTCTGG
TGTGGGGTCTTCCCTTACAGCAGCAAGCTCATAAACTTAACTTGACCGCAGGTGTGTTCCCTTGGTGGTCTT
TATCTGATAGACTTTTGTACATGGTACAATATTTCAAAGGCACATGGTACAATATTTCAAAGGCACATGG
TACAATATTTCAAAGGTATGAGTATACAAAAAAATCTCCACCTACTTCTCTCCCTAGTCACCCAATTC
TCAAGAGGCAATCACTGTTGCCATTTTCCCTGAGTATTTCCACAGATAGTCAACTATCTACAGAAAAAT
ATATACATTAATAAAGGCCAGGCACAGTGGCATGCCTGTAGTCCCAGCTACTCAGAAGGTTGAAGCAG
GAGGATCACTTGAGGCCAGGAGGGCCAGGAATTTGAAGCTGTAGTGTGCTATGATAGCAGGCCTGTGAAT
AGCCATTGCACCTCCATCCTGGGCAACATAGCCAGACCCTGTCTCTAAAAAAAAAAAAAAAAAAAAAAAAAAG
ACAAAAAACAAATCCATATATACATATATATGTATATATATAGTATGCTACCAGTTGTGTGAAAAAAGG
GGCAAAATATATATTTTATTTAACTTATATACAAATAACATTTCTAGAAATAAATTGTTTAAATATAAAATAC
ACTAATATATAATAATGTATTACCTAACATATGATTATATATAACTATAATGTGTACTGTTTTACATACA
TATTTCCAAAGTATACTATAAATGCACCTCCGCACCTTGTCTTTTTACTAAATATATCTTGAAATCAT
CCTTTATTCGTACATAAAAAGCTTCATAGTTCCTTTTTATGGCTGCAAAATGTTCCAGCTTATGGATGGA
CTGATTTCTCTATCGAGCAACATTAAGATTGTGTCCATTTTTACTATTCCCTAATTTTGTGGAAGTGAATTT
CTTTTGCCATGTGATTTCCACAGGTGTATATATGTAGCGTAATTAGTACTAGTAGAAAGTAGAATTGCTA
GATCAAAGAGTATGTGCCTTTGTAATTTTGATGATATTGTGAAATCTTCTTCCACAGAAGTTGTTGCAAC
TTTCACTCTCACCAGCAATGTGTAAGATAACTTGCTTCTTCATACCCTCACCAACAATGAGCTATCAAAT
GTTATTTTTTGCCAAGCAAATAAATTAAGTGTATCTCAGTTTTTGTGTTGTTGTTGTTTTTGGAGAC
AGAGTTTCACTCTTTGTTGCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCTCACTGTAACCTCCGCCTC
CCAGGTTCAAACGATTTCTCTGTCTCACCCTCCCGAGTAGCTGGGATTACAGGCACCTGCCACCACACCC
AGTTAATTTTTGTATTTTTAGTAGAGACAGGGTTTACCATGTTGGCTAGGCTGACCTCGAACTCCTAAC
CTCAGGTGATCCACCTGCCTCGGCCCTCCCAAAGTGTGGAATTACAGGCGTGAGCCACCATGCCTGGCTC
TTCTTGCTTGTTTTGTTTTGAGACAGAGTCTTGCTCTGTGAGCCAGGCTGGAGTGCAGTGGCACGATATTG
GCTCACTGCAACCTCTGTCTCCAGGCTCAAGCAATTTCTCCTGCCTCAGCCTCCTGAGTAGCTAGGATTA
CAGATGTGTGCCACCACGCTGGCTAATTTTTGTACTTTTTAGTAGAGACGGAGTTTACCATGTTGGCCA
GGCTGGTCTCAAACCTCCTGACCTCAGGTAATCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATGACAGGC
GTGAGCCACCACGCTGGGCCCTTCTTGGTTTTTTAATCCATCTTCTGCCATATGGTGAATTAATCTAA
GTTGTATTGTAGGATTAATGGAGAGCAGCTAAAGAGGAATAAAGCCATGTGTTTCCAGCCAAGGATAGTTAT
CATGAAGGATAAAAAGCCTCAAATCCTGATTTTGAATACAAAGTGCTGATTTTAAATCAGTACATGTT
GAGCTGAGCCCTGAGGCCTTTGTTTTCAGTCCCAACCAGAGCAGAAGTAAGCATTGTTGAAACCGTAGTAA
AAAGATTATTTGAGAAATAATGGAATGGGACAAAGGATTATAATGTGCTACAGCCCTCTGACCATTACAC
TTTCTAAACTATTAATAAAATTTTACTTTTTCAAAAATGATTTCAAAAATTGCAAAATTTAGAAATATGCT
GAAATAAAAACAGCTTCATCCCAA

Supplementary data 3. Sequence alignment of *pble* transposases and PGBD proteins used to calculate the phylogenetic tree shown in Figure 4. The black shade is for residues which are present in more than 60% at a position. Yellow is for the position of the catalytic triad. The blue line is for β -sheet and red line for α -helix.

Tni-pble 1 NIYDPLLCFKLFFFTDEIISEIWKVNTNAEISLKRRESM--TGA-----TFRDTNEDIYAFFGILVMTAVRK---DNHMSTDDLFDRS--LSMVY
Hmel2-pble 1 QDKRPIEWFENFLDDEDVISLIVSESNKYAVKKNLPG-----DITTEDMKCFIGILLVSGYSW---LP---RRRMWYENSPTKNELI
Csu-pble 1 ---RPIEWFENFLDDEDVISLIVSESNKYAVKKNLPG-----DITTEDMKCFIGILLVSGYSW---LP---RRRMWYENSPTKNELI
Smi7-pble 1 KVRTPTTEYFELFLNDDVMELIVTYSNLYASTKGVNLS-----LTNSELKCFLGIIIFLSGYVS---IP---RRRMFWEQRTDSQNALV
Lfu-pble 1 EVGTPTELLERFLNDEVIELIVYNSLYAASKGVNLS-----LSNSELKCFLGIIIFLSGYVS---VP---RRRMFWEQRADAHNGLV
Hvu-pble 1 TAKTPTTEIFELFLDDKVEIIVTYTNLYASENGLTLG-----LSNSEFKFLLGIIILSGYVP---LL---RRRMFWEQRPDTHNKMV
Ppa-PGBD3 1 VMRTPTEILELFLDDEVIELIVKYSNLYACSKGVHLG-----LTSSEFKCFLGIIIFLSGYVS---VP---RRRMFWEQRTDVHNVLV
Hsa-PGBD3 1 VMRTPTEILELFLDDEVIELIVKYSNLYACSKGVHLG-----LTSSEFKCFLGIIIFLSGYVS---VP---RRRMFWEQRTDVHNVLV
Ota-pble 1 EASTAIDFFELFFDAELVDKIVLETNRYASLQKKRAL-----NVCTSEELKCFIGILLVSGYTO---VP---RRRMFWEDNSDAHHEIV
Hsa-PGBD2 1 QELSPVGLFELFFDEGTINFIIVNETNRYAWQKNVNLG-----LTAQELKCVLGIILISGYIS---YP---RRRMFWETSPDSHHHLV
Ppa-PGBD2 1 QELSPVGLFELFFDEGTINFIIVNETNRYAWQKNVNLG-----LTAQELKCVLGIILISGYIS---YP---RRRMFWETSPDSHHHLV
Tma-PGBD2 1 QELSPVGLFELFFDEGTINFIIVNETNRYARQKNVNLG-----LTAQELKCVLGIILISGYIS---YP---RRRMFWETSPDSHHHLV
Tma-PGBD1 1 EKLNPVELFELFFDDDTFNLIIVNETNRYASQKNVNLG-----VTVQEMKCVFVGLLLSGFVR---RP---RRGMWYEIS-DADKNLV
Hsa-PGBD1 1 EKLNPVELFELFFDDDTFNLIIVNETNRYASQKNVSLG-----VTVQEMRCVFGVLLSGFMR---HP---RREMWEVS-DTDQNLV
Oaf-PGBD1 1 EKLNPVELFELFFDDDTFNLIIVNETNRYASQKNVNLG-----VTIQEMKCVFVGLLLSGFVR---RP---RRGMWYEIS-DADQNLV
Nvi-pble 1 KKNVESFLFSQIFSTSLDSDVVTETNKYAFQKKSKN-----WHDNTNEEVSALIGMWILMGIHK---LP---RVRMWESSDPFLKVDV
Smi2-pble 1 STDKEKDYFGTIFTEEILQNIIVDETNRYANQPKKNKRLGSKNTKDIILNWKKEITSEELKAWIGMHILMGIHQ---LP---ELKNYWSTDPILNVQSV
Oaf-PGBD7 1 HTDTEIDYFELSVFTEEMLETIKDQTNLYATQERITRSRGQEIRRSTENWQPTTVEEIKAFIAVHILMGIHI---LP---ELRHYWSSDPLLGVASV
Cas-PGBD7 1 HTDTEVDYFELSVFTEEMLETIKNQTNLRYATQERTIRSRGKEIKRSAENWQPTSVEEIKAFIAVHILMGIHV---LP---ELRHYWSSDPLLGLASV
Laf-PGBD7 1 HTDTEVDYFELSVFTEEMLETIKDQTNLYATQERTTRSRGQEIRRPTQNWQPTTVEEIKAYIAVHIVMGIHV---LP---ELRYWSSDPSLGVASV
Cgi-pble 1 ADKNEIDFLNLLFPEELYSKISETNRYAQLCIRAKAN-----PGWYDTTPEEIRAFGLCLVVMGIIS---AP---AQDLWYSSDKLFLWSCI
Bgl-pble 1 EDCSPLDYFKLFIITDEMIGLIVTETNRYAEQQQVAKG-----QRDSQWKPVTAQDIYTFLYINIMFGIHW---IP---STKLWYSSDKSRVPAI
Pfl-pble 1 VQAQPLDYFLLISPSFLQLVIVDETNRYATQSQRING-----ADPSWYETKVDEMRAFIAVNILMGIHT---LP---DVNDYWSTDDRLOVPGI
Ola-PGBD9 1 DDADTRDYFRLLFPDSLFEHMVEQTNNYALYRQRSG-----KSDPHWHPTDVREMKAYVGLNILMGINQ---LP---DTGMWASDIFIGNAGF
Hbu-PGBD9 1 DDADTRDYFRILFPDSLFEHMVEQTNNYALYRQRSG-----KSDPHWHPTDVREMKAYVGLNILMGINQ---LP---DTGMWASDIFIGNAGF
Pny-PGBD9 1 DDADTRDYFRILFPDSLFEHMVEQTNNYALYRQRSG-----KSDPHWHPTDVREMKAYVGLNILMGINQ---LP---DTGMWASDIFIGNAGF
Tcr-pble 1 STESPVDTFLCLFPDSFIEHLVFOQTNLYATQLGKPF-----VPVTEEEMKIFIAINIMMGIKR---LP---SYRDHWSSYPQLRDSFI
Bbe-PGBD5 1 PNTQPIDFFHMLVPESLLSAIVQTNRYAHKYQLKNG-----RDEEWRNVDLAELKAFMAIFMYTSIVHGIFSE---DITTVWEDNSITYNPVI
Pma-PGBD5 1 ASATPADFFQLFVDPDGLLKIMVSQTNAYARKFQERFG-----QDPAWHEVDVAEMRAFLGVVASTATHH---CE---SVLSIWGGG-FYGNRSV
Hsa-PGBD5 1 PSASAVDFFQLFVDPDNLKNNMVQTNMYAKKFQERFG-----SDGAWVEVTLTEMKAFGLGYMISTSISH---CE---SVLSIWGGG-FYSNRSL
Oaf-PGBD5 1 PDASAMDFFQLFVDPDNLKNNMVQTNMYAKKYQERFG-----SDGAWVEVSLAEMKAFGLGYMISTSIIYH---CE---SVLSIWGGG-FYSNRSI
Api1-pble 1 NNGTPIIDFYFLFIDDEVISLIVTETNKYANKKLASPN---IKRHSRLQKWIDTNESEIKNFLGIIIMYMGITVN---MP---TIASWYWTSS-HLLTSHL
Api3-pble 1 ANATVWDCFNLFFTDDIIEELIVTETNRYAEQFLSSNR---ISKSSRFTKWWPTDSKEIKLFLSGLLIWMGLVE---MP---NLGSWYSTKHRYKNYVA
Atr1-pble 1 ANATPIDFYSLLVTDQLFENIAHQTNIFAQQSIDKGG---IKDKSRLRLWPTDKEEMKRFGLGVIIWMGLIVK---LP---KIVDYWSSNDMLGQSF
Psi-PGBD8 1 HEKTPYDFFKFFFITDDVIDYMKETNIYASQTLQKGNV---KPHSRLKKWYDVTSQEMERFLGVLFWMGLSK---KP---RLPDYWSKN-ILYDNSV
Mmu-PGBD8 1 HEKTPYDFFKFFFITDDVIDYMKETNIYASQTLQKGNV---KPHSRLKKWYDVTSQEMERFLGVLFWMGLSK---KP---RLPDYWSKN-ILYDNSV
Dma-PGBD8 1 HEKTPYDFFKFFFITDDVIDYMKETNIYASQTLQKGNV---KPHSRLKKWYDVTSQEMERFLGVLFWMGLSK---KP---RLPDYWSKN-ILYDNSV
Pva-pble 1 DTTDPLQYFELFFTEELVSKITSETNAQAALLASKPPGPRGFPR-MDKWKDNDDELKVFVFAVMLLQRIVQ---KP---ELEMFWSTRPLLDTPYL
Oab1-pble 1 FRSKCSSYVEKYLDHLSIIVKETNLRYADQFLQSHPNLKPRSR-MRKWYSTNNEVRCFIAAILLQGIIVK---KP---ALDMWFLKREIICSPPF
Hsa-PGBD4 1 DITDPLQYFELFFTEELVSKITRETNAQAALLASKPPGPKGFSR-MDKWKDNDDELKVFVFAVMLLQGIIVQ---KP---ELEMFWSTRPLLDTPYL
Uribo1 1 TNFQIIDEFNQFITEASLQDMVHFTNLRYAEQYLASHPLPGYSRA---HAWYPTNVGEIKRFLALTLAMGLIVE---RN---TLASWYDTTTTVLSIPLF
Uribo2 1 SNFEPINFFQLFMTEAILQDMVLYTNVYAEQYLTQNPLPRYARA---HAWHPTDIAEMKRFVGLTLAMGLIK---AN---SLESYWDTTTTVLSIPVF
Pxu1-pble 1 PNSTPLDAFQLFFSENLSNIVRETNNYKYVIANSSNKIKSRI---NSWKDTDISEMYAFFALTMLMPRVK---KL---AINEYWSTDELLQNTTF
Xtr-PGBD6 1 NNFELVDFFHVVFVTEIDILQNMVIFYTNLYAEQFLARCPLPEYSQG---HAWHPTNTHEIKRFLALTIAMDLEH---GH---TLASWYDTTTTIISIPLF

Tni-pble 83 VSVMSRDREFDFLIRCLRMDDKSIR--PTLRENDVFTPVVKIWDLFIHQCIQNYTPGAHLTIDDEQLL-----GFRGRCPFRMYIPNKPSKYG
Hmel2-pble 77 SSAMTRDRDFDFIFRHLHVNNDLQ-----DKYTKVVRPLVTLNKKFLEFSPLEEHYSVDEAMI-----PYYGRHGCKQHIKPKPIRYG
Csu-pble 74 SSAMTRDRDFDFIFRHLHVNNDLQ-----DKYTKVVRPLVTLNKKFLEFSPLEEHYSVDEAMI-----PYYDRHGCKQHIKPKPIRYG
Smi7-pble 77 SSAMRRDRFETIFSNLHVADNANLDP-----DKFAKLRPLINKLNELCMKFVPNETHFSIDESMV-----PYFGRHGCKQFIRGKPIRF
Lfu-pble 76 SGAMRRDRFETIFSHLHVADNANLDP-----DKFAKLRPLIINKLNERCMLFVPNETCFSDDESMI-----PYFGRHGCKQFIRGKPIRF
Hvu-pble 77 SSAMRRDRFESSIFTNLHVSNNANLDHE-----DKFAKLRPLISELNNRCLKFTPMETNFSFDESMV-----PYFSRHGCKQFIRGKPIRF
Ppa-PGBD3 77 SAAMRRDRFETIFSNLHVADNANLDPV-----DKFSKLRPLISKLNRCMKFVPNETYFSFDESMV-----PYFGRHGCKQFIRGKPIRF
Hsa-PGBD3 77 SAAMRRDRFETIFSNLHVADNANLDPV-----DKFSKLRPLISKLNRCMKFVPNETYFSFDESMV-----PYFGRHGCKQFIRGKPIRF
Ota-pble 77 ANGMKRDRFEAILSHFYLANNYELDAE-----DKYAKIRPLVMTINEKFLCYAPIEEYYSFDESMC-----KYFGRHGCKQFIRGKPIRF
Hsa-PGBD2 77 ADAIRDRFELIFSYLHFADNNELDAS-----DRFAKVRPLIIRMNCNFQKHAPLEEFYSFGEPMC-----EYFGRHGCKQLHRGKPVRLG
Ppa-PGBD2 77 ADAIRDRFELIFSYLHFADNNELDAS-----DRFAKVRPLIIRMNCNFQKHAPLEEFYSFGEPMC-----EYFGRHGCKQLHRGKPVRLG
Tma-PGBD2 77 ADAIRDRFELIFSYLHFSDNNELDAS-----DRFAKVRPLIIRMNCNFQKHAPLEEFYSFGEPMC-----EYFGRHGCKQLHTGKPIRLG
Tma-PGBD1 76 RDAIRDRFESIFSYLHFADNSHLDQK-----DKFTKLRPLIKQMNKNFLLYAPLEEYCFDKSMC-----ECFD---SDHCLNGKPTRIG
Hsa-PGBD1 76 RDAIRDRFELIFSNLHFADNGHLDQK-----DKFTKLRPLIKQMNKNFLLYAPLEEYCFDKSMC-----ECFD---SDQFLNGKPIRIG
Oaf-PGBD1 76 RDAIRDRFELIFSHLHFADNSCLDQK-----DKFTKLRPLIKQMNKNFLLYAPLEEYCFDKSMC-----ECFD---SDQFLNGKPIRIG
Nvi-pble 79 ANIMPYTRFSKLTETHFNNDNSMLPRTDPNYDKLHKIRPLIDHINQTSNKSYNTSKTVSVDDESMI-----PFSGRSSFIQYMPNPKIKRG
Smi2-pble 91 ARVMTAKRFKKITETLHVNDNSTNPPRGQPGHDKLHKVVRPLSOLNDSICKAYEPSSTLSIDECMI-----AFKGRSLSLKYMPMKPIKRG
Oaf-PGBD7 91 PDLMTKARFKKLTENIHCNDNSKAVPKGEPGYDWLHKLRPINALNSRLKEVYMPSSVMAVDETMV-----PFKDRSSLKQYMPIKPVKRG
Cas-PGBD7 91 PDLMTKARFKKLTENIHCNDNSKAVPKGEPGYDRLHKLRPINALNSRLKEVYVPSVMAIDESMV-----PFKDR-SIKHHIPIKPVQRG
Laf-PGBD7 91 PELMTKARFKKLTENIHCNDSSKAVPKGEPGYDWLHKLRPINALNTRLKEVYVPSVMAVDESMV-----PFKGRHSSVKQYMPIKPVKRG
Cgi-pble 83 EERFTRRFENIQRYFHVADTTQNPPRNQPGHDKLVHVRTIMETIRENFKROYNAHKEVVIDEAMI-----GFSGRLGFKQYVPLKPTKRG
Bgl-pble 85 ADVMGRTRFEKIRQYFHLNDSAHFPNRNSDNYDPLYKIRPLLDHVRSTCQAVYKPKQRKLSVDEAMV-----GFRGRLAMKQYLKPKPTPWG
Pfl-pble 84 AKVLPKNQYKKLWHYFHINNNELDLPAGDPNRDRMFKVRPLLDLVHFSFAASYVPFRELWIDEAMV-----KFKGHSYMIQYMPAKPVKKG
Ola-PGBD9 85 KKTMTARFEKLTQYLQLCDREAEVVRGERCYDGLYKIRPLLDVVENTMWDTYTPNRCLTIDKCSI-----ATKGRFSPTQYMSSKPVKKG
Hbu-PGBD9 85 KKTMTARFEKLTQYLQLCDRESEVVRGERGYDGLFKIRPLLDVVENTMWDAYVTPNRCLTIDKCAI-----VTKGRFSPTQYMSSKPLRKG
Pny-PGBD9 85 KKTMTARFEKLTQYLQLCDRESEVVRGERGYDGLFKIRPLLDVVENTMWDAYVTPNRCLTIDKCAI-----VTKGRFSPTQYMSSKPLRKG
Tcr-pble 78 SPLMPVNRFGWLLTHLHVNNVLPKRDDNAFDKLYKLRPMLDTLSEFCKTCKFGNHSVIDESMI-----KFKGRSSFRQFMPNPKIKRG
Bbe-PGBD5 87 ASIMTQRRFQHIMQYFHVVSS-----AQQQSDSLRKIRPFLEAVQNSFRNEFKTAQTHVHHEPPEEPP-----SVWSPRL-ERKRRKKR
Pma-PGBD5 83 AVAMTRIRFEKLLKYFHVVAFRS-----AREDLPLGLYKVPFLDSLKSSFEASFMPSTQVHLHEPLIEEDPTPLAACASGGCASSEGGERRRRKKR
Hsa-PGBD5 83 ALVMSQARFEKILKYFHVVAFRS-----SQTTHGLYKVPFLDSLQNSFDSAFRPSQTQVHLHEPLIDEDP-----VFIATCTER--ELRKRKKR
Oaf-PGBD5 83 ALVMSQARFEKILKYFHVVAF-----RSSQTTHGLYKVPFLDSLQNGFDSAFRPSQTQVHLHEPLIDEDP-----VFIATCTER--ELRKRKKR
Api1-pble 88 SHIMSRNRFELLSHIHFQNNND-----VANLNNRLYKIQNILDLNQLKFKQWVIPSHDMCIDESMI-----AFQGRLLVFKQYIQTKRYRYG
Api3-pble 89 KKTMSRNRFELLRFWHFSDNK-----KAPEGDRIYKIRNIIIDKIVNRFQNVMEPGEVLAVDETMV-----PFRGRLFRQYIPGKAHKYG
Atr1-pble 89 KKTMSRNRFVLLRMLHFSNNE-----EVPKNDRLYKIRRLVEILNQTFSIHYTPDECLVDDESMI-----PFRGRIIFRQYINKSKRHYG
Psi-PGBD8 88 KKLISRNRFEMLLRMWHFSNNEECPPG-----DRLFKIQPLLDRLERFQVAVVPDKEICIEETMV-----PFRGRLSFIQYVKNKCHKFG
Mmu-PGBD8 88 KKLISRNRFEMLLRMWHFSNNEECPPG-----DRLFKIQPLLDRLERFQVAVVPDKEICIEETMV-----PFRGRLSFIQYVKNKCHKFG
Dma-PGBD8 88 KKLISRNRFEMLLRMWHFSNNEECPPG-----DRLFKIQPLLDRLERFQVAVVPDKEICIEETMV-----PFRGRLNFVQYVKNKCHKFG
Pva-pble 90 RQIMTGERFLLLRCLHFVNNSHISAGQSKAQVSMQKIKPVDFLVNKFSTVYTPNRNIAVNESLM-----LFKRQLALROYIPMKRAQCG
Oab1-pble 90 GKIFSADRFLLLCKFLYFENNAPHN--DMPSKCLKIKTVLEYVINKCKSLYTPKMDICIDESLL-----MWKGRLSWRQYIPSKRSRFG
Hsa-PGBD4 90 RQIMTGERFLLLRCLHFVNNSHISAGQSKAQISLQKIKPVDFLVNKFSTVYTPNRNIAVNESLM-----LFKGRPLAMKQYLPKRVRF
Uribo1 89 SAVMPNRNYQILLRFLHFNDNAAAVPPNEPQHDRLYKLRPLIDSLSRFAEAVYTPSQNICVDESLL-----LFKGRLLFRQYIPSKRARYG
Uribo2 89 SATMSRNRYQLLLRFLHFNNNATAVPPDQPGHDLHKLRLPLIDSLSERFAAVYTPCQNICIDESLL-----LFKGRLLQFRQYIPSKRARYG
Pxu1-pble 89 RKTMRVDRYLLLMQLHFND-----NNIHCEDPLIKIRPIVDDLKKSFAESFYPGKKLCIDESLM-----LFKGRVYFKQYIPSKRSRFG
Xtr-PGBD6 89 SAIMPINRYQILLRFLHFNDNATAAPPNEPSHDLHKLRLPLIESLCKRFAEAVYTPQNICVNESPV-----FKKERLMFRQNNPKKRSRYG

β2

β3

α1

β4

α2

Tni-pble	167	IKILMMCD	SGTKY	MINGMPY	LGRGTQT	-----	NGV-PL	GEYYV	KEL	SKPV	HGSC	---	RNITC	DNWFT	-	SIP	LAKNL	LQEPY	KLTV	GTVRS						
Hmel2-pble	156	FKAVWGATRL	GYVLWMEPY	QGATTMCN	---PI-	YKELGL	GASVV	LTFCDV	LISR	GFDLPY	HVVV	-	NFFT	-	GTP	LEEIT	TKK-	GLR	CT	GTVRE						
Csu-pble	153	FKAVWGATRL	GYVLWMEPY	QGATTM	---CN	PIYKELGL	GASVV	LTFCDV	LISR	GFDLPY	HVVV	-	NFFT	-	GTP	LEEIT	TKK-	GLR	CT	GTVRE						
Smi7-pble	158	YKFWCAATRL	GYICWFEPY	QGNL	--A--	NK--	YEEYGV	GASVV	LRFTDA	LTEQH	PGQY	HVFV	-	DNFFT	-	SIAL	LDKLG	SM--	GHQ	ATGTVRK						
Lfu-pble	157	YKFWCGATRL	GYICWFQPY	QGNP	--N--	TN--	YEGYGV	GASVV	LQFTEA	LTKEH	PGQY	HVFV	-	DNFFT	-	SIVL	LDKLS	SM--	GHQ	ATGTVRK						
Hvu-pble	158	FKFWCGATPL	GCICWFQPY	QGNP	--N--	TQ--	FQEYGV	GASVV	LQFVEA	LTEEH	PGEY	HVFV	-	DNFFT	-	SISL	LDKLS	SM--	GHV	TGTVRK						
Ppa-PGBD3	158	YKFWCGATCL	GYICWFQPY	QGNP	--N--	TK--	HEEYGV	GASLV	LQFSEA	LTEAH	PGQY	HVFV	-	NNFFT	-	SIAL	LDKLS	SM--	GHQ	ATGTVRK						
Hsa-PGBD3	158	YKFWCGATCL	GYICWFQPY	QGNP	--N--	TK--	HEEYGV	GASLV	LQFSEA	LTEAH	PGQY	HVFV	-	NNFFT	-	SIAL	LDKLS	SM--	GHQ	ATGTVRK						
Ota-pble	158	FKIWCATRL	GYLVNVDPY	TGKNY	---KE	KDDLNLGL	GGQLV	YKFAKI	LKSQN	PNPFH	LFC	-	DNFFT	-	SIA	LSE	QKI-	GIR	ATG	TIKE						
Hsa-PGBD2	158	YKIWCCTSR	GYLVWFEPY	QGTFL	--T--	KP--	DRSLDL	GGSMV	IKFVDA	LQER	GFLPY	HIFD	-	DKVFT	-	SVKL	MSIL	RKK-	GVK	ATGTVRE						
Ppa-PGBD2	158	YKIWCCTSR	GYLVWFEPY	QGTFL	--T--	KP--	DRSLDL	GGSMV	IKFVDA	LQER	GFLPY	HIFD	-	DKVFT	-	SVKL	MSIL	RKK-	GVK	ATGTVRE						
Tma-PGBD2	158	YKIWCCTSR	GYLVWFEPY	QGTFL	--T--	KP--	DKGLDL	GGSMV	IKFVDA	LQER	GFLPY	HIFD	-	DKVFT	-	SVKL	MSIL	RKK-	GVK	ATGTVRE						
Tma-PGBD1	154	YKIWCCTTQ	GYLVWFEPY	QEESTVTA	---DK--	DLDLGI	GGNLV	MNFADV	LLER	GQPY	HLCF	-	DSFFT	-	SVKL	MSAL	KKK-	GLR	AT	GAIRE						
Hsa-PGBD1	154	YKIWCCTTQ	GYLVWFEPY	QEESTMKV	---DE--	DPDLGL	GGNLV	MNFADV	LLER	GQPY	HLCF	-	DSFFT	-	SVKL	LSAL	KKK-	GVR	AT	GAIRE						
Oaf-PGBD1	154	YKIWCCTTQ	GYLVWFEPY	QEESTMKA	---DK--	DLDLGL	GGNLV	MNFADV	LLER	GQPY	HLCF	-	DSFFT	-	SVKL	VSAL	KKK-	GLK	AT	GAIRE						
Nvi-pble	165	FKVWCLADSS	TYVVKSEVY	TGKARD	GS-----	KNPL	GERVV	LNLSN	ILGV	---	YLVAF	-	DNFFT	-	SVNL	VEEL	LDK-	GIY	SV	GTVRN						
Smi2-pble	177	YKVVWCLADSKT	GYIQKFEIY	CGKETS	DN--K--	AKKLN	LGETV	VTKL	TEDV	KHTG	---	RLIAH	-	DNFFT	-	SFRL	LAKK	LLAS	-	GLF	SVGTVKP					
Oaf-PGBD7	177	YKVVWCLADSKT	GFVSQFDIC	CGKKGV	---DN-	NSSLS	FGESV	VGL	CHSWY	HSH	---	RLIS	-	DSYFT	-	SYHL	MKSM	KER	-	GLY	AVGTVRA					
Cas-PGBD7	176	YKVVWCLADSKT	GFVSQFDIC	CGKKGI	---DN-	DSTFSL	FGESV	VGL	CHSWY	HSH	---	RLIAF	-	DNFFT	-	SYHL	MKSM	NER	-	GLY	AVGTVRA					
Laf-PGBD7	177	YKVVWCLADSKT	GFVTQFDIY	CGKKGV	VD--H--	DSSLS	LFGESV	VGL	CHSWY	HSH	---	RLIS	-	NNFFT	-	SYHL	MKSM	MYER	-	GLY	AVGTVRA					
Cgi-pble	169	IKVWVRADPHN	GFMNDFQVY	TGKEGN	---VPE	AN--	GGRVV	LDL	MDSV	LNLC	---	HHVY	-	DSFFS	-	SPDL	FLKL	WNE	-	GTV	ACGTVKP					
Bgl-pble	171	IKIWCCEASS	GYVLFNFRVY	TGKSNY	---S	PNPGL	GYNVV	LDL	MSPE	YLDKY	---	HEVY	-	DNFFS	-	SLKL	LAED	LLER	-	GLT	YSCATIRA					
Pfl-pble	170	FKVWMLACPR	TYCLSLDPY	TGKKGGQ	---VA	QKGL	GYSV	VWNL	SRMY	HNMW	---	HILFY	-	DNFFS	-	SVLS	RALD	LQKQ	-	GT	LCCSTVVRT					
Ola-PGBD9	171	LTVMMLCDSRS	GYCHRTKIHV	GRPS	---ED-	EGAAS	LGH	RVV	TFLV	RGLERQY	---	HHLFM	-	DSFFT	-	SVPL	LQRL	LRD-	-	GLY	ACGPTQP					
Hbu-PGBD9	171	LTVMMLCDSRS	GYCHRAKIYV	GRPS	---ED-	EAAACL	GYR	VV	TSLV	RGLEQY	---	HHLFM	-	DSFFT	-	SVPL	LQRL	LRD-	-	GQY	ACGPTQP					
Pny-PGBD9	171	LTVMMLCDSRS	GYCHRAKIYV	GRPS	---ED-	EAAACL	GYR	VV	TSLV	RGLEQY	---	HHLFM	-	DSFFT	-	SVPL	LQRL	LRD-	-	GQY	ACGPTQP					
Tcr-pble	164	YKVVVRADSE	GFVCQFIY	TGK	---V-	EG--	NIEKLL	GAR	VKDL	TLELV	GLG	---	HHVY	-	DNFFT	-	SVDL	MISL	KNQ-	-	AIY	ACGTVRN				
Bbe-PGBD5	168	FNLFVRQCTAT	GFISQIYIHV	K-----	DKND	NQTLHEK	GYSI	ARAL	CSDL	INSG	---	RNHIF	-	CEPLHT	-	SLRL	FREL	ESR	-	NIY	GCGVITD					
Pma-PGBD5	174	FGLWVRQCSQT	GFICQIHVHL	KESGSS	GS	DCLD	AFRS	KSQ	LHGL	VARR	LCHNL	-	-	PPSG	---	YVIF	C	PSIT	-	SLPL	FQEFQEK	-	GIH	CCGLLSP		
Hsa-PGBD5	166	FSLWVRQCSST	GFIIQIYVHL	KEGGGP	--DGL	DALK	NKPK	QLHSM	VARS	LCRN	--AAG-	-	-	KNYIIFT	-	GPSIT	-	SLTL	FEFE	EKQ-	-	GIY	CCGLLRA			
Oaf-PGBD5	166	FSLWVRQCSST	GFIIQIYVHL	KEGGGP	--DGL	DALK	NKPK	QLHSL	VAKSL	CRNA	AAGKN	---	-	YIIFT	-	GPSIT	-	SLTL	FEFE	EKQ-	-	EIY	CCGLLSS			
Api1-pble	169	VKV-FKLCVSP	CYTLKFKIY	SGKESV	V-----	DKNDS	VSSRI	V	MNLV	DEYLDL	G	---	-	RTLVT	-	DNWYT	-	SVSL	AHQ	LLER	-	KTH	LVGTLRS			
Api3-pble	170	VKL-FKLCGTN	GYTYNVQVY	YGK	--SQ-	-----	VDGKGL	GCR	VV	LDL	SRRYLN	MG	---	-	RTMVT	-	DNFYT	-	SITL	ANEL	LSY-	-	DTH	LVGTLRS		
Atr1-pble	170	IKL-FKLCCTP	GYTKVFRVY	YGK	--TL-	-----	DSQDNT	PTNV	VNL	LNARY	FGK	G	---	-	HTMFV	-	DNWYT	-	SIDL	AEKL	VQH-	-	DTH	LVGTLRK		
Psi-PGBD8	169	IKLFLKCLKD	GYTWNVKIY	CGK	-----	EA-	RSGIP	VSSSV	V	MKLI	IEGL	LLDSG	---	-	-	HVLYT	-	DNYYT	-	SIHL	AHQ	LLER	-	STY	LVGTLRS	
Mmu-PGBD8	169	IKLFLKCLKD	GYTWNVKIY	CGK	-----	EA-	RSGIP	VSSSV	V	MKLI	IEGL	LLDSG	---	-	-	HVLYT	-	DNYYT	-	SIHL	AHQ	LLER	-	STY	LVGTLRS	
Dma-PGBD8	169	IKLFLKCLKD	GYTWNVKIY	CGK	-----	EA-	RSGIP	VSSSV	V	MKLI	IEGL	LLDSG	---	-	-	HVLYT	-	DNYYT	-	SIHL	AHQ	LLER	-	STY	LVGTLRS	
Pva-pble	176	LKLYVLCESQ	SGYVWVALVHT	GPSMNLK	--NS-	ADGL	KSSRI	V	TLV	SDLL	LGQ	G	---	-	YCVFI	-	GNFYT	-	SPML	FREL	HQN-	-	RTD	AVGTACL		
Oab1-pble	173	IQFFVLCESSE	GYIWNFFIY	TGKETYCD	--SQ-	YSEFN	SARIV	LQL	CDEL	FERG	---	YRLYL	-	DNWYT	-	GVPF	IEKL	CAH-	-	KTD	VVGT	IRK				
Hsa-PGBD4	176	LKLYVLCESQ	SGYVWVALVHT	GPGMNLK	--DS-	ADGL	KSSRI	V	TLV	NDLL	LGQ	G	---	-	YCVFL	-	DNFNI	-	SPML	FREL	HQN-	-	RTD	AVGTARL		
Uribo1	175	MKLYKLCESST	GYTSFFMFY	EGKDSNLD	--PPG	PLD	TASG	KIV	WELI	TPL	LG	G	---	-	YHLYV	-	DNFYT	-	SILL	FRT	LYCL	-	-	DTLACGTVRQ		
Uribo2	175	IKFYKLCESSS	GYTSYFLIY	EGKDSKLD	--PPG	CPD	LT	VS	GKI	V	WELI	SP	-	-	-	FHLYV	-	DNFYS	-	SIP	L	F	TALYCL	-	-	DTPACGTINR
Pxu1-pble	169	IKSFVLCCKT	NVYVLYNIIY	TGSKTECD	--ST-	FSNVG	KSGD	V	V	M	SLIK	PYLDK	G	---	-	HTLIT	-	DNWYS	-	SPNL	F	D	LLHKM	-	-	KTNAFGTVRK
Xtr-PGBD6	175	MKLYKLCESKT	GYSGNFMIE	GTDSFLD	--PPG	CLD	F	T	S	D	KI	V	W	-	-	YHLYV	-	DNFYT	-	SIP	L	F	RALHSL	-	-	DTPACGIVSH

α3

α4

Tni-pble 311 MYYNQTKMGVDTLDQMCSVMTCSRKT-NRWPAL-LYGMINIAACINSFIIYSHNVSSK- - - - -GKVVQS-RKKFMRNLYMSLTSSFMKRRL
Hmel2-pble 316 SIYNKYMGVDRSDENISHYRIGIRG-KKWY-MPLLTHMIDLAEHNAWQLYKIN-H- - - - -GKLD-HLGFRRRVAIALIESNRKNAK
Csu-pble 316 SIYNKYMGVDRSDENISHYRIGIRG-KKWY-MPLLTHMIDLAEHNAWQLYKIN-H- - - - -GKLDH-LGFRRRVAIALIESNRKNAK
Smi7-pble 317 KVYNQFMGVDRADENIDKYRASICG-KKWYSSPLLFC-FELALQNAWQLHKTY-DA- - - - -KPM-DLLEFRRRVCHYLETHGHPVE
Lfu-pble 316 KGYNQYMGVDRADENIDKYRASIRG-KKWYSSPLLFC-FELALQNAWQLHKTY-DT- - - - -KPM-DLLEFRRRVACHYLETHGYPVQ
Hvu-pble 317 KLYNQYMGGLDRADENIGQYQASICG-KKWYSSPLLFC-FEVILQNAWQLLHRTI-EE- - - - -KPLD-LLEFHRRRIASNYLETYGPPE
Ppa-PGBD3 317 KVYNQFMGVDGRADENIDKYRASIRG-KKWYSSPLLFC-FELVLQNAWQLHKTY-DE- - - - -KPWD-FLEFRRRVVCHYLETHGHPPE
Hsa-PGBD3 317 KVYNQFMGVDGRADENIDKYRASIRG-KKWYSSPLLFC-FELVLQNAWQLHKTY-DE- - - - -KPWD-FLEFRRRVVCHYLETHGHPPE
Ota-pble 320 TL- - - - -
Hsa-PGBD2 318 KLYQEKVGVGGRMDQNIACYKVKIRG-MKWYSSFIGYV-IDAALNNAWQLLHRICCD- - - - -AQVD-LLAFRRRYIACVYLESNADTTS
Ppa-PGBD2 318 KLYQEKVGVGGRMDQNIACYKVKIRG-MKWYSSFIGYV-IDAALNNAWQLLHRICCD- - - - -AQVD-LLAFRRRYIACVYLESNADTTS
Tma-PGBD2 318 RLYQEKVRGVGGRMDQNIACYKVKIRG-MKWYSSFIGYV-IDAALNNAWQLLHRICCD- - - - -AQVD-LLAFRRRYVACVYLESNADTSS
Tma-PGBD1 315 KLYDECREGVAKMDQIISKYRVIRRS-KKWYSILVSYM-IDLAMNNAWQLLHRACNPG- - - - -ASLD-LLDFRRRCVAHFYLEHNAHLS
Hsa-PGBD1 316 KVYDECKEGVAKMDQIISKYRVIRRS-KKWYSILVSYM-IDVAMNNAWQLLHRACNPG- - - - -ASLD-PLDFRRRFVAHFYLEHNAHLS
Oaf-PGBD1 315 KLYDECREGVAKMDQIISKYRVIRRS-KKWYSILVSYM-IDVAMNNAWQLLHRACNPG- - - - -ASLD-LLDFRRRCVAHFYLEHNAHLS
Nvi-pble 316 QAYNKIMGVDVDFDLREVYMIERRS-KKWWHSIMYYY-IDMAIVNCYILYKCYKEG- - - - -ECED-QLAFRIQLAKQLIGDYNGRKR
Smi2-pble 331 AVYNDVMGVDVDFDQRKERYQIGRRS-VKWWHRIFYFL-LDLGIINSFILWQVNRN- - - - -RNL-DLQTFRIALARQLIDGYS-SRK
Oaf-PGBD7 330 AEYNAIMGVDVDFDQSCERYAIGRRS-RKWWHRLHFL-VDLAIVNSFTMWNCH-G- - - - -GRCN-QLSFRLLALIRQLTVGRKRKR
Cas-PGBD7 329 AEYNSIKGVDVDFDQRRERYAIGRRS-RKWWHRLHFL-IDLAIVNSFIMWNCN-HG- - - - -GRCN-QLSFRLLALSRLTVGRKRKR
Laf-PGBD7 330 AEYNAIMGVDVDFDQRCERYAIGRRS-RKWWHRLHFL-IDLAIVNSFIMWNCN-HG- - - - -GRCN-QLSFRLLALSRLTVGQKRKR
Cgi-pble 314 KLYNAYMNGVDHADQLRSTYNIARKS-LKWWKYLFL- - - - -YDVCIVNAFILMRESPNHGPRTKNRRVTRTQLEFRKVAHQLILFS- - - - -
Bgl-pble 317 VYNSGMGVDLADQHRSYYPVGRVS-NKWWRYMLWYL-IDISMINSFLIMKKS- - - - -PNDSTDSLQNHLLFHKKVADQLLTA-TGVTL
Pfl-pble 315 IDYNKFMGVDLSDQLATYYPGGK-HYKYWKYLFWFI-INRAICNAWIMYKIAN- - - - -PP- - - - -PRHRYR-QKDFRLDLSTQLVAGFTARKR
Ola-PGBD9 336 LLYQENMRGVDLSDQLRECYQVGRPC-KKWWRYFLWFI-VNLCIVNAYIVMRESRGGAPPAGFN-GKQFTQRHFRVRLAQLI-GD-YQGA
Hbu-PGBD9 339 LLYQENMRGVDLSDQLRECYQVGRPC-KKWWRYFLWFI-VNLCIVNAYIILRESRGGAPPAGFN-GKQFTQRHFRVRLAQLI-GD-YQGA
Pny-PGBD9 339 LLYQENMRGVDLSDQLRECYQVGRPC-KKWWRYFLWFI-VNLCIVNAYIILRESRGGAPPAGFN-GKQFTQRHFRVRLAQLI-GD-YQGA
Tcr-pble 315 KDYNQHMGYVDKADMLKSLYEISRKC-RKWWHRIFWHF-VDVALTNGFVIYALQSEG- - - - -THLS-LKKFRISVVDGLIGATQSSPR
Bbe-PGBD5 329 QAYQNHICYIDKYNKFKSWYIISYKP-NKTWQQVFWLA-LSIAINNAWILYKLSPAY- - - - -QAQRYG-RAQFGERLVQQLRIPEANGP
Pma-PGBD5 338 EAFAAHLGYICKYDDKYSRYFISHKP-NKTWQQVFWFA-LSIAINNAWILYKMSDAY- - - - -HVKRYS-RAQFGERLVRELLGLEDASPT
Hsa-PGBD5 325 EAFAAHLSYICKYDDKYSRYFISHKP-NKTWQQVFWFA-LSIAINNAWILYKMSDAY- - - - -HVKRYS-RAQFGERLVRELLGLEDASPT
Oaf-PGBD5 325 EAFAAHLSYICKYDDKYSRYFISHKP-NKTWQQVFWFA-LSIAINNAWILYKMSDAY- - - - -HVKRYS-RAQFGERLVRELLGLEDASPT
Api1-pble 313 LDYNRCKSFIDLSDQMASYGSPLRKS-LKWYRKV-FELLNLSVVALCLFQKTTA- - - - -TRIK-ITDFRSNLIKYLTFKPNMNQE
Api3-pble 315 VDYNSGKAGIDLSDQLSSYSTAVRKS-IRWYHKVATEILFGTAMVVALIVYNTINS- - - - -DKMK-ITQFRETLDVDTILGLDNQGPQ
Atr1-pble 314 IDYNKAKAAVDLSDQMSAYSSPLRKT-LKWWKLGIELLNTAVVNAVLYESTTK- - - - -KKIS-VVNFRRLLAKQLIEGQENTG
Psi-PGBD8 312 VDYNSKAFIDLSDQLKSYSCSLRRG-IKWFRLKLAIELITGSALVNSLVLYRVV-NK- - - - -KKIS-ITKFKEQVCLKLLQTDDETSTA
Mmu-PGBD8 312 VDYNSKAFVDSLSDQLKSYSCSLRRG-IKWFRLKLAIELITGSALVNSLVLYRVV-NK- - - - -KKIS-ITKFKEQVCLKLLQTDDETSTV
Dma-PGBD8 312 VDYNSKAFIDLSDQLKSYSCSLRRG-IKWFRLKLAIELITGSALVNSLVLYRVV-NK- - - - -KKIS-ITKFKEQVCLKLLQTDDETPTA
Pva-pble 323 VDYNNKMGVVDLADQMLASYPTECKRHKFWYKFFRHL-LNITVLNSYILFKKDNPE- - - - -HTIS-HVNFRLTLIERMLEKHHKPGQ
Oab1-pble 321 FDYNNKMGVVDMSDLSIIAYSTARKRLKYYKKIFLHL-LDVICLNSYLIYKMN- - - - -GKLS-RIHFLLLEYIEDTIASYPIESR
Hsa-PGBD4 323 VDYNNEMGAVDSDQMLTSYPSERKRHKVYKFFHHL-LHITVLNSYILFKKDNPE- - - - -HTMS-HINFRLALIERMLEKHHKPGQ
Uribo1 326 KDYSKHMGGVDKTDQIQTYYPDATRKT-RAWYKKAAYM-IQMALNSYVYKAAVPG- - - - -PKLS-YYNFLLQLLPAFLFGD- - - - -VQE
Uribo2 326 KEYSKYMGGVDRTDQLQHYYNATRKT-RAWYKKGVIYL-IQMALNSYIVYKAAVPG- - - - -PKLS-YYKYQLQILPALLFGGVVEEQT
Pxl1-pble 319 VDYNNLMGSDVDMVDMVLSAINSCRKS-LKWYKKYFFHL-LDMSIYNSYILYQOVTN- - - - -KCLK-YSKPHLILIKQILQKYPQENK
Xtr-PGBD6 327 KEYSRHIDGVNKTDDQIQYY-NAARKT-KAWYKKAAYM-IQMALNSYVYKATVPG- - - - -PKLS-FYNYQVQLLPSLLFGD- - - - -VEE

Supplementary data 4 : Transposon donor sequences and primers used in the study.

a. Sequence of the *Ifp2-TIR5'-NeoR-TIR3* donor (the cassette containing the promoter SV40 - NeoR ORF - sv40 terminator is highlighted in yellow)

```
TTAACCCCTAGAAAGATAATCATATTTGTGACGTACGTTAAAGATAATCATGCGTAAAAATTGACGCAT
GTGTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTTATTATAT
TTACACTTACATACTAATAATAAAATTCACAAACAATTTATTTATGTTTATTTATTTATTAAAAAA
AAACAAAAACTCAAATTTCTTCTATAAAGTAACAAAACTTTTGGTATCGATGTTCGACGAATTC TG
TGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGCTCCCCAGGCAGGCAGAAGTATGCAAAGC
ATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATG
CAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTA
ACTCCGCCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTTTATTTATGCAGAGGCC
GAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTTGGAGGCCTAGGCTTT
TGCAAAAAGCTCCCGGGAGCTTGGATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCG
TTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTC
GGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAG
GGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAAGTGCAGGACGAGGCA
GCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACGTGAA
GCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTATCTCACCTTGCT
CCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTT
GTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTGTTCGCCAGGCTC
AAGGCGCGCATGCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGCCGAATATC
ATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGGCAATGGGCTGACCGCTTC
CTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAG
TTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCACGAG
ATTTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTTCCGGGACGCCGGCT
GGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTATTATGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGC
ATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCCCAgctagcTGG
CTCGAGCAATTGGCAGATCTAAGCTTTGATATCTATAACAAGAAAATATATATATAATAAGTTATCA
CGTAAGTAGAACATGAAATAACAATATAAATTATCGTATGAGTTAAATCTTAAAGTCACGTAAAAG
ATAATCATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAATCAGTGACACTTACCGCATTG
ACAAGCACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCTAAATGCACAGCGACGGATTCCG
CGCTATTTAGAAAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTC
TAGGGTTAACCGCGGTGGAGCTCCAGCTTTTGTTCCTTTAGTGAGGGTTAA
```

b. *Ter-pble-TIR5'-NeoR-TIR3* donor (the cassette containing the promoter SV40 - NeoR ORF - sv40 terminator is highlighted in yellow)

```
TTAACCCCTTTGAAGCACACATTAAGGATTTAATTTATTTTTTAATTTTTTACAATGAAAACCTATTT
CTAAATACATTTAAAATTTAATTTGAGCAAAAATAAAATGTTTTTCATTAATATTGAGTGTGCTAA
AAAATGGTGGTCATTCAGTGAACCACTATGCAACTAAGCTTCACTATATTTTTTACATGGTGTGGT
ATGCTACAAAGCAATTTTTAGATTCATTGAGACACAATGCCACTTCGCATGCTTTACAAGACCATT
TACTTCGGTGTGGCTGTAACCTTCTGTGCTGCAGTGGGTTAACATGAATTCGTGGAATGTGTGTCA
GTTAGGGTGTGGAAAGTCCCCAGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA
GTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTC
CGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCC
GGGAGCTTGGATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGA
ACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGC
```

ACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCT
TTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTG
GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGT
ATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCA
CCAAGCGAAACATCGCATCGAGCGAGCACGTA CTGGATGGAAGCCGGTCTTGTTCGATCAGGATGA
TCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGAGGCTCAAGGCGCGCATGCC
CGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGG
CCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTT
GGCTACCCGTGATATTGCTGAAGAGCTTGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGG
TATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGG
ACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCACGAGATTTTCGATTCACC
GCCGCCTTCTATGAAAGGTGGGCTTCGGAATCGTTTTCCGGGACCGCGGCTGGATGATCCTCCAG
CGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTAC
AAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGT
TTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCACGTGACAATAGTATTAAATAATGA
TTTGAATACTTAGGGCTACATAGTTATATATTTATTCCTTGAAAAATATAACAAAACCAATTAGA
AGAAACGTGTCACCTCAATAACAAAACAAAATACATACTATATTGGAGGCATGAATACCAACAAA
AATTACTAATAATAAATTTTTACAAGAATAAACTAGCAAATAAATTCAGATAGATGATTCATAAG
TGCTGCCATACAGTTAAATGTATCTCAAATAAATAAAGGTGGTTACCAAGTGAAACCACTA
TGCTTCAAAGGGTTAA

c. Sequences used as baits to identify transposon breakpoints in analyses done with lumpy.

Sequence of the *Ifp2-TIR5'-NeoR-TIR3* sequence flanked upstream and downstream by 100-bp plasmid regions. The cassette containing the promoter SV40 - NeoR ORF - sv40 terminator was highlighted in yellow. TTA TSD were bolded while 100-bp plasmid backbone sequences were typed in blue.

ATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGC
GTAATACGACTCACTATAGGGCGAATTGGGTACC**TTAAC**CCCTAGAAAGATAATCATATTGTGACGT
ACGTTAAAGATAATCATGCGTAAAATTGACGCATGTGTTTTATCGGCTGTATATCGAGGTTTATT
TATTAATTTGAATAGATATTAAGTTTTATTTATTTTACACTTACATACTAATAATAAATTCACAA
ACAATTTATTTATGTTTATTTATTTATTTAAAAAAAACAAAACCTCAAATTTCTTCTATAAAGTA
ACAAAACCTTTTGGTATCGATGTCGACGAATTCTGTGGAATGTGTGTGTCAGTTAGGGTGTGGAAAGTC
CCCAGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGG
AAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAT
AGTCCCGCCCCTAACTCCGCCATCCCGCCCCTAACTCCGCCAGTTCCGCCATTCTCCGCCCA
TGGCTGACTAATTTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCTCTGAGCTATTCCAGAA
GTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGGATATCCAT
TTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGC
AGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTG
CTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCT
GTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGT
TCCTTGCGCAGCTGTGCTCGACGTTGTACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGT
GCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGC
AATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCAT
CGAGCGAGCACGTA CTGGATGGAAGCCGGTCTTGTTCGATCAGGATGATCTGGACGAAGAGCATCA
GGGGCTCGCGCCAGCCGAACTGTTCCGAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGT
CGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCAT
CGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGC
TGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATT
GCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATG
ACCGACCAAGCGACGCCAACCTGCCATCACGAGATTTTCGATTCACCGCCGCTTCTATGAAAGG

TTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTG
GAGTTCCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATC
ACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAAT
GTATCTTATCATGTCTGGATCCCCAgctagcTGGCTCGAGCAATTGGCAGATCTAAGCTTGATATC
TATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGAACATGAAATAACAATATAATTA
TCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTCATTTTACTCACGCGGTCCG
TTATAGTTCAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCAAGCGGCG
ACTGAGATGTCCTAAATGCACAGCGACGGATTTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAG
AATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGG**TTAACCGCGGTGGAGCTCCAGCTTTTG**
TTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAA
TTGTTATCCGCT

Sequence of the *Ter-pble-TIR5'*--*TIR3* sequence flanked upstream and downstream by 100-bp plasmid regions. TTAA TSD were bolded while 100-bp plasmid backbone sequences were typed in blue.

CGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGGTACGGCC
GTCAAGGCCAAGCTTCCCGGGATACTTACCATGG**TTAAC**CCCTTTGAAGCACACATTAAGGATTTAA
TTTTATTTTTAATTTTTTTACAATGAAAACCTTATTTCTAAATACATTTAAAATTTAATTTGAGCAAAA
ATAAAATGTTTTTCATTAATATTGAGTGTGCTAAAAAATGGTGGTCATTCAGTGAAACCACTATGC
AACTAAGCTTCACTATATTTTTACATGGTGTGGTATGCTACAAAGCAATTTTTAGATTCATTGAGA
CACAATGCCACTTCGCATGTCTTACAAGACCATTACTTCGGTGTGGCTGTAACCTCTGTGCTGCA
GTGGGTAAACATGAATTCGAGCTCGGATCCACGTGACAATAGTATTAATAATGATTTGAATACT
TAGGGCCTACATAGTTATATATTTATTCCTTGAAAAATATAACAAAACCAATTAGAAGAAACGTGT
CACTTCAATAACAAAAACAAAATACATACTATATTTGGAGGCATGAATACCAACAAAATTACTAAT
AATAAATTTTTACAAGAATAAACTAGCAAATAAATTCAGATAGATGATTCATAAGTGCTGCCATA
CAGTTAAATGTATCTCAAATAATTAATAAAAGGTGGTTACCAAGTGAAACCACTATGCTTCAAAG
GG**TTAACTCGAGTAAGTATCACGTGAAGCTTTGCAAGCTCCAGCTTTTGTTCCTTTAGTGAGGGT**
TAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCC

Sequence of the *Ifp2-TIR5'*-*NeoR-TIR3* sequence flanked upstream and downstream by 100-bp plasmid regions. This sequence was that contained in the *Ifp2* plasmid construct used in Henssen et al, 2015 to monitor integration assays in HEK293 cells. The cassette containing the promoter EF1 - NeoR ORF - sv40 terminator was highlighted in yellow. TTAA TSD were bolded while 100-bp plasmid backbone sequences were typed in blue.

TTGAACCCGTGGAGGACGGGACGACTCGCGGTGCAAATGTGTTTTACAGCGTGATGGAGCAGATGA
AGATGCTCGACACGCTGCAGAACACGCAGCTAGAT**TTAAC**CCCTAGAAAGATAATCATATTGTGACGT
ACGTTAAAGATAATCATGTGTAATAATTGACGCATGTGTTTTATCGGTCTGTATATCGAGGTTTATT
TATTAATTTGAATAGATATTAAGTTTTATTATTTTACACTTACATACTAATAATAAATTTCAACAA
ACAATTTATTTATGTTTTATTTATTTATTTAAAAAAAACAAAACCTC**AAAATTTCTTCTATAAAGTAA**
CAAACTTTTATGAGGGACAGCCCCCCCCAAAGCCCCCAGGGATGTAATTACGTCCCTCCCCCGC
TAGGGGGCAGCAGCGAGCCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGAGCC
GGCAGCGTGCGGGGACAGCCCGGGCACGGGAAGGTGGCACGGGATCGCTTTCCTCTGAACGCTTC
TCGCTGCTCTTTGAGCCTGCAGACACCTGGGGGATACGGGGAAAAGGCCTCCACgGCCaaggatc
tgcgatcgcctccggtgcccgcagtgggcagagcgcacatcgcccacagtccccgagaagttgggg
ggaggggtcggcaattgaacgggtgcctagagaaggtggcgcggggtaaacgaggaaagtgatgctc
gtgtactggctccgctttttcccaggggtgggggagaaccgtatataagtgcagtagtcgcccgtg
aacgctctttttcgcaacgggtttgcccgcagaacacagctgaagcttcgaggggctcgcactctc
ccttcacgcgcccgcgcccctacctgaggccgccatccacgcccgttgagtcgcttctgccgctc
cccgcctgtggtgcctcctgaactgcgtccgcccgtctaggttaagtttaagctcaggtcagagaccg
ggcctttgtccggcgctcccttgagcctacctagactcagccggctctccacgctttgcctgacc
ctgcttgctcaactctacgtctttgtttcgttttctgttctgcgcccgttacagatccaagctgtga
ccggcgccctactctagagctagcgaattcgaatttaaatcggatccGCGGCCGCgcccctctcct

ccccccccctaactggtactggccgaagccgcttgggaataaggccgggtgtgcggtttgtctatatgt
tattttccaccatattgcccgtcttttggcaatgtgagggcccggaaacctggccctgtcttcttga
cgagcattcctaggggtctttcccCtctcgcCAAaggaatgcaaggtctggtgaatgtcgtgaagg
aagcagttcctctggaagcttcttgaagacaacaacgtctgTagcgacccttgcaggcagcga
acccccacctggcgacaggtgacctctgcggccaaaagccacgtgtataagatacacctgCaaagg
cggcacaaccccagtgccacgttgtgagttggatagttgtggaaagagtcaaattggctctcctcaa
gcgtattcaacaAggggctgaaggatgccagaaggtaccccattgtatgggatctgatctggggc
ctcgggtgcacatgctttacatgtgtttagtCgaggttaaaaaaacgtctaggccccccgaaccacg
gggacgtggttttcctttgaaaaaCACGATGATAATATGGCCACAACCATGgcgtccggaATGATT
GAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGG
GCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGGCGCCCGGTT
CTTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCG
TGGCTGGCCgCGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGAC
TGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCACTCACCTTGCTCCTGCCGAGAAA
GTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCATTCGAC
CACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGAT
GATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGACAGGCTCAAGGCGCGCATG
CCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCGTGCTTGCCGAATATCATGGTGGAAAAT
GGCCGCTTTTCTGGATTCACTGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCG
TTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTAC
GGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGtc
gacaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatggtgctcct
tttacgctatgtggatacgtgctttaatgcctttgtateatgcGTTAACTaaactTGTTTATTGC
AGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACT
GCATTCTAGTTGTGGTTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGAATTGACTCAAATG
ATGTCAATTAGTCTATCAGAAGCTCATCTGGTCTCCCTTCCGGGGGACAAGACATCCCTGTTTAAAT
ATTTAAACAGCAGTGTTCCTAAACTGGGTCTTATATCCCTTGCTCTGGTCAACCAGGTTGCAGGG
TTTCTGTCTCACAGGAACGAAGTCCCTAAAGAAACAGTGGCAGCCAGGTTTAGCCCCGGAATTG
ACTGGATTCTTTTTTAGGGCCCATTTGGTATGGCTTTTTCCCGTATCCCCCAGGTTGTCTGCAGG
CTCAAAGAGCAGCGAGAAGCGTTCAGAGGAAAGCGATCCCGTGCCACCTTCCCCGTGCCCGGGCTG
TCCCCGCACGCTGCCGGCTCGGGGATGCGGGGGGAGCGCCGGACCGGAGCGGAGCCCCGGGCGGCT
CGCTGCTGCCCCCTAGCGGGGGAGGGACGTAATTACATCCCTGGGGGCTTTGGGGGGGGGCTGTCC
CTGATATCTATAACAAGAAAATATATATAATAAGTTATCACGTAAGTAGAACATGAAATAACAA
TATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAGATAATCATGCGTCATTTTGACTCA
CGCGGTGCTTATAGTTCAAATCAGTGACACTTACCGCATTGACAAGCACGCTCACGGGAGCTCC
AAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGCAAT
ATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTAATCTAGCTGCATCAGG
ATCATATCGTCGGGTCTTTTTCCGGCTCAGTCATCGCCAAAGCTGGCGCTATCTGGGCATCGGG
AGGAAGAAGCCCGTGCCTT

d. Primers used to amplify the Mm524 cDNA by RT-PCR.

Mm524 (Accession N°XM_006530804.1)

Forward primer (flanked with *EcoRI*) : 5'-CCGAA**GAATTC**ATGGCCGAGGGCGGTGG-3'

Reverse primer (flanked with *XbaI*) : 5'-CCGTT**TCTAGAT**CAGTGGGCTGGCGATGAG-3'

e. Primers to amplify loci containing *Ifp2* or *Ter-pble* insertions

First linear amplification (B = biotin)

(B)-NeoR 5' : (B)-TTCTTCGCCACCCCAACTTGTTTATTGC-3'

(B)-NeoR 3' : (B)-GACTTTCCACACCCTAACTGACA-3'

Linkers

Oligo I : 5'-GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG-3'

Oligo II DpnII : 5'-GATCCCTAACTGCTGTGCCACTGAATTCAGATC-3'

Oligo II PNB : 5'-CATGCCTAACTGCTGTGCCACTGAATTCAGATC-3'

First round of nested PCR

Primer anchored in linker

LC1 : 5'-GACCCGGGAGATCTGAATTC-3'

Primers anchored in the 5' or 3' end of the *IFP2* vector

(B)-TIR-UTR 5' PB III : (B)-CGATAAAACACATGCGTCAAT -3'

(B)-TIR-UTR 3' PB III : (B)-GCTATTTAGAAAGAGAGAGCA -3'

Primers anchored in the 5' or 3' end of the *Tcr-pble* vector

(B)-TIR-UTR 5'Tcr-pbleII : (B)-GTTTTTCATTAATATTGAGTGTTGC-3'

(B)-TIR-UTR 3'Tcr-pbleII : (B)-GCAGCACTTATGAATCATCTATC-3'

Second round of nested PCR

Primer anchored in linker

LC2 : 5'-GATCTGAATTCAGTGGCACAG-3'

Primers anchored in the 5' or 3' end of the *IFP2* vector

TIR-UTR PB 5'III : 5'-CGATAAAACACATGCGTCAAT -3'

TIR-UTR PB 3'III : 5'-GCTATTTAGAAAGAGAGAGCA-3'

Primers anchored in the 5' or 3' end of the *Tcr-pble* vector

TIR-UTR 3' Tcr-pble III : 5'-CACCTTTTATTTAATTATTTGAGATAC-3'

TIR-UTR 5'Trc-pble III : 5'-CAATGAAACTTATTTCTAAATAC-3'

>GFPT01020379.1

ttaaccctttgaagcacacattaaggatttaattttatTTTTTaaTTTTTacaatgaaaacttatttct
aaatacatttaaaatTTAATTTGAGCAAAAATAA--TTTTT-
cattaatattgagtggtgctaaaaaatggtggtcattcagtgaaaccactatgcaactaagcttact
atatgtttacatggtggtatgcttcaaagcaatTTTTtagcttcattgagacacaatgccacttcgc
atgtcctacaagaccatttacttcggtgtggctgtaacttc-
gtgctgcagtgggcacaacgtctgcttgaactcc----
ttacaggtaaagtagccaccttgtcaaactTTTTTcaagaggtacttgaggcttgaatatagctcact
ggctttggaattggttttcttctcttaggtgaagattggtgtgcaccaatcaaaccatctactacaga
aatacgaaatTTCTTTAGTGATAAATGGGTGCCCTCCTTTGGAGTGATAAATGACAAAGCCATTG
TCAGTGCAACATCCACAAAGTGCCAAAATATACGGTGCCACCCTTGCGG--ATTTT----
cggcttatttcgtatagggatt-
caacatatcagctttatccacatagcccatgtgctgattgtaatctttcactagttgtggacaggtta
ctatttgtattgttccatctttgttttttcttggcttcagaaataagcgaaggattgtggaaattg
gacaaaaatgtaatacatttgttgtctttccacttttagccaagcaactcctgtaaattgaggtgcgaa
ctcactatcacctttcttcatgttcttttcta---
tcttttgttcttttggaagccagttctcacattccttactgttccgcatgcatatatcgcttggttc
tttagtgaaatcattagatcaacacttgtaaataattgtcgaaataaacatgatgaccaacccaac
aagttcaagagtaaggctttgacaactcgggcacctaataagcttctcaatattgccttcaaccttac
ctgtgtatatttgaactgacagacaaacccgactcatccgcacgcacccaaacctttagcctcgt
tttattggcttgttgggcataaattgtctgaaagaactcctcctttaaatttgatcatactttcatc
gatggactgatgattccctggttgaagcaggttttgaagctctgaaagtgtatcaagcattggct
ttagttttagagcttatcaaagcattgtcgtctctttttggctgcaatacattgttgtttacatgt
aaatgcgtcaacaaccaacaaaacgattcactggcattaggggagagataaatgaatccctaagctg
agggtaagatgaccaatggtccctgtaggagggtaaacgttttatccccatcataatgtttattgcta
tga---
aaatcttcatttcttcttcagttacagggacgaatggtttccccaactgtgtagcataaaggttagtt
tgaaaaactaaatgttcaataaaaagaatctggaaataaacacaaaaaagtatcgacaggagattctgt
acttgttggtagcatttagggccaaagtcaatttgaataatttgaatgttccaccgccttattttgttt
cccacacaccatcta--atTTTTTaaatcgctcactaaggggttcatcatcatcatcatc-----
tgtattacaatcttctgatttatatcagaataatttaggacattattgttattattgaaaccgttat
cttctgtataaaaacaattgatctgtcctcactttcctgtaagcaatcttctaataatccgaatcacat
tc----
tccatcagaatcacactgtccatcaatagagtcagtatcactaggcacttctatttagcaaacgttcaa
catcttctctttgagtgatagcttttagcaataattttcttttat-----
ttatagacatgacgacactgaaatgagtaaaaatatacacatgaccttacttgcacagtggtagcaat
ttgaaaccacctagtaattag---g--
ttattaaattaatataatttagataaaatTTAATTATGAATGTAATAATTATAAAGCAAAAAATAATA
GTATTAATAATGATTGAATACT-----
tacatagttatataatttattccttgaaaaatataacaaaaccaattagaataaacgtgtcacttcaat
aacaaaaacaaaatacatactatattggaggcatgaataccaacaaaaattactaataataaatttt
acaagaataaaaactagcaataaattcagatagatgattcataagtgtgcca-
acagttaaatgtgtctcaataattaaataaaaaggtggttaccaagtgaaccactatgcttcaaagg
gtaa

>MSSY03005913.1b

ttaaccctttgaagcacacattaaggatttaattttatTTTTTaaTTTTTacaatgaaaacttatttct
aaatacatttaaaatTTAATTTGAGCAAAAATAA--TTTTT-
cattaatattgagtggtgctaaaaaatggtggtcattc-----

-----agtgaaaccactatgcttcaaagggttaa
>MSSY03005238.1b
ttaaccctttgaagcacacattaaggatttaattttatttttaattttttacaatgaaaacttatttct
aaatacatttaaaatttaatttgagcaaaaataa--tgttt-
cattaatattgagtgttgctaaaaaatggtggtcattcagtgaaaccac-----

aacaaactacatactatattatagggcatgaataccaacaaaaattactaataataaatttttacttga
ataaaaactagcaataagttatgatagatgattaataagtgtcgca-
acagttaaatgtattttcaataatgaaaggtggttactaggggaaaccactatgcttcaa-
ggtttaa

>PGTA01000295.1

TTAACCCTTTTGAAGCACACATTAAGGATTTAATTAATTTTTAATTTTTTACAATGAAAACCTTATTTCT
AAATACATCTAAAATTTAATTTGAGCAAAAATAA--TGTTT-
CATTAATATTGAGTGTTGCTAAAAAATGGT----
CATTCAAGTAAAAGCACTATGCAACTAAGCTTCACTATATGTTTACATGGTGTGGTATGCTACAAATCA
ATTTTTAGATTCATTGAGACACAATGCCACTTCGCATGTCTTACAAGACCATTGCTTCGGTGTGGCT
GTAACCTCT-TGCTGCAGTGGGCACAACGTCTGCTTGAACCTCC----
TTACAGGTAATGAGCCACCTTGTCAAATCTTTTTCAATAGGTACTTGAGGCTTAAAATAGCTCACT
GGCTTTGGAATTGGTTTTCTTCTCTAGGTGAAGATTGTGTTGCACCAATCAAACCATCTACTACAGA
AATACGAAATTTCTTTAGTGATAAATGGGTGCCCTCACTTTGGAGTGCATAAATGACAAAAGCCATTG
TCAGTGCACATCCACAAAAGTGCCAAAATATACGGTGCCACCACCTTGCGGC--
ATTTCTTCCGGCTTATTTCTGTATAGGGATTTTCAACATATCAGCTTTATCCACATAGCCCATGTGCT
GGTTGTAATCTTTCAGTAGTGTGGACAGGTTACTATTTGTATTGTTCCATCTTGTTTTTTCTCTTG
GCTTCAGAAATAAGCGAAGGATTGTGGAAACTGGAGAAAAATGTAATACATTTGTTGCTTTCCACTT
TAGCCAAGCAACTCCTGTAATGATGTGCGAAACTCACTATCACCTTTCTTCATGTTCTTTTCTACC-
--

TTTTGTTCTTTTGGCAAGCCAGTTATCACATTCCTTACTGTGCCGCATGCATATATCGCTTGGTTCTT
TAGTGAAATCATTAGATCAACACTTGTAATAAATTGTCGAAATAAACATGATGACCCAACCCAAAA
GTCCAAAAGTAAAGTCTTTGACAACCTTGAACACTTCAATAGCTTCTCAATATTGCCTTCAACCTTACCT
GTGTATATTTGAAACTGACAGATAAACCCTGACTCATCCGCACGCACCCACACCTTGTAGCCTCGTTT
TATTGGCTTGTGGCATAAATTGTCTGAAAGAACTCCTTCTTTAAATTTGATCATACTTTCATCAA
TGGACTGATGTTTTCTGGTTTGAAGCAGGTTTTGAAAGTCTCTGAAAGTGATCAAGCATTGGTCTT
AGTTTGTAGAGCTTATCAAAGCATTGTGCTCTTTTTGGCTGCAATACATTGTTGTTTACATGTAA
ATGCGTCAACAACCAACCAAAAACGATTCACTGCCATTAGGGGAGAGATAAATGAATCCCTAAGCTGAG
GGTAAGATGACCAATGGTACCTGTAGGAGGGTAAACGTTTTATTCCCATCATTATGTTTATTGCTATG
A---

AAATTTTCATTTCTTCTTCAGTTACA-----

>PGTA01003554

TTAACCCTTTTGAAGCACACATTAAGGATTTAATTTATTTTTATTTTTTTGCTATGAAAACCTTATTTCT
AAATACATTTCAAATTTAATTTTAGCAAAAAATAAAGTTTTTCATTGAGATTGAGTGGTCTAAAAA
TGGTGGTCTTTAGTAAAACCACTATTCAAATAAATTCAGTATATTTTTACATGGTGTGGTATGCTA
CAAAGCAATTTTATATTCATTGAGACACAATGCCACTTCGCATGTCTTACAAGACCATTTACTTCGG
TGTGGCTGTAACCTCT-TGCTGCAGTGGGCACAACGTCTGCTTGAACCTCC----
TTACAGGTAATGAGCCACCTTGTCAAATCTTTTTTCAAGAGGTACTTGAGGCTTAAAATAGCTCACT
GGCTTTGGAAGTGGTTTTTCTTACTCCAGGTGAAGATTGTGTTGCACCAATCAATCCATCTACTACAGA
AAT--GA-----GA--

AATGGGAGCCCTCACTTTAGAGTGCATAAATTACAAGGCCATTTGACAGTGCAACACCCACAAAGTGT
CAGAATATACGATGCGG-CATTTCCGGCTTATTTCC-----
AACATATCAGCTTTATCCACATAGCCCATGACTGGTTGTAATCTTTCACTAGTTGTGGA-----
CC---
TGTATTGTTCCATCTTCGTTTTTCTCTTGACTTCAGAAATAAGCGAGGGATTGTGGGAATT-----

ACATTTGTTGTCTTTCCACTTTAGCCAAGCAACTCCCGTAAATTAAGTGTGAAACTCACTATCACCTT
TCTTCGTGTTCTTTTCTACC---
TTTTGTTCTTTTGGCAAGCTGGTTCTCACATTCATCACTGTGCCGCATGCATGTATCGATTGGTTCTT
TAGTGAAATAATTGAATCAACACTTATAAAAATAGTTGTC---
ATAAACATCATGACCCAAACCAAAAAGTTCAAAGTTAGACCTTTGGCAACTCGAGCACCTAATAGTT
TCTCAATATTGCCCTCAACCTTCCCTATGTATATCTGAAACTGAC-----
TTACCCGCACGCACCCACACCTTGTAGCCTCGTTTTATTGGATTGTTGGGCATAGATTGTCTGAAAGA
ACTCCTTCCTTTAAATATGATCATACTTCCATCAATGGACTGATGTTTTCTGGTTTGAAGCAGGTTA
TGAAAGTCACTGAAAGTGTATCAAGAATTGGTCGTAGTTTGTAGAGTTTGTCAAATGCATTGTCTCCT
CTTTTTAGCT-
TAATACTTTGTTGTTTACATGTAAATGTGTCAACAACCAACCAAAAACGATTCACTGGCATCAGGGGAG
AGATAAATGTATCCCTAAGCTGAGGTTGATATGACCAATGGTCCCTTTAGGAGGGTAAACGTTTTATC
CCCATCATAATGTTTATTATTATGA---
AAATTTTTCATTTATTCTGCAGTTACAAGGATGAATAGTTTCCCCATCTGTGTAGAGTAAAGGTAAGTT
TGAAAACTAAAGGTTCACTGAAAGAATCTGGATATAAACACAAAAAAGTATTGACAGGAGATTCTGT
ACTTGTGGTACATTAGGGCCAAAGTCATTTGTCAAATTTTGAATGTTCCACCACATTATTTTGT
CCCACAAACCATATA-TTTTTTAAATCCCTCACTAAGTGGTTCATCATCATCATCAT-----
ATTTATTACTATCTTCTGACTTACATCAGAATAATTTAGGACATTATTGTCATTATTGAACCCGTCA
TCTTCTGTATAAAAAACACTTGATCTGTCCTCACTTTCTGTAAGCAATCTTCTCCTTA-
GAATCACACTCACCTTCC-----
ATAGAGTCAGTATCCCTAGGCACTTCTACTAGCAAACGTTCAACATCTTCTCTTTGAGTGATAGACT
AAGTTTAGCAATCATTTTATTTTA-----TTTATAGACATGAC---
ACTGAAATGATTACAAATATGCACACAGCTTAACTCGCACAAATGGTAGCAATTTGAAGCCACCTAGTA
ATTAG-----TTTATTTAAATATAT-TATATAT-
TATATATATATTT-----

>PGFK01000005.1

TTAACCTTTTGAAGCACACATTAAGGATTTAATTTATTTTTATTTTTTGTCTATGAAAACCTTATTTCT
AAATACATTTCAAATTTAATTTTAGCAAAAAAATAAAGTTTTTCATTGAGATTGAGTGTGCTAAAAA
TTGTGGTCTTTCAGTAAAACCACTATTCAAATAAATTTCAATATATTTTTACATGGTGTGGTATGCTA
CAAAGCAATTTTATATTGAGACACAATGCCACTTCGCATGTCTTACAAGACCATTTACTTCGG
TGTGGCTGTAACCTTCT-
TGCTGCAGTGGGCACAACGCTGCTTGAACCTCCCTCCTTACAGGTAATGAGCTACCTTGCCAATCT
TTTTTCAAGAGGTACTTGAGGCTTGAATAGCTCACTGGCTTTGGAAGTGGTTTTCTTACTCCAGGTG
AAGATTGTGTTGCACCAATCAATCCATCTACTACAGAAATGAGAAAT-----
GGGAGCCCTCACTTTAGAGTGCATAAATTACAAGGCCATTTGACAGTGCAACACCCACAAAGTGTGAG
AATATACGATGCGG-CATTTCCGGCTTATTTCC-----
AACATATCAGCTTTATCCACATAGCCCATGACTGGTTGTAATCTTTCACTAGTTGTGGA-----
C--
CTGTATTGTTCCATCTTCGTTTTTCTCTTGACTTCAGAAATAAGCGAGGGATTGTGGGAATT-----

ACATTTGTTGTCTTTCCACTTTAGCCAAGCAACTCCCGTAAATTAAGTGTGAAACTCACTATCACCTT
TCTTCGTGTTCTTTTCTACC---
TTTTGTTCTTTTGGCAAGCTGGTTCTCACATTCCTCACTGTGCCGCATGCATGTATCGATTGGTTCTT
TAGTGAAATAATTGAATCAACACTTATAAAAATAGTTGTC---

ATAAACATCATGACCCAAACCAAAAAGTTCAAAGTTAGACCTTTGGCAACTCGAGCACCTAATAGTT
TCTCAATATTGTCCTCAACCTCCCTATGTATATCTGAAACTGAC-----
TTACCCGCACGCACCCACACCTTGTAGCCTCGTTTTATTGGATTGTTGGGCATAGATTGTCTGAAAGA
ACTCCTTCCTTTAAATATGATCATACTCCATCAATGGACTGATGTTTTCTGGTTTGAAGCAGGTTA
TGAAAGTCACTGAAAGTGTATCAAGAATTGGTCGTAGTTTGTAGAGTTTGTCAAATGCATTGTCTCCT
CTTTTTAGCT-
TAATACTTTGTTGTTTACATGTAAATGTGTCAACAACCAACCAAAAACGATTCACTGGCATCAGGGGAG
AGATAAATGTATCCCTAAGCTGAGGTTGATATGACCAATGGCCCTTTAGGAGGGTAAACGTTTTATC
CCCATCATAATGTTTATTATTATGA----
AAATTTTCATTTATTCTGCAGTTACAAGGATGAATAGTTTCCCATCTGTGTAGAGTAAAGGTAAGTT
TGAAAACTAAATGTTCACTGAAAGAATCTGGATATAAACACAAAAAAGTATTGACAGGAGATTCTGT
ACTTGTGGTACATTAGGGCCAAAGTCATTTGTCAAATTTTGAATGTTCCACCACATTATTTGTTTT
CCCAAAAACCATATATATTTTTTAAATCCCTACTAAGTGGTTCATAATCATCATAATCATCATA
TTTATTACTATCTTCTGACTTACATCAGAATAATTTAGGACATTATTGTCATTATTGAACCCGTCAT
CTTCTGTATAAAAACACTTGATCTGTCTCCTCACTTTCCTGTAAGCAATCTTCTCCTA-
GAATCACACTCACCTTCC-----
ATAGAGTCAGTATCCCTAGGCACTTCAACTAGCAAACGTTCAACATCTTCTCTTTGAGTGATAAGTT
TAG-----CAATCATTTTATTTTA-----TTTATAGACATGAC----
ACTGAAATGATTACAAATATGCACACAGCTTAACTCGACAATGGTAGCAATTTGAAGCCACCTAGTA
ATTAG-----TTTATTTAAATATAT-TATATAT-
TATATA-----

>MSSY03008721

TCAGCCCTTTGACGCACACATTAAGGA--
TAATTTATTTTTATCTTTTTACAATTAACCTTATTTCTAAATACATTTAAAATTTTATTTTAGCAA
--TAATATGTTT-CATTAAGATCGAGTGTGCTAAAAAATGGTG-----
AAACCACTATGCAACTAAGCTCCACTATATTTTTACATGGTGTGGTATGCTACAAAGCAATTTTTAGA
NTCGTTGAGAAACAATGCAACTTCGCATTTCTTACAAGACCATTTACTTCGGTGTGGCTGTAACCTC-
GTGCTGCAGTGGGCACAACGTCTGCTTGAACCTC-----
TTACAGGTAATGAGCTACTTTTCAAATCTTTTTTCAAGAGGTAICTCGAGACTGAAAATAGCTCACT
GGCTTTGGAATTGGTTTTCTTCTCTAGGTGAAGTTTTGTTGGACTAATCAATTCATCTACTACAGA
AATACGAAATTCCTTTAGTGATAAATGGGAGCCCTCACTTTAGAGTGCATAAATGACAAAACCATTTG
TCAGTGAACATCCACAAAGTGCCAAAATATACGGTGCCATCATTTGCGGCTTATTTT-----
-----GTAAAGGGATTTT-
AACATATCAGCTTTATCCACATAGCCCATGTGCTGGTTATAATCTTTTATTAGTTATGGACAGACTAC
CATCTGTATTATCCATCTTTGTTTTTTCTCTTGCTTCATAAATAAGCGAAGGATTGCGGAAATTGG
CCAAAAATGTAATACATTTGTTATCTTTCCACTTTAGCCAAGCAACTCCTGTAATGAGGTGCGAAAC
TCACTATCACCTTTCTTCATGTTCTTTTCTACCTTTTTTTTTTTTTTTTTGGCAAGCCAGTTCTCACATT
CCTTACTGTTCCGCATGCATATATCGCTTGTTCTTTAGTGAAATCATTAGCTCAACACTTG-----

CTATGATTTTTTTTTTTTCAATTTCTTCTTCAATTACAGGGACGAATGGTTTTCCACAACCTGTGTAGCATAA
AGGTTAGTTTGAAAACTAAATGTTCAATTAAGAATCTGGAAATAAACACAAAAAATTATCAACAGG
AGATTATTTACTTGTGGTACAATAGGGCCAAAGTAATTAGAAAAATGTTGAATGATCCACCACCTTA
TTTTGTTTTCCACACACTATCTA-

>MSSY03005913

TTAACCCCTTTGAAGCACACATTAAGGATTTAATTTATTTTTAATTTTTTACAATGAAAACCTTATTTCT
AAATACATTTAAAATTTAATTTGAGCAAAAATAA--TTTTTC-
ATTAATATTGAGTGTTGC-----

-----TAAA-
AAATGGTGGTCATTCAGTGAAACCACTATGCTTCAAAGGGTTAA

>Consensus60

TTAACCCCTTTGAAGCACACATTAAGGATTTAATTTATTTTTAATTTTTTACAATGAAAACCTTATTTCT
AAATACATTTAAAATTTAATTTGAGCAAAAATAAAATGTTTTCATTAATATTGAGTGTTGCTAAAAAA

TGGTGGTCATT CAGTGAAACCACTATGCAACTAAGCTNCACTATATTTTTACATGGTGTGGTATGCTA
CAAAGCAATTTTTAGATTCATTGAGACACAATGCCACTTCGCATGTCTTACAAGACCATTTACTTCGG
TGTGGCTGTAACCTCTGTGCTGCAGTGGGCACAACGTCTGCTTGAACCTCCCTCCTTACAGGTAAATGA
GCNACCTTGTCAAATCTTTTTCAAGAGGTACTTGAGGCTTGAATAGCTCACTGGCTTTGGAANTGG
TTTTCTTCTCTAGGTGAAGATTGTGTTGCACCAATCAATCCATCTACTACAGAAATACGAAATTNCT
TTAGTGATAAATGGGAGCCCTCACTTTAGAGTGCATAAATGACAAAGCCATTTGTCAGTGC AACATCC
ACAAAGTGCCAAAATATACGGTGCCACCATTTGCGGCTTATTTCTTCCGGCTTATTTCTGTANAGGGA
TTTTCAACATATCAGCTTTATCCACATAGCCCATGTGCTGGTTGTAATCTTTC ACTAGTTGTGGACAG
NNTACCATCTGTATTGTTCCATCTTTGTTTTCTCTTGGCTTCAGAAATAAGCGAAGGATTGTGGAA
ATTGGNCAAAAATGTAATACATTTGTTGTCTTCCACTTTAGCCAAGCAACTCCTGTAAATGANGTGN
GAACTCACTATCACCTTTCTCATGTTCTTTCTACCTTTTTTTGTTCTTTTGGCAAGCCAGTTCTC
ACATTCCTTACTGTGCCGCATGCATATATCGCTTGGTTCTTTAGTGAAATCATTAGATCAACACTTGT
AAAATANTTGTGCAAAATAACATNATGACCCAANCCAAAAGTTCAAAGTNAGNNCTTTGNCAACTC
GAGCACCTAATAGNTTCTCAATATTGCCNTCAACCTTNCCTNTGTATATNTGAACTGACAGANAAC
CCCGACTNANCCGCACGCACCCACACCTTGTAGCCTCGTTTTATTGGNTTGTGGGCATANATTGTCT
GAAAGAACTCCTTCTTTAAATNTGATCATACTTNCATCAATGGACTGATGTTTTCTGGTTTGAAGC
AGGTTNTGAAAGTCNCTGAAAGTGTATCAAGNATTGGTCNTAGTTTGTAGAGNTTNTCAAANGCATTG
TCNNTCTTTTTNGCTGNAATACNTTGTGTTTACATGTAATGNGTCAACAACCAACCAAAAACGATT
CACTGGCATNAGGGGAGAGATAAATGNATCCCTAAGCTGAGGNTNANATGACCAATGGTCCCTNTAGG
AGGGTAAACGTTTTATCCCCATCATAATGTTTATTNCTATGATTTAAATTTTCATTTCTTCTTCAGTT
ACAGGGACGAATGGTTTCCCAACTGTGTAGCGTAAAGGTTAGTTTGAAAACTAAATGTTCAATNAA
AGAATCTGGAAATAAACACAAAAAAGTATCGACAGGAGATTCTGTACTTGTGGTACATTAGGGCCAA
AGTCATTTGTAAAATTTTGAATGTTCCACCACCTTATTTTGTTTTCCACACACCATCTATTTTTTTT
AAATCGCTCACTAAGTGTTTCATCATCATCATCATCATCATATTTTATTACNATCTTCTGATTTA
TATCAGAATAATTTAGGACATTATTGTTATTATTGAAACCGTTATCTTCTGTATAAAAAACANTTGAT
CTGTCCTCACTTTCTGTAAAGCAATCTTCTAAATCCGAATCACNTTACCTTCCNTCAGAATCACACT
GTCCATCAATAGAGTCAGTATCACTAGGCACCTTCTANTAGCAAACGTTCAACATCTTCTCTTTGAGT
NATACGTTTAGCAATTATTTATTTTATTTTATTTTATTTATAGACATGACGACACTGAAATGAGTAA
AAATATACACATGACCTTACTTGCACAGTGGTAGCAATTTGAAACCACCTAGTAATTAGTTAGATTTA
TTTAAATATATATATTTAGATAAATTTAATTATGAATGTAATAATTATAAAGCAAAAAACAATAGTA
TTAAATAATGATTTGAATACTTAGGGCCTACATAGTTATATATTTATTCCTTGAAAAATATAACAAAA
CCAATTAGAAGAAACGTGTCACCTTCAATAACAAAAACAAAATACATACTATATTGGAGGCATGAATAC
CAACAAAAATTACTAATAATAAATTTTTACAAGAATAAAACTAGCAAATAAATTCAGATAGATGATTC
ATAAGTGCTGCCATACAGTTAAATGTATCTCAATAATTAATAAAAAGGTGGTTACCAAGTGAAACCA
CTATGCTTCAAAGGGTTAA