Supplementary figure legends

Supplementary Figure 1

Alignment of the nucleotide sequences of MITE700 elements identified in the ATCC 33277 genome. Asterisks indicate the nucleotides conserved in all the MITE700 elements.

Supplementary Figure 2

Alignment of the nucleotide sequences of MITE*PgRS* elements identified in the ATCC 33277 genome. Terminal inverted repeats (IRs), Repeat A, Repeat B, and Repeat C regions are indicated. A unique nucleotide sequence detected in MITE*PgRS*_08 is underlined.

Supplementary Figure 3

The *rrn* operons on *P. gingivalis* chromosomes. (A) Locations of the *rrn* operons (arrow heads) and the PCR primers used to examine each *rrn*-flanking region (arrows) on the ATCC 33277 and W83 chromosomes are shown. The *dnaA* gene is located at the top of each chromosome circle. (B) Structures of *rrn* operon-flanking regions of ATCC 33277, W83, TDC60, TDC117, TDC275, SU63 and GAI7802 were examined by PCR using a set of orientation-specific primer pairs indicated in panel (A) (primer sequences are available in Supplementary Table 1). PCR products were separated on a 0.8% agarose gel and visualized by ethidium bromide staining.

Supplementary Figure 4

Comparison of the loci for capsular polysaccharide synthesis of strains ATCC 33277, W83 and 381. All the CDSs on the ATCC 33277 and W83 chromosomes are drawn to scale. Strain 381 contains a nearly identical capsular polysaccharide synthesis locus as that of ATCC 33277 except that a CDS corresponding to PGN_0223 has been truncated by a premature stop codon. Amino-acid sequence identities (%) are also indicated between homologous CDSs.

Supplementary Figure 5

Comparison of the nucleotide sequences of the *glk* genes in seven *P. gingivalis* strains. (A) Schematic presentation of the *glk* genes of seven *P. gingivalis* strains. (B) Nucleotides conserved in all the seven strains are indicated by asterisks below the alignment. Bold nucleotides indicate silent substitutions. Nucleotide substitutions which cause amino acid substitutions or nonsense mutations are highlighted. Locations of the ribosome-binding sequence (SD) and start and stop codons of *glk* are also indicated. MITE239 inserted into the the *glk* gene in strain ATCC 33277 and the CDSs located just upstream and downstream of the *glk* gene are indicated by lines, the terminal inverted repeats (IRs) of MITE239 by boxes, and the target site duplication generated by the MITE239 insertion by double lines.

Supplementary Figure 1

```
                  IR 12 bp                                 variable region
            1 *   *  ****                                                                               ** * ***    *******   * ****** 120
MITE700_01    ACGTCA-TTCGACGTAAAAAAATGCCAATGTATTCGTTCCCTGATTAGCGGTTTACACACTAAGATCTTGGGAGTTATAGGTTGTTCCGTTCCGGCAATTGCATAGGGGGCAATCCTGCT
MITE700_05    ACGTCAGTTCGGCGTAAGAAAACGCCAATGTATTCGTTCCCTGATTAGTGGTTAACGCACTAACATTTTGTATGTTATAGGTTGTACCGTTCAGGCAATTGTATAGGGGGCAATCCTGCT
MITE700_09    ACGTCAGTTCGGCGTAAGAAAACGCCAATGTATTCGTTCCCTGATTAGCGGTTAACGCACTAATATTTTGGATGTTATAGGTTGTTCCGTTCAGGCAATTGTATAGGGGGCAATCCTGCT
MITE700_02    ACCCCTGTTCGGCGTAAGAAAACGCCAATGTATTCGTTCCCTGATTAGCGGGTTACACACTAAGATCTTGGGAGTTATAGGTTGTTCCGTTCAGGCAATTGTATAGGGGGCAATCCTGCT
MITE700_10    ACGTCAGTTCGGCGTAAGAAAACGCCAATGTATTCATTCCCTGATTAGCGGTTAACGCACTAATATTTTGGATGTTATAGGTTGTTCCGTTCAGGCAATTGTATAGGGGGCAATCCTGCT
MITE700_04    ACGTCAGTTCGACGTAAGAAAACGCCAATGAATAATTCTTTTCCTGATCAGAGATTGATGCCCTAAGATTC-------------------TCCGGCAAATGTATAGGGGG-TATCCTGCT
MITE700_12    ACGTCAGTTCGACGTAAGAAAACGCCAATGAATAATTCTTTTCCTGATCAGAGGCTAATGCCCTAAAATTC-------------------TCAGACAA----ATAGGGG--TACCCTGCT
MITE700_13    ACGTCTGTTCGACGTA--------------------------------------------------------------------------------------------------------
MITE700_14    ------------------------------------------------------------------------------------------------------------------------
MITE700_08    ------------------------------------------------------------------------------------------------------------------------
MITE700_07    ------------------------------------------------------------------------------------------------------------------------
MITE700_03    ATGTCAGTTCGATCTAAGCGCAAATAGGGAAGTTGGGTCTGGGTCGAACTGACGACTAA-------------------------------------------------------------
MITE700_11    ------------------------------------------------------------------------------------------------------------------------
MITE700_06    ------------------------------------------------------------------------------------------------------------------------

          121 * ***  ****** ************ ********* ****  *** ******* ******* ****  *********** *** * ** *  ***    **  ** *** ****   *  240
MITE700_01    TCTGCCCGGATCGGACCAACATTGAATTTCGGATTG-TTTTTTGACTGAATGTTGCAGTCCGTACAGGAAGATGGGCTATACCACAGGGCA-CATGGCACTACTCCGTAGAGGATGACGG
MITE700_05    TTTGCTTGGATCGAACCAACATTGAAATTCGGATTGCTTTT--GACAGAATGTTGCAGTCCGCACAGGAAGATGGGCTATACCATAGGGCAAGATGGCGGTACTCCGTAGAGGATGACGG
MITE700_09    TTTGCTTGGATCGAACCAACATTGAAATTCGGATTGCTTTT--GACAGAATGTTGCAGTCCGCACAGGAAGATGGGCTATACCATAGGGCAAGATGGCGGTACTCCGTAGAGGATGACGG
MITE700_02    TCTGCCTGGATCGAACCAACATTGAAATTCGGATTGCTTTTTTGACCGAATGTTGCAGTCCGTACAGGAAGATGGGCTATACCACAAGGCA-GATGGCGGTACACCGTAGAGGATAACGG
MITE700_10    TCTGCTTGGATCGGACCAACATTGAAATTCGGATTGCTTTT--GACCGAATGTTGCAGTCCGCACAGGGAGATGGGCTATACCACAAGGCA-GATGGCGGTACACCGTAGAGGATAACGG
MITE700_04    TCTGCTTGGATCGAACCAACATTGAAATTCGGATTGCTTTT--GACAGAATGTTGCAGTCCGCACAGGAAGATGGGCTATACCATAGGGCAAGATGCCGGTACTCCGTAGAGGATGATGG
MITE700_12    TCTGCTTGGATCGAACCAACATTGAAATTCGGATTGCTTTTTTGACCGAATGTTACAGTCCGTACAGAAAGATGGGCTATACCATAGGGCAC-ATGGAGGTACTCCATAGGGGATGACGA
MITE700_13    -------------------------------------------------ATGTTACAGTCCGTACAGAAAGATGGGCTATACCATAGGGCA-GATGGCGGTACACCATAGAGGATGGCGA
MITE700_14    ------------------------------------------------------------------------------------------------------------------------
MITE700_08    ------------------------------------------------------------------------------------------------------------------------
MITE700_07    ------------------------------------------------------------------------------------------------------------------------
MITE700_03    ----------------------------------------------------------------------------CTATGCCACAAGGCA-GATGGCGGTATTCCGTAGAGGATGATGA
MITE700_11    ----------------------------------------------------------TCCGAACAGGAAGATGGGCTATACCACAGGGCA-GATGGTGGTACTCCGTAGAGGATAACGG
MITE700_06    -------------------------------------------------------------------------GGGCTATACCATAGGG-AGCATGGCGGTATGCCGTAGAGGATGACGA

          241  *** ******  ********** ** ******* ****  ****************************** ****************** ******** **    ************** 360
MITE700_01    GCTATCCCGTACAGCTCTTCGAGGTACGCTGTAGAGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGTGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_05    GCTACCCCGTATGGCTCTTCGAGGTACGCTGTAGAGCTCTTCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGGGAGCCGTAGAGGTAGGTGGCCGAGCTGAG
MITE700_09    GCTACCCCGTATGGCTCTTCGAGGTACGCTGTAGAGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGGGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_02    GCTATCCCGTACGGCTCTTCGAGCTACGCTGTAGAGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGTGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_10    GCTATCCCGTACGGCTCTTCGAGCTACGCTGTAGAGCTCTTCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGGGAGCCGTAGAGGTAGGTGGCCGAGCTGAG
MITE700_04    GCTACCCCGTATGGCTCTTCGAGGTACGCTGTAGAGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGTGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_12    GCTATCCCGTACAGCTCTTCGAGCTACGCTGTAG--------------------------------------------------------------------------------------
MITE700_13    GCTATCCCGTACGGCTCTTCGAGCTACGCTGTAGAGCTCAGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGTGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_14    ------------------------------------------------------------------------------------------------------------------------
MITE700_08    ------------------------------------------------------------------------------------------------------------------------
MITE700_07    ------------------------------------------------------------------------------------------------------------------------
MITE700_03    GCTATCCC----------------------------------------------------------------------------------------------------------------
MITE700_11    ACTATCCCGTACAGCTCTTCGAGCTATGCTGTAGAGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAGCCGTAGAGTTCCGCTCCGGGAGCCGTAGAG----GTGGCCGAGCTGAG
MITE700_06    GCTATCCCGTACGGCTCTTCGAGCTACGCTGTAGGGCTCGGCGAGGTACGCTCTACAGCGTAGCGAGCCGAACCGTAGAGTTCCGCTCCGGGAGCCGTACAG----GTGGCCGAGCTGAG

          361 ** *** **  ***** * *************** ******  ** ******   *****    ************ **************************** *** *** * * ** 480
MITE700_01    CCGTAGAGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAGGGCCGTAGAGCGCAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCGGCGT
MITE700_05    CCCTAGAGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAGAGCCGTAGAGCACAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCGGCGT
MITE700_09    CCGTAGAGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAGAGCCGTAGAGCGCAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCTGCGT
MITE700_02    CCGTAGAGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAGAGCCGTAGAGCGCAGCTCAGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCGGAGT
MITE700_10    CCGTAGAGCCCGGCGCAGCGAGCCCTACAGCGTAGCTCGAAGAGCTGTAGAGCGCAGCTCGGTACGCTCTACGGCGCAGCCCGAGTCGCTGTAGGGCGTAGCTCGCCCCCCTT-CGGAGT
MITE700_04    CCGTAGAGCCCGGCGCAGCGAGCCCTACAGCGTACCTCGAAGAGCCGTAGAGCGCAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCGGAGT
MITE700_12    --------------------------------------------------AGTGCAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCTCTTTCGGAGT
MITE700_13    CCGTAGAGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAGAGCCGTAGAGCACAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTTCGGCGT
MITE700_14    ------------------------------------------------------------------------------------------------------------------------
MITE700_08    ------------------------------------------------------------------------------------------------------------------------
MITE700_07    ------------------------------------------------------------------------------------------------------------------------
MITE700_03    -------------------------TACAGCGTAACTCGAAGAGCTGTAGAGCGTAGCTCGACCC-------------------------------------------------------
MITE700_11    CCGTAGAGCTTGGCGCAGTGAGCCCTACAGCGTAGCTCGAAGAGCCGTAGAGCGCAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCCTTCCGGAGT
MITE700_06    CCGTAGGGCCCGGCGCGGTGAGCCCTACAGCGTAGCTCGAAAAGCCGTAGAGCGTAGCTCGGTACGCTCTACGGCGTAGCCCGAGTCGCTGTAGGGCGTAGCTCGACCCCC---------

          481 ** ******* ************** ******* ***  *******  **************** ***  *********** ** ***** *  ****  ***** *    * **** ** 600
MITE700_01    ATGTTCCGAACGGCTATAAAACAATGCGGATTCGATCC-AAGAAAAACATGCAATATGCTGTTTGTCAATGTATTTCATTTGTCTGTTGAGAGGTTTCAGCTTCGAGGGGCTCCGACGTC
MITE700_05    ATGTTCCGAACGGCTATAAAACAATGCGGATTCGATCTGAAGAAAA-TATGCAATATGCTGTTTGTCAGCGTATTTCATTTCTCGGTTGAGAAATTTCAGCTTCGAGAGGCTTCGACATC
MITE700_09    ATGTTCCGAATGGCTATAAAACAATGCGGATTCGATCTGAAGAAAAATATGCAATATGCTGTTTGTCAGCGTATTTCATTTCTCGGTTGAGAAATTTCAGCTTCGAGAGGCTTCGACATC
MITE700_02    ATGTTCCGAACGGCTATAAAACAATGCGGATTCAATCC-AAGAAAAATATGCAATATGCTGTTTGTCAGCGTATTTCATTTCTCAGTTGAAAGATTTCGACTTCGAGTGGATCCGACGTC
MITE700_10    ATGTTCCGAACGGCTATAAAACAATGCGGATTCGATCTGAAGAAAAATATGCAATATGCTGTTTGTCAGCGTATTTCATTTCTCGGTTGAGAGATTTCAGCTTCGAGAAGCTCCGACGTC
MITE700_04    ATATTCCGAACGGCTATAAAACAATACGGATTCGATCC-AAGAAAAACATGCAATATGCTGTTTGTCAATGTATTTCATTTGTCTGTTGAGAGGTTTCAGCTTCGAGAGGCTCCGACGTC
MITE700_12    ATGTTCCGAACGGCTATAAAACAATGCGGATTCGATCC-AAGAAAAACATGCAATATGCTGTTTATCAACGTATTTCATTTCTCGGTTGAGAGATTTCAGCTTCGGGAGACTTCGACATC
MITE700_13    ATGTTCCGAACGGCTATAAAACAATGCGGATTCGATCC-AAGAAAAATATGCAATATGCTGTTTGTCAATGTATTTCATTTGTCTGTTGAGAGGTTTCAGCTTCGAGGGGCTCCGACGTC
MITE700_14    ------------------------------------------------------------------------------------------------------------------------
MITE700_08    ------------------------------------------------------------------------------------------------------------------------
MITE700_07    ------------------------------------------------------TATGCTGTTTGTCAACGTATTTCATTTCTCGGTTGAGAGGTTTCAGCTTCGAGAAGCTTCGACATC
MITE700_03    ------------------------------------------------------------------------------------------------------------------------
MITE700_11    ATATTCCGAAAGGCTATAAAACAATGC---------------------------------------------------------------------------------------------
MITE700_06    ------------------------------------------------------------------------------------------------------------------------

                                                                                                                                 IR 12 bp
          601 * ** ********** ******************  ** ** * ***** *    ****  ** *                         **  * * **** ** ****** ** ******** * 727
MITE700_01    CAATGAGACGTCAAACTTTTTCCCTTTCATTTTCTGATTCGGTTATAGTCAATATCCCATCAAGAGTTTGATCGATAAGGCGGATCTTATGAGGATCGACACCATATCGCTTACATCGAACTCACGT
MITE700_05    CAATGAGACGTCAAACTTTTTCCCTTTCATTTTCTGATTCGATTATAGTCAATATCCCAACAAAAG-------------------------AGAGTTGGCACCATATCGCTTACATCGAACTCACGT
MITE700_09    CAATGAGACGTCAAACTTTTTCCCTTTCATTTTCTGATTCGATTATAGTCAATATCCCAACAAGAG-------------------------AGAGTCGGCACCATATCGCTTACATCGAACTCACGT
MITE700_02    CAATGAGACGTCAAACTTTTTCCCTTTCATTTTCTGAT-CGGTTATAGTCAATGCCCCAACAAGAGCTTGATCGATAAGGTGGCTCTTATGAGAATCGACACCATATCGCTTACATCGAACTCACGT
MITE700_10    CAATGAGACGTCAAA----------------------------------------------------------------------------------------------------------------
MITE700_04    CGATGAGACGTCAAATTTTTTCCCTTTCATTTTCTGAT-CGCTTATAGTCA-TATCCCAACAAGAGCTTGATCGATAAGGCGGCTCTCATGAGAATCGACACCATATCGCTTATATCGAACTCACCT
MITE700_12    CGATGAGACGTCAAATTTTTTCCCTTTCATTTTCTGAT-CGGTTATAGTTAA---CCCAACAAAGGATTGATCGACAAA----TACTTATGAGAGTCGGCACCATAACGCTTATATCGAACTCACGT
MITE700_13    CGATGAGACGTCAAACTTTTTCCCTTTCATTTTCTGATTCGGTTATAGTCAATATCCCATCAAGAG-------------------------AGAATCGGCACCATATCGCTTACATTGAACTCACGT
MITE700_14    -----------CAAACTTTTTCCCTTTCATTTTCCAATTCGGTTATAGTCAATATCCCAACAAGTGTTTGATCGATATGGCGGATCTTATGAGAGTCGGCACCATATCGCTTACATCGAACTCACGT
MITE700_08    -----------CAAACTTTTTCCCTTTCATTTTCCAATTCGGTTATAGTCAATATCCCAACAAGTGTTTGATCGATATGGCGGATCTTATGAGAGTCGGCACCATATCGCTTACATCGAACTCACGT
MITE700_07    CGAT-AGACGTCAAACTTTTTCCCTTTCATTTTCTGAT-CGGTCATAGTCAATATCCCAATAAGAGCTTGATCGATAAGGCGGCTCATATGAGAATCGGCACCGTATCGCTTACATCGAACTCACGT
MITE700_03    -------------------------------------------------------------------------------------------------------------------------------
MITE700_11    -------------------------------------------------------------------------------------------------------------------------------
MITE700_06    -------------------------------------------------------------------------------------------------------------------------------
```

supplementary Figure 2

```
                    IR                         Repeat A
                1                                                                                                                     120
MITEPgRS_04     GAGACTGTTGCGGGGGAGTTTCATTGAGCTCTTTTGCTGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTCCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CACCAAGT
MITEPgRS_07     GAGACTGTTGCAGGGGATTTTCATTGAGCTCTTTTGCTGCAGAGCAGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTCCGGTATATGGACACAGCAAGCAAACAGAAATTT-CATCAAGT
MITEPgRS_16     AAGACTGTTGCATGGGAGTTTCATTGAGCTCTTTTGCTGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTCCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CACCAAGT
MITEPgRS_13     -GAACTGTTGCATGGGAGTTTCATTGAGCTCTTTTGCTGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGATCAAACCTCCGGTATATGGACAT--CGAGCAAACAGAAATTT-CCCCAAGT
MITEPgRS_18     GAGACTGTTGCATGGGAGTTTCATTGAACTCTTTTGCTGCAGAGCTGATTCTTAGTGCCTTTGGGAAAGATCAAACCTCCGGTATATGGACAT--CGAGCAAACAGAAATTT-CCCCAAGT
MITEPgRS_14     GAGACTGTTGCATGGGATTTTCATTGAGCTCTTTTGCTGCAGAGCAGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTCCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CACCAAGT
MITEPgRS_12     GAGACTGTTGCGGGGGAGTTTCATTGAGCTCTTTTGCTGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTTCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CCCCAAGT
MITEPgRS_01     GAGATTGTTGCATGGGAGTTTCATTGAGTTCTTTTGCTGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTTCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CCCCAAGT
MITEPgRS_11     GAGACTGTTGCATGGGAGTTTCATTGAACTCTTTTGCTGCAGAGCAGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTCCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CCCCAAGT
MITEPgRS_02     GAGACTGTTGCAGGGGATTTTCATTGAGCTCTTTTGCTGCAGAGCAGATTCTTAGTGTCTTCGGGAAAGGTCAAACCTTCGGTATATGGGCAC--CGAGCAAATAGAAATTT-CCCCAAGT
MITEPgRS_06     GAGACTGTTGCATGGGAGTTTCATTGAACTCTTTTGCCGCAGAGCTGATTCTTAGTGTCTTCGGGAAAGATCAAACCTC-AAGATATGGCTAC--CAAGCAAACAGAAATTT-CCCCAAGT
MITEPgRS_08     GAGACTGTTGCAGGGGATTTTCGTTGAACTCTTTTTCCGCAGAGCTGATTCTTTGTGCCTTCGTGAAAGGTTAAACCTCCAGGACATGAGCAC--CGAGCAAATAGATATTTTCCCCCAGT
MITEPgRS_19     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_10     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_03     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_05     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_09     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_15     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_17     ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_20     ------------------------------------------------------------------------------------------------------------------------

                                                             Repeat B
                121                                                                                                 220
MITEPgRS_04     TTCCATCAGA---------------------------------GAAGTATTCCTTTCCTCGTCAAATCAATGCTTGTGCCGATAGGCGAGAAATAAGGAA
MITEPgRS_07     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAATCAATGCTTGTGCCGATAGGCGAGAAATAAGGAA
MITEPgRS_16     TTCCATCAGA---------------------------------GAAGTACCCCTTTCCTTGTCAAAT----------------AAGTGAGAAATAAGGAA
MITEPgRS_13     TTC-TTTCAGAGGTGGATACTTGAGGTTCGCAGAGACTCACATGAAGCACTCCTTTTTTCGTCAAATCAATGCTTGTGCCGATAGGTGAGAAACAAGAAA
MITEPgRS_18     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_14     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_12     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_01     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_11     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_02     TTCCATTAGA---------------------------------GAAGTACTCCTTTCCTCGTCAAAT----------------AGGCGAGAAATAAGAAA
MITEPgRS_06     TTCCATCAGA------------------------------------------------------------------------------------------
MITEPgRS_08     TTC-TTTCATAGGTGTATACTTGAGGTTTGCAGAGATCCGCATGAAGCATTCCTTTCTTCGTCAAATCAATGCTTGTGTCGATAGGCGAGAAAT------
MITEPgRS_19     ----------------------------------------------------------------------------------------------------
MITEPgRS_10     ----------------------------------------------------------------------------------------------------
MITEPgRS_03     ----------------------------------------------------------------------------------------------------
MITEPgRS_05     ----------------------------------------------------------------------------------------------------
MITEPgRS_09     ----------------------------------------------------------------------------------------------------
MITEPgRS_15     ----------------------------------------------------------------------------------------------------
MITEPgRS_17     ----------------------------------------------------------------------------------------------------
MITEPgRS_20     ----------------------------------------------------------------------------------------------------

                         Repeat C regions
                221                                                                                                                   344
MITEPgRS_04     TGATTGTCAGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_07     TGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_16     TGATTGTAGGCCGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_13     TGATTGTCAGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCGGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_18     CGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCGGCTGATTCTTGCTCCCTGCACGATGCAGGACG CGATTGTCGGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_14     CGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_12     CGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGAC-- TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_01     CGATTGTCGGCTGTTTCTTGCTCCCTGCACGATGCAGGACG- CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG -----------------------------------------
MITEPgRS_11     CGATTGTCGGCTGTTTCTTGCTTCCTGCACGATGCAGGACG- ----------------------------------------- -----------------------------------------
MITEPgRS_02     CGATTGTCAGCCGTTTCTTGCTTCCAGCACGATGCAGGACG- CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_06     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_08     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_19     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_10     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_03     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_05     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_09     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_15     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_17     ------------------------------------------ ----------------------------------------- -----------------------------------------
MITEPgRS_20     ------------------------------------------ ----------------------------------------- -----------------------------------------


                345                                                                                                                  467
MITEPgRS_04     CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_07     TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG
MITEPgRS_16     CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_13     CGATTGTCAGTTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_18     CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGTTGATTCTTGCTTCCCGCACGATGCAGGACG
MITEPgRS_14     CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGTTTCTTGCTTCCCGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_12     CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_01     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_11     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_02     CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCATCGTGCAGGACG -----------------------------------------
MITEPgRS_06     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_08     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_19     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_10     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_03     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_05     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_09     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_15     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_17     ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_20     ----------------------------------------- ----------------------------------------- -----------------------------------------
```

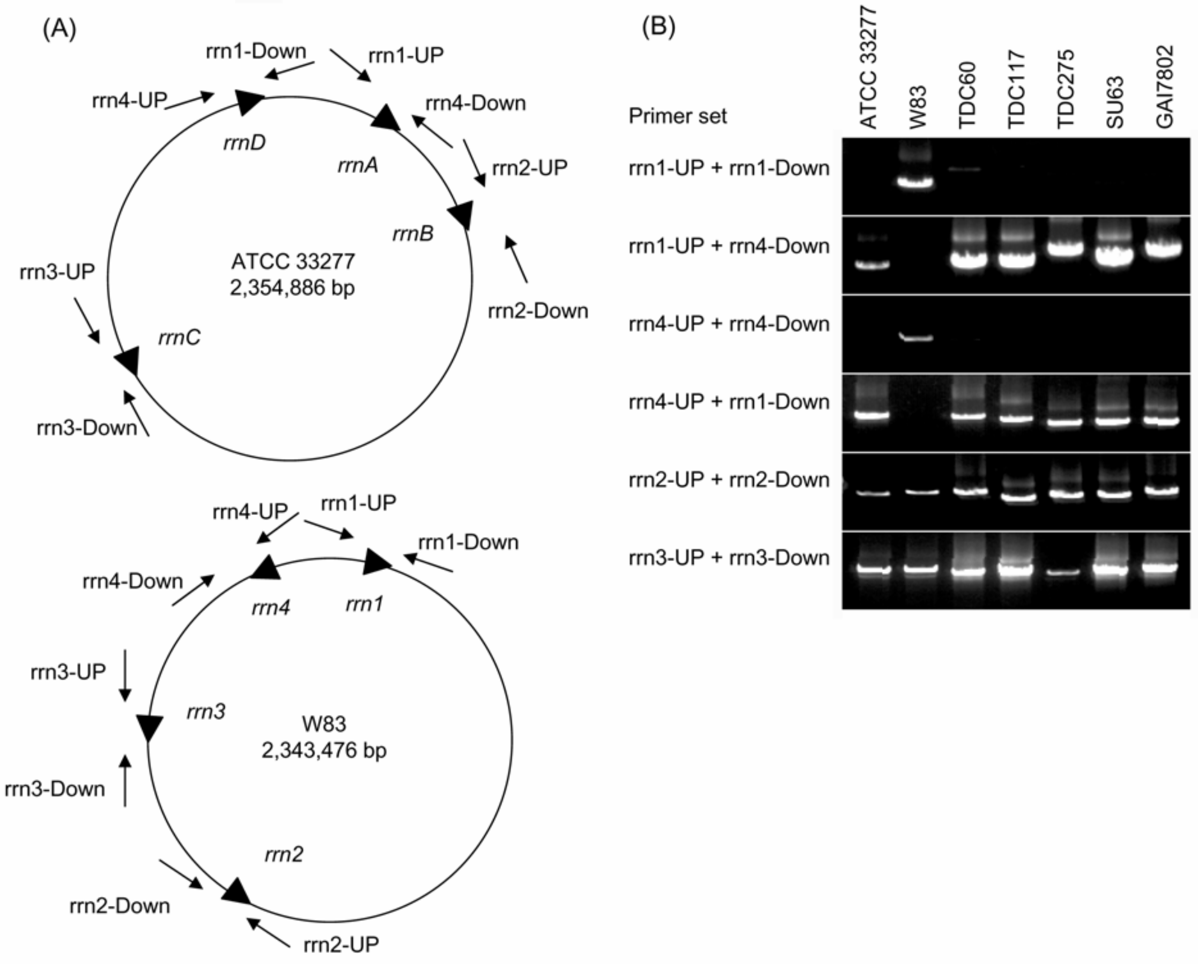
supplementary Figure 2   (continue)

```
             468                                                                                                                     590
MITEPgRS_04  CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGGTGCCGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_07  CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACA CAATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_16  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_13  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_18  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAATTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_14  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_12  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_01  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_11  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_02  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_06  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_08  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_19  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_10  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_03  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_05  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_09  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_15  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_17  ----------------------------------------- ----------------------------------------- -----------------------------------------
MITEPgRS_20  ----------------------------------------- ----------------------------------------- -----------------------------------------


             591                                                                                                                       715
MITEPgRS_04  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAACTGATTCTTGCTTCCTGCAC-GATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_07  CGATTGTCAGCTGATTCTTGCTCCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCAC-GATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG
MITEPgRS_16  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_13  TGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG ------------------------------------------ ------------------------------------------
MITEPgRS_18  CGATTGTCAACTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTTCCTGCACCGATGCAGGACG ------------------------------------------
MITEPgRS_14  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_12  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_01  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_11  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_02  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_06  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_08  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_19  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_10  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_03  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_05  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_09  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_15  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_17  ----------------------------------------- ------------------------------------------ ------------------------------------------
MITEPgRS_20  ----------------------------------------- ------------------------------------------ ------------------------------------------


             716                                                                                                 821
MITEPgRS_04  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_07  CGATTGTCAGCTGATTCTTGCTTCCTGCACGATGCAGGACG CGATTGTCAGCTGATTCTTGCTCCCTGCAC-GATGCAGGACG CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_16  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_13  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_18  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_14  ----------------------------------------- ------------------------------------------ CGATTGTCAACTGATTCTGCTCC
MITEPgRS_12  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_01  ----------------------------------------- ------------------------------------------ CAATTGTCA---------GCTCC
MITEPgRS_11  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_02  ----------------------------------------- ------------------------------------------ CGATTGTCAGCTGATTCTGCTCC
MITEPgRS_06  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_08  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_19  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_10  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_03  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_05  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_09  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_15  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_17  ----------------------------------------- ------------------------------------------ -----------------------
MITEPgRS_20  ----------------------------------------- ------------------------------------------ -----------------------


                                                             Repeat A
             822                                                                                                                 941
MITEPgRS_04  CATCAATGCGCTAACTATCAGCTGTTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_07  CATCAATGCGCTAACTATCAGCTGTTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_16  CATCAATGCGCTAACTATCAGCTGTTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAGTTTTTTTTCAGAT-------------------------
MITEPgRS_13  CATCAATGCGCTAACTATCAGCTGTTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_18  CATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTAAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGGAAAGATCAAACCTCATGCATAAGGGC
MITEPgRS_14  AATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_12  CATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGGAAAGATCAAACCTCAAG-ATATGGCT
MITEPgRS_01  CATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_11  CATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGGAAAGATCAAACCTCATGCATAAGGGC
MITEPgRS_02  CATCAATGCGCTAACTATCAGCTATTTGCAACTATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_06  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_08  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_19  --------------------------------TATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_10  ---------------------------------ATTTTATAGGACTTTCATTGAAGTCTTTTGCCGCAGAGCTGATTCTTAAGTGTTTTTCAGAT-------------------------
MITEPgRS_03  ----------------------------------------------------------------------------------GTGTTTT-CAGGAAAGGTCAAACCTCAAGTATATGGAC
MITEPgRS_05  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_09  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_15  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_17  ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_20  ------------------------------------------------------------------------------------------------------------------------
```

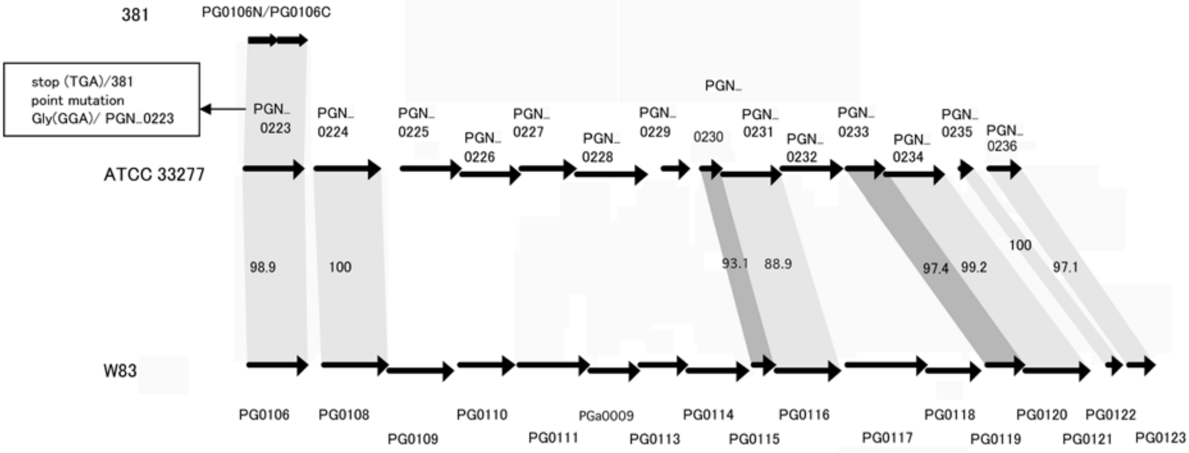
supplementary Figure 2   (continue)

```
                                                                                                       Repeat B
                 942                                                                                                             1061
MITEPgRS_04      -------------------------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_07      -------------------------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCATGTATCTGTCTTGATCGA
MITEPgRS_16      ------------------------------------------------------------------------------------------------------------------------
MITEPgRS_13      -------------------------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCAA
MITEPgRS_18      ACCAAACAAATAGAAAT--------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCACT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_14      -------------------------------------------TACTTGAGGTTTGCAGAGAGAGCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTATCTGTCTTGATCGA
MITEPgRS_12      ACCAAGCAAACAGAAATTTCCCCAAGTTTCCATCAGAGGTA--TACTTGAGCTTTGAAGAGA--TTGCATGAAATACTTCCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_01      -------------------------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCATGTATCTGTCTTGATCGA
MITEPgRS_11      ACCAAACAAATAGAAAT--------------------------TACTTGAGGTTTGCAGAGAGACCGCATGAAGCTCT-CCTTTCTTCGTCGAATCAATGCTTGTATCTGTCTTGATCGA
MITEPgRS_02      -------------------------------------------TACTTGAGGTTTGCAGAGAGATCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_06      -------------------------------------GGTA--TACTTGAGCTTTGAAGAGA--TTGCATGAAATACTTCCTTTCTTCGTCGAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_08      -----------------CCCCGGATTTGGGAGATGAGCTTTTCGAATTTTGGGTCGGGGGAAGTCAAGTCCAATAGGAAAAGCAGACGAAGAGAAGCCAAACTCATCCTGTCTTGATCAA
MITEPgRS_19      -------------------------------------------TACTTGAGGTTTGCAGAGAGACCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_10      -------------------------------------------TACTTGAGGTTTGCAGAGAGACCGCATGAAGCTCT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCGA
MITEPgRS_03      ACCAAGCAAACAGAAATTTCCCCAAGCTTCCATCAGAGGTATATACTTGAGGTTTGCAGAGA--CTGCATGAAGCACT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGATCAA
MITEPgRS_05      -----------------TTCCCCAAGTTTCCA-----------TACTTGAGGTTCGCGAAGA--CTGCATGAAGCACT-CCTTACTTCGTCGAATCAATGCTTGTGTCTGTCTTGATCAA
MITEPgRS_09      ----------------------------------------------TTGAGGTTCGCAGAGA-CCTACATGAGGCTCT-CTTTTCTTCGTCAAATCAATGCTTGTGTTTGTCTTCATCAA
MITEPgRS_15      -------------------------------------------TA-TTGAGGTTTGCAGAGA–CTGCATGAAGTACTA-CTTT-CTTCGTCAAATCAATGCTTATGTCTGTCTTGATCAA
MITEPgRS_17      -----------------------------------------------------------------TGCATAGAGTACT-CCTTTCTTCGTCAAATCAATGCTTGTGTCTGTCTTGA----
MITEPgRS_20      -----------------------------------------TATACTTGGGGTTCGCAGAGA-TCCACATGAAGTTCT-CTTTTCTTCTTCAAATCAATGCTTGTATC------------


                                             IR
                 1062                            1097
MITEPgRS_04      TATGAGAGGGGG---GGGTTATTGTGCAACAGTCTC
MITEPgRS_07      TATGAGGGGG----AGGTTATTGTGCAAAAGTCTC
MITEPgRS_16      ------------------------------------
MITEPgRS_13      TATGAGAGGG-----GGGTTATTGTGCAACGGTCTC
MITEPgRS_18      TATGAGGAGGTGGG-AGGTTATTGTGCAACAGTCTC
MITEPgRS_14      TATGAGGGGGTGGAAGGGTTTTTGTGCAACGGTCTC
MITEPgRS_12      TATGAGGAGGTGG-AAGGTTATTGTGCAACAGTCTC
MITEPgRS_01      TATGAGGGGGGGG-AGGGTTATTGTATAACGGTCTC
MITEPgRS_11      TATGAGGGGGGG-AAGGGTTATTGTGCAACGGTCTC
MITEPgRS_02      TATGAGAGGGGGG-AGGGTTATTGTGCAACGGTCTC
MITEPgRS_06      TATAAGGAGGTGGG-AGGTTATTGTGCAACAGTCTC
MITEPgRS_08      TATGATGGGGTGGAAGGGTTATTCTGCAACGGCCTC
MITEPgRS_19      TATGAGAGAGGGGGGAGGTTATTGTGCAACAGTCTC
MITEPgRS_10      TATGAGAGAGGGGGGAGGTTATTGTGCAACAGTCTC
MITEPgRS_03      TATGAGGGGGTAGGAGGGTTATTGCGCAATGGTCTC
MITEPgRS_05      TATGAAAGGATAGAAGAGTTATTGTGCAATGGTCTC
MITEPgRS_09      TATGAG---GGGGGAGGGTTATTGCGCAAAGGTCTC
MITEPgRS_15      TATGAGGGGATGGAAGGGTTATTGTGCAATGGTCTT
MITEPgRS_17      ------------------------------------
MITEPgRS_20      ------------------------------------
```

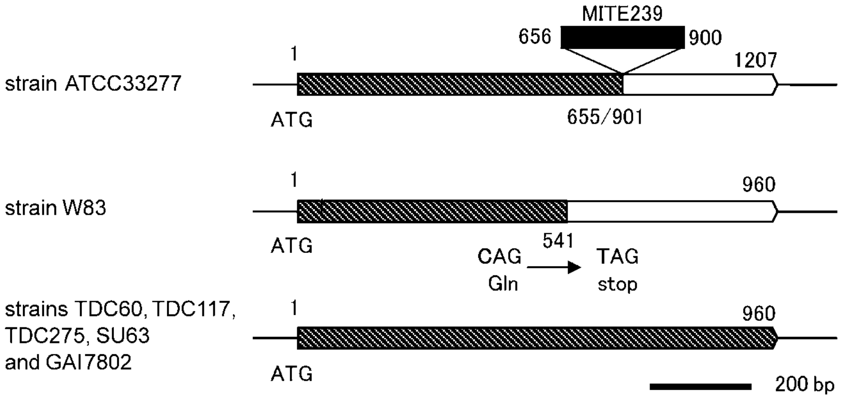
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5 A



Supplementary Figure 5B

```
                      PGN_0379

          -202                                                                                                             -83
33277     TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTACGGAGGAAACAAACTGTATTCGTGGCTGGATATCGTCATCAGTGGCCGGAGGAGCAAACTGCAAAGCAT
GAI7802   TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTGTTCTATTACGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGTCGGAGGAGCAAACTGCAAAGCAT
TDC60     TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTACGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGTCGGAGGAGCAAACTGCAAAGCAT
W83       TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTATGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGCCGGAGGAGCAAACTGCAAAGCAT
SU63      TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTATGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGCCGGAGGAGCAAACTGCAAAGCAT
TDC275    TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTACGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGTCGGAGGAGCAAACTGCAAAGCAT
TDC117    TTATTTCATCGATGTATTGGCCGGTCTGGTTTCGGGACTCTTATTCTATTACGGAGGGAACAAACTGTATTCGTGGCTGGACATCGTCATCAGTGGTCGGAGGAGCAAACTGCAAAGCAT
          ****************************************** ******** ***** *********************** ************** ***********************

          -82                                                                                   1                               34
33277     ACCGGCCGGCCGGTGAACGAA----CGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAATATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
GAI7802   ACCGGCCGGTCGGTGAACGAA----CGAAAAAACGAATATCTTTGTGGATATAGCTAAAGTGAAACATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
TDC60     ACCGGCCGGCCGGTGAACGAA----CGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAACATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
W83       ACCAGCCGGCCGGTGAACGAACGGACGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAACATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
SU63      ACCGGCCGGCCGGTGAACGAACGGACGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAACATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
TDC275    ACCGGCCGGCCGGTGAACGAA----CGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAACATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
TDC117    ACCGGCCGGCCGGTGAACGAA----CGAAAAAACGAATATCTTTGCGGATATAGCTAAAGTGAAATATCTAAAAGACACTATTTCAATGGAAAAGCCTTATGTATTAGGAGTGGATGTAG
          *** ***** ***********    ******************** ******************* ****AAAAG***********ATG*******************************
                                                                                 SD            start

          35                                                                                                               154
33277     GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
GAI7802   GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
TDC60     GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
W83       GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
SU63      GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
TDC275    GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
TDC117    GCGGTACCAACACCGTATTCGGAGTGGTGGACGCTCGTGGCAATTTGGTTATAAGTTCGTCCATCAAGACCGGAGCGCACAACGATCTGAACGATTATATCAAAGATCTGACTGCCGGCA
          ************************************************************************************************************************

          155                                                                                                              274
33277     TCAATCAGTTGATCGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
GAI7802   TCAATCAGTTGATCGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCCCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
TDC60     TCAATCAGTTGATCGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
W83       TCAATCAGTTGATTGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
SU63      TCAATCAGTTGATTGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
TDC275    TCAATCAGTTGATTGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
TDC117    TCAATCAGTTGATTGAACAGGTAGGAGGGAAAGAGAAGATAAAGGGTATCGGTGTAGGTGCTCCCAATGGAAACTATTATACCGGATCGATCGAGTTTGCGCCCAATTTGCCTTGGAAAC
          ************* *********************************************** **********************************************************

          275                                                                                                              394
33277     AGACCAAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGGGAGATGACCTATGGAGCAGCTCGTGGCATGA
GAI7802   AGACCAAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGGGAGATGACCTATGGAGCAGCTCGTGGCATGA
TDC60     AGACCAAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGAGAGATGACCTATGGAGCAGCTCGTGGCATGA
W83       AGACCCAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGAGAGATGACCTATGGAGCAGCTCGTGGCATGA
SU63      AGACCCAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGAGAGATGACCTATGGAGCAGCTCGTGGCATGA
TDC275    AGACCCAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGAGAGATGACCTATGGAGCAGCTCGTGGCATGA
TDC117    AGACCAAAATACCCTTCGCCCAAATGCTCACCGATTCATTAGGTATCCCCACGACACTGACCAATGATGCCAATGCTGCTGCCATCGGAGAGATGACCTATGGAGTAGCTCGTGGCATGA
          ***** ********************************************************************************** **************** **************

          395                                                                                                              514
33277     AGGACTTCATCGTTATCACGCTCGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
GAI7802   AGGACTTCATCGTTATCACGCTCGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
TDC60     AGGACTTCATCGTTATCACGCTTGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
W83       AGGACTTCATCGTTATCACGCTCGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
SU63      AGGACTTCATCGTTATCACGCTCGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
TDC275    AGGACTTCATCGTTATCATGCTTGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
TDC117    AGGACTTCATCGTTATCACGCTCGGTACGGGGGTCGGAAGCGGCATAGTAGTCAATGGCAGCTTGGTGTACGGACACGATGGATTTGCCGGTGAGCTGGGACACATGATCGTTCGCCGCA
          ****************** *** *************************************************************************************************

          515                                                                                                              634
33277     ATGGTCGCATGTGCGGTTGTGGCAGACAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGTGAATATCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
GAI7802   ATGGTCGCATGTGCGGTTGTGGCAGACAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGCGAATATCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
TDC60     ATGGTCGCATGTGCGGTTGTGGCAGGCAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAACTCGCGAATATCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
W83       ATGGTCGCATGTGCGGTTGTGGCAGATAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGCGAATACCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
SU63      ATGGTCGCATGTGCGGTTGTGGCAGACAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGCGAATACCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
TDC275    ATGGTCGCATGTGCGGTTGTGGCAGACAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGCGAATATCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
TDC117    ATGGTCGCATGTGCGGTTGTGGCAGACAGGGCTGTCTCGAGACTTACACCTCTGCTACAGGAGTAGCCCGTACAGCTCGCGAATACCTGGACATCCGTTCGGACAAGAGCTTGCTCCGGA
          *************************  *********************************************** **** ***** **********************************

          635                                                                                                              754
33277     ACATTCAGCCCGATCTGATACGTGAGTTCGATATAAAGAAATGGCACAACACATTGGTTATTAGTGTGTTTTATTTTCCACACAGTGTTTTTCAGGCACTTATATTGCTTTCAGATGGCT
GAI7802   ACATTCAGCCCGATCTGAT-----------------------------------------------------------------------------------------------------
TDC60     ACATTCAGCCCGATCTGAT-----------------------------------------------------------------------------------------------------
W83       ACATTCAGCCCGATCTAAT-----------------------------------------------------------------------------------------------------
SU63      ACATTCAGCCCGATCTGAT-----------------------------------------------------------------------------------------------------
TDC275    ACATTCAGCCCGATCTGAT-----------------------------------------------------------------------------------------------------
TDC117    ACATTCAGCCCGATCTAAT-----------------------------------------------------------------------------------------------------
          **************** **

          755                                  MITE239                                                                     874
33277     TTTTATCTACATCTGAAACCCAATGACTTAGAAGCCGCCCCTTCTTCCCGATCATTATGCAGCCGGAGAATCACAAAAGAGTATGATTTACAGCTACTTGTTATTTTTATGGTATGTTCG
GAI7802   ------------------------------------------------------------------------------------------------------------------------
TDC60     ------------------------------------------------------------------------------------------------------------------------
W83       ------------------------------------------------------------------------------------------------------------------------
SU63      ------------------------------------------------------------------------------------------------------------------------
TDC275    ------------------------------------------------------------------------------------------------------------------------
TDC117    ------------------------------------------------------------------------------------------------------------------------
```

Supplementary Figure 5B (continue)

```
          875                                                                                                                    994
33277     CTTAGATCGAACTGGCGTGATCTGATCACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
GAI7802   --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
TDC60     --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
W83       --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
SU63      --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
TDC275    --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGTGGCGATGGCCTTGCTCAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
TDC117    --------------------------CACTTCGAAAGATGTTTACGATGCAGCTATCAGCGGCGATGGCCTTGCACAAGAGATATTCGAGACTACGGGAGCCATCCTCGGAGAGGCTTTT
                                    ********************************* ************** *********************************************

          991                                                                                                                   1114
33277     GCCGACTTCGTTACATTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTTACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
GAI7802   GCCGACTTCGTTACCTTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTTACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
TDC60     GCCGACTTCGTTACATTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTGACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
W83       GCCGACTTCGTTACCTTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTTACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
SU63      GCCGACTTCGTTACATTCTCCAGTCCGGAAGCGATCATCCTCTTCAGTGGTCTTACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
TDC275    GCCGACTTCGTTACATTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTTACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
TDC117    GCCGACTTCGTTACCTTCTCCAGTCCGGAAGCGATCATCCTCTTCGGTGGTCTGACCAAAGCAGGTGACCTGCTGATGAATCCGATCCGCCACCACATGGAAAAGAATGTGCTGAATATC
          ************** ****************************** ******* ******************************************************************

                                                                                                        PGN_0382
          1115                                                                                                                  1234
33277     TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
GAI7802   TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
TDC60     TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
W83       TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
SU63      TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
TDC275    TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
TDC117    TACCGAGGGAAGACCAAGCTGCTCTTCTCTCAGCTCAAAGAGAGCGATGCCGCTGTGCTCGGGGCAAGTGCTTTGGGATGGGAGGCAAAATAAACGCTGAGCATCCGCTCGCACCTTTCA
          ******************************************************************************************TAA***************************
                                                                                                    stop

          1235                                                                                                                  1354
33277     GATACTGTCCGCGCTGCGGAAACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
GAI7802   GATACTGTCCGCGCTGCGGACACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
TDC60     GATACTGTCCGCGCTGCGGACACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
W83       GATACTGTCCGCGCTGCGGAAACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
SU63      GATACTGTCCGCGCTGCGGAAACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
TDC275    GATACTGTCCGCGCTGCGGAAACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
TDC117    GATACTGTCCGCGCTGCGGAAACGATGGTTTTGCGGAGAGAAACCCGAAAGCCAAATCGTGTCCGAAGTGCGGATTGATTTATTATGCCAATCCTTCTGCTGCGACGGCTTGTTTCATTA
          ******************** ***************************************************************************************************

          1355                                                                 1426
33277     CCGATAGTGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGCACACTCGATTTGCCCG
GAI7802   CCGATAGCGCAGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGCACACTCGATTTGCCCG
TDC60     CCGATAGCGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGCACACTCGATTTGCCCG
W83       CCGATAGCGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGCACACTCGATTTGCCCG
SU63      CCGATAGCGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGTACACTCGATTTGCCCG
TDC275    CCGATAGCGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGTACACTCGATTTGCCCG
TDC117    CCGATAGCGCGGGCAGACTATTGGCCGTGCGCCGAGCCAAAGATCCGGCCAAAGGCACACTCGATTTGCCCG
          ******* ** ******************************************* ****************
```

Supplementary Table 1. Primers used in this study

target region	primer name	sequence
rrn operon	rrn1-Up	CACAAGTCAGAACATGGCCGATGAC
	rrn1-Down	GAGATGTCCGAAAGTCCATGTCACC
	rrn2-Up	GTTGTAGGACAGCAACCTTTTGGCG
	rrn2-Down	GCTATGGATATTCTGCGGTGTACGG
	rrn3-Up	AGATAGCCAGTTTCGTTACGTCCGC
	rrn3-Down	ATACAGCAACGGTTACTTCCGCGAC
	rrn4-Up	AACACTCCCCACCGGCAAAAACATC
	rrn4-Down	TATCGCAGGCACAAACCGCTTTACC
CTnPg1-a	CTnPg1-up	ATCCGTCCACGTGTGTGATATTGT
	CTnPg1-down	GATTTCATACGGCGGGAGTAAC
	CTnPg1-left	TATACCTCCTCATCGGACAGACC
	CTnPg1-right	GTTTTGGTAACGCTGCTCTGTC
Glucose kinase	Glyco-up	TATTCTGGTACCTCCTGCCC
	Glyco-down	ATTCCTGTCTCCTCACGGAT
	Glk-Real-F	ATGTAGGCGGATCCAACACC
	Glk-Real-R	CCGGCAGTCAGATCTTTGAT
Kgp	Kgp-Real-F	ATGTGGGCAGGTAATCTTGC
	Kgp-Real-R	AGCCTGATTCTGAGGCAGAG
16S rRNA	16S-Real-F	CTTGACTTCAGTGGCGGCA
	16S-Real-R	AGGGAAGACGGTTTTCACCA

Supplementary Table 2. Newly annotated 114 CDSs (PGa No.) of the W83 genome in this study

PG#	start	stop	strand	length(aa)	product
PGa0001	6487	6642	+	51	hypothetical protein
PGa0002	35136	34594	-	180	putative SpoU rRNA methylase family protein
PGa0003	41766	41984	+	72	hypothetical protein
PGa0004	62287	62556	+	89	IS*Pg1*, transposase, partial
PGa0005	62562	63281	+	239	IS*Pg1*, transposase, partial
PGa0006	82439	82020	-	139	hypothetical protein
PGa0007	119278	118853	-	141	hypothetical protein similar to PG2076
PGa0008	128357	128061	-	98	hypothetical protein similar to PG1787
PGa0009	135230	136030	+	266	conserved hypothetical protein
PGa0010	154352	153042	-	436	putative CDP-glycerol glycerophosphotransferase
PGa0011	168495	168109	-	128	putative protein-export transmembrane protein
PGa0012	189608	189348	-	86	hypothetical protein
PGa0013	198679	199101	+	140	partial Mfa1 fimbrilin
PGa0014	200458	201843	+	461	partial Mfa1 fimbrilin
PGa0015	223466	223284	-	60	IS*Pg6*, transposase, partial
PGa0016	223791	223576	-	71	IS*Pg6*, transposase, partial
PGa0017	236419	236577	+	52	hypothetical protein similar to PG1426
PGa0018	299068	298868	-	66	IS*Pg1*, transposase, partial
PGa0019	326843	327433	+	196	putative chromate transport protein
PGa0020	327452	328045	+	197	chromate transport protein
PGa0021	335816	335691	-	41	hypothetical protein similar to PG0297
PGa0022	336545	336207	-	112	IS*Pg3*, transposase, partial
PGa0023	422084	422332	+	82	putative preprotein translocase SecE subunit
PGa0024	461786	461989	+	67	hypothetical protein
PGa0025	514604	515005	+	133	hypothetical protein
PGa0026	586913	586698	-	71	hypothetical protein
PGa0027	607412	607624	+	70	hypothetical protein
PGa0028	696004	694832	-	390	tonB-linked receptor Tlr, frame-shifted with PGN_0683
PGa0029	697076	695718	-	452	tonB-linked receptor Tlr, frame-shifted with PGN_0683
PGa0030	715388	715894	+	168	conserved hypothetical protein frame-shifted with PGN_0702
PGa0031	715858	716475	+	205	conserved hypothetical protein with a phosphotransferase enzyme family domain, frame-shifted with PGN_0702
PGa0032	782156	780975	-	393	pseudouridine synthase
PGa0033	803453	804211	+	252	pseudouridine synthase
PGa0034	811778	811059	-	239	IS*Pg1*, transposase, partial
PGa0035	812140	811784	-	118	IS*Pg1*, transposase, partial
PGa0036	814830	815186	+	118	IS*Pg1*, transposase, partial
PGa0037	815192	815911	+	239	IS*Pg1*, transposase, partial
PGa0038	858188	857958	-	76	hypothetical protein similar to PG1251
PGa0039	873915	873736	-	59	IS*Pg1*, transposase, partial
PGa0040	874162	873938	-	74	IS*Pg1*, transposase, partial
PGa0041	874751	874251	-	166	IS*Pg1*, transposase, partial
PGa0042	875588	874839	-	249	IS*Pg1*, transposase, partial
PGa0043	895133	894084	-	349	integrase
PGa0044	906634	905915	-	239	IS*Pg1*, transposase, partial
PGa0045	906996	906640	-	118	IS*Pg1*, transposase, partial
PGa0046	907656	907018	-	212	probable integrase
PGa0047	913013	913369	+	118	IS*Pg1*, transposase, partial
PGa0048	913375	914094	+	239	IS*Pg1*, transposase, partial
PGa0049	988353	987640	-	237	hypothetical protein
PGa0050	1001219	1000500	-	239	IS*Pg1*, transposase, partial
PGa0051	1001581	1001225	-	118	IS*Pg1*, transposase, partial
PGa0052	1004952	1004410	-	180	IS*Pg1*, transposase, partial
PGa0053	1025199	1024489	-	236	probable CDP-diacylglycerol-serine O-phosphatidyltransferase

PGa0054	1049668	1049243	-	141	IS*Pg1*, transposase, partial
PGa0055	1050326	1049757	-	189	IS*Pg1*, transposase, partial
PGa0056	1085800	1085201	-	199	probable para-aminobenzoate synthase component I
PGa0057	1086820	1085807	-	337	putative para-aminobenzoate synthase component I
PGa0058	1184183	1183506	-	225	conserved hypothetical protein
PGa0059	1184906	1184205	-	233	conserved hypothetical protein
PGa0060	1198996	1197677	-	439	pseudouridine synthase
PGa0061	1228291	1227332	-	319	peptide chain release factor RF-2
PGa0062	1235147	1235752	+	201	conserved hypothetical protein
PGa0063	1241429	1240071	-	452	cobyric acid synthase
PGa0064	1269922	1269659	-	87	IS*Pg2*, transposase, partial
PGa0065	1279266	1278262	-	334	putative transcriptional regulator
PGa0066	1299868	1301499	+	543	transglycosylase
PGa0067	1305183	1304464	-	239	IS*Pg1*, transposase, partial
PGa0068	1305545	1305189	-	118	IS*Pg1*, transposase, partial
PGa0069	1313669	1312986	-	227	putative pseudouridine synthase
PGa0070	1410604	1411761	+	385	putative NAD(P)(+) transhydrogenase (AB-specific)
PGa0071	1411871	1412185	+	104	probable NAD(P) transhydrogenase subunit alpha part 2
PGa0072	1430600	1429881	-	239	IS*Pg1*, transposase, partial
PGa0073	1430956	1430606	-	116	IS*Pg1*, transposase, partial
PGa0074	1463878	1464234	+	118	IS*Pg1*, transposase, partial
PGa0075	1464240	1464959	+	239	IS*Pg1*, transposase, partial
PGa0076	1482344	1481625	-	239	IS*Pg1*, transposase, partial
PGa0077	1482706	1482350	-	118	IS*Pg1*, transposase, partial
PGa0078	1595556	1594936	-	206	IS*Pg3*, transposase, partial
PGa0079	1618623	1619399	+	258	tRNA pseudouridine synthase B
PGa0080	1624846	1627503	+	885	PrtT cysteine protease/hemagglutinin
PGa0081	1724068	1721861	-	735	cation-transporting ATPase
PGa0082	1724367	1724104	-	87	hypothetical protein
PGa0083	1759812	1760174	+	120	probable partial hemagglutinin-related protein
PGa0084	1760051	1760386	+	111	probable partial hemagglutinin-related protein
PGa0085	1765568	1766065	+	165	putative ABC-type transporter ATP-binding protein
PGa0086	1793919	1792945	-	324	probable pseudouridine synthase
PGa0087	1823492	1823157	-	111	putative partial transcriptional repressor
PGa0088	1824230	1823574	-	218	putative partial transcriptional repressor
PGa0089	1843634	1843843	+	69	hypothetical protein
PGa0090	1939821	1945016	+	1731	lysine-specific cysteine proteinase
PGa0091	1945284	1945640	+	118	IS*Pg1*, transposase, partial
PGa0092	1945646	1945915	+	89	IS*Pg1*, transposase, partial
PGa0093	1988553	1989611	+	352	hypothetical protein
PGa0094	2005261	2005001	-	86	IS*Pg3* transposase, partial
PGa0095	2007246	2006695	-	183	IS*Pg3* transposase, partial
PGa0096	2009798	2009079	-	239	IS*Pg1* transposase, partial
PGa0097	2039449	2039727	+	92	hypothetical protein similar to PG1547
PGa0098	2046082	2045909	-	57	hypothetical protein
PGa0099	2048288	2048139	-	49	hypothetical protein
PGa0100	2052066	2051440	-	208	conserved hypothetical protein
PGa0101	2055075	2056247	+	390	partial chloride channel protein
PGa0102	2080809	2079226	-	527	conserved hypothetical protein
PGa0103	2101989	2101270	-	239	IS*Pg1*, transposase, partial
PGa0104	2102351	2101995	-	118	IS*Pg1*, transposase, partial
PGa0105	2154260	2153955	-	101	IS*Pg2*, transposase, partial
PGa0106	2156466	2156783	+	105	IS*Pg1*, transposase, partial
PGa0107	2156789	2157508	+	239	IS*Pg1*, transposase, partial
PGa0108	2222925	2223965	+	346	hypothetical protein
PGa0109	2230037	2229726	-	103	IS*Pg2*, transposase, partial
PGa0110	2274662	2274886	+	74	IS*Pg1*, transposase, partial

PGa0111	2275269	2275988	+	239	IS*Pg1*, transposase, partial
PGa0112	2300430	2299711	-	239	IS*Pg1*, transposase, partial
PGa0113	2300792	2300436	-	118	IS*Pg1*, transposase, partial
PGa0114	2303753	2304634	+	293	immunoreactive 32 kD antigen PG25

Supplementary Table 3. The multi-copied CDS groups in ATCC 33277

CDS group	product	representative CDS					
Cpgn01	transposase in IS*Pg1*	PGN_0027	PGN_0104	PGN_0587	PGN_1077	PGN_1319	PGN_1839
			PGN_0108	PGN_0602	PGN_1110	PGN_1397	PGN_1901
			PGN_0130	PGN_0605	PGN_1118	PGN_1420	PGN_1926
			PGN_0131	PGN_0644	PGN_1132	PGN_1433	PGN_1957
			PGN_0210	PGN_0749	PGN_1161	PGN_1442	
			PGN_0212	PGN_0834	PGN_1177	PGN_1619	
			PGN_0218	PGN_0842	PGN_1191	PGN_1644	
			PGN_0324	PGN_0918	PGN_1216	PGN_1669	
			PGN_0575	PGN_0971	PGN_1276	PGN_1727	
			PGN_0577	PGN_1008	PGN_1280	PGN_1838	
Cpgn02	transposase in IS*Pg3*	PGN_0442	PGN_0106	PGN_0790	PGN_0953	PGN_1066	PGN_1911
			PGN_0107	PGN_0864	PGN_0955	PGN_1278	PGN_1913
			PGN_0425	PGN_0879	PGN_1006	PGN_1279	PGN_1922
			PGN_0454	PGN_0909	PGN_1060	PGN_1428	
			PGN_0459	PGN_0934	PGN_1064	PGN_1726	
			PGN_0585	PGN_0944	PGN_1065	PGN_1836	
Cpgn03	conserved hypothetical protein	PGN_1478	PGN_0363	PGN_0305	PGN_0337	PGN_0551	PGN_1778
Cpgn04	transposase in IS*Pg2*	PGN_1394	PGN_0839	PGN_1160			
Cpgn05	conserved hypothetical protein	PGN_0951	PGN_0105	PGN_1277	PGN_1915		
Cpgn06	hypothetical protein	PGN_0127	PGN_0443	PGN_0603	PGN_1887		
Cpgn07	conserved hypothetical protein	PGN_0216	PGN_0439	PGN_1399	PGN_2048		
Cpgn08	conserved hypothetical protein	PGN_0217	PGN_0440	PGN_1398	PGN_2049		
Cpgn09	conserved hypothetical protein	PGN_0215	PGN_0438	PGN_2047			
Cpgn10	conserved hypothetical protein	PGN_1477	PGN_0304				
Cpgn11	hypothetical protein	PGN_1082	PGN_0479	PGN_1083			
Cpgn12	conserved hypothetical protein	PGN_0652	PGN_0536	PGN_1592			
Cpgn13	hypothetical protein	PGN_0979	PGN_1102	PGN_1257			
Cpgn14	conserved hypothetical protein	PGN_0946	PGN_1920				
Cpgn15	ABC transporter ATP-binding protein	PGN_0949	PGN_1917				
Cpgn16	carboxyl-terminal processing protease	PGN_0952	PGN_1914				
Cpgn17	ABC transporter ATP-binding protein	PGN_0950	PGN_1916				
Cpgn18	conserved hypothetical protein	PGN_1821	PGN_1824				
Cpgn19	conserved protein found in conjugative transposon, TraM	PGN_0060	PGN_1281				
Cpgn20	putative TonB	PGN_0103	PGN_1275				
Cpgn21	conserved hypothetical protein	PGN_0948	PGN_1918				
Cpgn22	conserved hypothetical protein	PGN_0050	PGN_1291				
Cpgn23	conserved hypothetical protein	PGN_0053	PGN_1288				
Cpgn24	diaminopimelate decarboxylase	PGN_0100	PGN_1272				
Cpgn25	hemagglutinin protein HagB	PGN_1904	PGN_1906				
Cpgn26	conserved hypothetical protein	PGN_1697	PGN_0029				
Cpgn27	conserved protein found in conjugative transposon, TraN	PGN_0059	PGN_1282				
Cpgn28	putative	PGN_0101	PGN_1273				

	1,4-dihydroxy-2-naphthoate octaprenyltransferase		
Cpgn29	probable conserved protein found in conjugative transposon, TraP	PGN_0057	PGN_1284
Cpgn30	conserved hypothetical protein	PGN_0306	PGN_1237
Cpgn31	conserved hypothetical protein	PGN_0947	PGN_1919
Cpgn32	putative TetR family transcriptional regulator	PGN_0945	PGN_1921
Cpgn33	probable conserved protein found in conjugative transposon	PGN_0058	PGN_1283
Cpgn34	partial transposase in IS*Pg1*	PGN_0219	PGN_0966
Cpgn35	probable elongation factor P	PGN_0616	PGN_1405
Cpgn36	probable anti-restriction protein	PGN_0049	PGN_1292
Cpgn37	probable lysozyme	PGN_0055	PGN_1286
Cpgn39	DNA-binding protein, histone-like family	PGN_0614	PGN_1407
Cpgn40	DNA-binding protein, histone-like family	PGN_0326	PGN_0460
Cpgn41	probable conserved protein found in conjugative transposon	PGN_0056	PGN_1285
Cpgn42	partial transposase in IS*Pg1*	PGN_0220	PGN_0967
Cpgn43	conserved hypothetical protein	PGN_0048	PGN_1293
Cpgn44	hypothetical protein	PGN_0054	PGN_1287
Cpgn45	conserved hypothetical protein	PGN_0102	PGN_1274
Cpgn46	conserved hypothetical protein	PGN_0437	PGN_2046
Cpgn47	conserved hypothetical protein	PGN_1835	PGN_1007
Cpgn48	hypothetical protein	PGN_0046	PGN_1295
Cpgn49	conserved hypothetical protein	PGN_0051	PGN_1290
Cpgn50	hypothetical protein	PGN_0052	PGN_1289
Cpgn51	conserved hypothetical protein	PGN_0047	PGN_1294
Cpgn52	hypothetical protein	PGN_0615	PGN_1406
Cpgn53	conserved hypothetical protein	PGN_1910	PGN_1936

Supplementary Table 4. The multi-copied CDS groups in W83

CDS group	product	representative CDS					
Cpg01	IS*Pg1*, transposase	PG0184	PGa0004	PG0825	PG1177	PGa0075	PGa0103
			PGa0005	PGa0044	PG1197	PGa0076	PGa0104
			PG0460	PGa0045	PGa0067	PGa0077	PGa0106
			PG0549	PGa0047	PGa0068	PG1448	PGa0107
			PGa0034	PGa0048	PG1244	PG1624	PGa0110
			PGa0035	PGa0050	PG1320	PGa0091	PGa0111
			PGa0036	PGa0051	PGa0072	PGa0092	PGa0112
			PGa0037	PGa0052	PGa0073	PG1906	PGa0113
			PGa0042	PG1031	PGa0074	PGa0096	
Cpg02	IS*Pg5* transposase Orf2	PG0008	PG0041	PG0458	PG0943	PG1644	PG2057
			PG0426	PG0591	PG1420	PG1710	PG2128
Cpg03	IS*Pg4,* transposase	PG0019	PG0050	PG0225	PG0970	PG1658	PG2194
			PG0177	PG0487	PG1261	PG1673	
Cpg04	IS*Pg5* transposase Orf1	PG0590	PG0009	PG0427	PG1419	PG1709	PG2129
			PG0040	PG0942	PG1645	PG2058	
Cpg05	IS*Pg3*, transposase	PG0194	PG0261	PG1032	PG1262	PGa0094	PGa0095
			PG0798				
Cpg06	IS*Pg2*, transposase	PG0277	PG0865	PG1350	PG1746	PG2176	
Cpg07	hypothetical protein	PG0499	PG1059	PG1398	PG2220		
Cpg08	hypothetical protein	PG0101	PG1266	PG1671	PG2075		
Cpg09	hypothetical protein	PG0102	PG1265	PG1670	PG2074		
Cpg10	hypothetical protein	PG0103	PG1264	PG1669	PG2073		
Cpg11	IS*Pg1*, transposase, partial	PGa0054	PGa0039	PGa0040			
Cpg12	hypothetical protein	PG0100	PG1267	PG1672			
Cpg13	conserved hypothetical protein	PG1526	PG1512				
Cpg14	MATE efflux family protein	PG0827	PG1446				
Cpg15	integrase	PG0819	PG1454				
Cpg16	integrase	PG0820	PG1453				
Cpg17	hemagglutinin protein HagB	PG1972	PG1975				
Cpg18	hypothetical protein	PG1757	PG0031				
Cpg19	transcriptional regulator, AraC family	PG0826	PG1447				
Cpg20	conserved hypothetical protein	PG0824	PG1449				
Cpg21	IS*Pg1*, transposase, partial	PGa0055	PGa0041				
Cpg22	translation elongation factor P	PG0568	PG1274				
Cpg23	DNA-binding protein, histone-like family	PG0566	PG1276				
Cpg24	DNA-binding protein, histone-like family	PG0853	PG2152				
Cpg25	hypothetical protein	PG0829	PG1444				
Cpg26	hypothetical protein	PG0831	PG1442				
Cpg27	hypothetical protein	PG0822	PG1451				
Cpg28	lipoprotein, putative	PG0821	PG1452				
Cpg29	conserved hypothetical protein	PG0823	PG1450				
Cpg30	hypothetical protein	PG1527	PG1532				
Cpg31	hypothetical protein	PG0265	PG1426				
Cpg32	hypothetical protein	PG1574	PG0371				

Supplementary Table 5. ATCC 33277-specific CDSs

	PGN No.	product	
	PGN_0018	hypothetical protein	
	PGN_0019	hypothetical protein	
	PGN_0028	hypothetical protein	*
	PGN_0032	conserved hypothetical protein	*
	PGN_0036	hypothetical protein	
	PGN_0040	hypothetical protein	
CTnPg1-a	PGN_0046	hypothetical protein	
	PGN_0047	conserved hypothetical protein	
	PGN_0048	conserved hypothetical protein	
	PGN_0049	probable anti-restriction protein	
	PGN_0050	conserved hypothetical protein	
	PGN_0051	conserved hypothetical protein	
	PGN_0052	hypothetical protein	
	PGN_0053	conserved hypothetical protein	
	PGN_0054	hypothetical protein	
	PGN_0055	probable lysozyme	*
	PGN_0056	probable conserved protein found in conjugative transposon	*
	PGN_0057	probable conserved protein found in conjugative transposon, TraP	*
	PGN_0058	probable conserved protein found in conjugative transposon	*
	PGN_0059	conserved protein found in conjugative transposon, TraN	*
	PGN_0060	conserved protein found in conjugative transposon, TraM	*
	PGN_0061	hypothetical protein	
	PGN_0062	putative conserved protein found in conjugative transposon, TraK	*
	PGN_0063	conserved transmembrane protein found in conjugative transposon, TraJ	*
	PGN_0064	putative conserved protein found in conjugative transposon, TraI	*
	PGN_0065	conserved protein found in conjugative transposon, TraG	*
	PGN_0066	probable conserved transmembrane protein found in conjugative transposon, TraF	*
	PGN_0067	probable conserved transmembrane protein found in conjugative transposon, TraE	*
	PGN_0068	hypothetical protein	
	PGN_0069	probable conserved protein found in conjugative transposon, TraA	*
	PGN_0070	hypothetical protein	
	PGN_0071	hypothetical protein	*
	PGN_0072	hypothetical protein	*
	PGN_0073	putative conserved protein found in conjugative transposon, TraA	*
	PGN_0074	conserved hypothetical protein	*
	PGN_0075	conserved hypothetical protein	*
	PGN_0076	putative mobilization protein, TraG family	*
	PGN_0077	hypothetical protein	
	PGN_0078	hypothetical protein	
	PGN_0079	conserved hypothetical protein with DUF1016 domain	
	PGN_0080	probable tetracycline resistance element mobilization regulatory protein RteC	
	PGN_0082	probable transcriptional regulator, AraC family	*
	PGN_0083	conserved hypothetical protein	*
	PGN_0084	DNA topoisomerase I	*
	PGN_0085	hypothetical protein	*
	PGN_0086	putative DNA methylase	*
	PGN_0087	conserved hypothetical protein	
	PGN_0088	putative transcriptional regulator	*
	PGN_0089	hypothetical protein	
	PGN_0090	hypothetical protein	
	PGN_0091	hypothetical protein	
	PGN_0092	conserved hypothetical protein	*
	PGN_0093	conserved hypothetical protein	*
	PGN_0094	putative bacteriophage integrase	*
	PGN_0095	hypothetical protein	
	PGN_0098	hypothetical protein	

	PGN_0102	conserved hypothetical protein	*
	PGN_0105	conserved hypothetical protein	*
	PGN_0109	conserved hypothetical protein	*
	PGN_0110	hypothetical protein	
	PGN_0111	partial transposase in IS*Pg6*	*
	PGN_0112	partial transposase Orf1 in IS*Pg5*	*
	PGN_0113	hypothetical protein	
	PGN_0124	hypothetical protein	
	PGN_0146	hypothetical protein	
	PGN_0154	conserved hypothetical protein	*
	PGN_0155	hypothetical protein	
	PGN_0161	putative thiamine biosynthesis protein ThiS	*
	PGN_0162	conserved hypothetical protein	*
	PGN_0163	hypothetical protein	*
	PGN_0164	conserved hypothetical protein	*
	PGN_0165	conserved hypothetical protein	
	PGN_0177	hypothetical protein	
	PGN_0180	FimA type I fimbrilin	*
	PGN_0181	conserved hypothetical protein	*
	PGN_0182	conserved hypothetical protein	*
	PGN_0183	minor component FimC	*
	PGN_0184	minor component FimD	*
	PGN_0185	minor component FimE	*
	PGN_0186	hypothetical protein	
	PGN_0187	conserved hypothetical protein	*
	PGN_0214	conserved hypothetical protein	*
	PGN_0215	conserved hypothetical protein	*
	PGN_0222	conserved hypothetical protein	*
GP1 locus	PGN_0225	probable glycosyltransferase	*
	PGN_0226	conserved hypothetical protein	
	PGN_0227	probable glycosyl transferase, family 1	*
	PGN_0228	probable coenzyme F390 synthetase	*
	PGN_0229	putative DNA-binding protein, histone-like family	*
	PGN_0231	probable delta-aminolevulinic acid dehydratase	*
	PGN_0232	probable glycosyl transferase, family 2	*
	PGN_0236	conserved hypothetical protein	*
	PGN_0244	hypothetical protein	
	PGN_0273	conserved hypothetical protein	*
	PGN_0280	hypothetical protein	*
	PGN_0285	pyridine nucleotide-disulphide oxidoreductase	*
	PGN_0286	hypothetical protein frame-shifted with PGN0285	*
	PGN_0287	Mfa1 fimbrilin	*
	PGN_0292	hypothetical protein	
	PGN_0293	receptor antigen A	*
	PGN_0294	receptor antigen B	*
	PGN_0295	C-terminal domain of Arg- and Lys-gingipain proteinase	*
	PGN_0296	conserved hypothetical protein	*
	PGN_0304	conserved hypothetical protein	*
	PGN_0306	conserved hypothetical protein	*
	PGN_0323	conserved hypothetical protein	*
	PGN_0325	hypothetical protein	
	PGN_0339	hypothetical protein	
	PGN_0343	hypothetical protein	
	PGN_0344	probable haloacid dehalogenase-like hydrolase	*
	PGN_0349	upregulated in stationary phase protein A	*
	PGN_0362	hypothetical protein	
	PGN_0364	hypothetical protein	*
	PGN_0372	hypothetical protein	
	PGN_0378	putative exopolyphosphatase	*
	PGN_0379	conserved hypothetical protein	*

	PGN_0380	partial ROK family transcriptional repressor with glucose kinase domain	*
	PGN_0381	partial ROK family transcriptional repressor with glucose kinase domain	*
	PGN_0401	conserved hypothetical protein	*
	PGN_0402	hypothetical protein	
	PGN_0407	hypothetical protein	
	PGN_0417	conserved hypothetical protein	*
	PGN_0430	putative ATP-binding component of ABC transporter protein	*
	PGN_0435	probable partial hemagglutinin-related protein	*
	PGN_0437	conserved hypothetical protein	*
	PGN_0438	conserved hypothetical protein	*
	PGN_0441	hypothetical protein	
	PGN_0449	conserved hypothetical protein	*
	PGN_0452	conserved hypothetical protein	
	PGN_0453	partial transposase Orf1 in IS*Pg5*	
	PGN_0455	partial transposase Orf1 in IS*Pg5*	
	PGN_0458	hypothetical protein	
	PGN_0467	conserved hypothetical protein	
	PGN_0474	hypothetical protein	
	PGN_0475	hypothetical protein	
	PGN_0478	partial transposase in IS*Pg4*	*
	PGN_0479	hypothetical protein	
	PGN_0480	partial transposase in IS*Pg4*	*
	PGN_0481	hypothetical protein	
	PGN_0482	probable immunoreactive 23 kDa antigen	*
	PGN_0488	conserved hypothetical protein	*
	PGN_0494	hypothetical protein	
	PGN_0495	conserved hypothetical protein	*
	PGN_0505	hypothetical protein	
	PGN_0506	hypothetical protein	
	PGN_0536	conserved hypothetical protein	*
	PGN_0537	probable transcriptional regulator, Crp family	*
	PGN_0541	hypothetical protein	
	PGN_0552	hypothetical protein	
	PGN_0558	conserved hypothetical protein	*
	PGN_0560	hypothetical protein	
	PGN_0563	conserved hypothetical protein	
	PGN_0572	conserved hypothetical protein	*
	PGN_0576	hypothetical protein	*
CTnPg2	PGN_0578	conserved hypothetical protein found in conjugative transposon	*
	PGN_0579	conserved hypothetical protein	*
	PGN_0580	conserved hypothetical protein	
	PGN_0581	conserved hypothetical protein	*
	PGN_0582	DNA topoisomerase I	*
	PGN_0583	conserved hypothetical protein	*
	PGN_0584	conserved hypothetical protein	*
	PGN_0586	conserved hypothetical protein	
	PGN_0588	conserved hypothetical protein	
	PGN_0589	conserved hypothetical protein	*
	PGN_0590	putative Fic family protein	
	PGN_0591	conserved hypothetical protein	
	PGN_0592	putative conserved protein found in conjugative transposon, TraQ	*
	PGN_0593	putative conserved protein found in conjugative transposon, TraO	*
	PGN_0594	conserved protein found in conjugative transposon, TraN	*
	PGN_0595	putative conserved protein found in conjugative transposon, TraM	*
	PGN_0596	conserved hypothetical protein found in conjugative transposon	
	PGN_0597	putative conserved protein found in conjugative transposon, TraK	*
	PGN_0598	conserved transmembrane protein found in conjugative transposon, TraJ	*
	PGN_0599	putative conserved protein found in conjugative transposon, TraI	*
	PGN_0600	conserved hypothetical protein	*
	PGN_0601	conserved hypothetical protein	

	PGN_0615	hypothetical protein	
	PGN_0648	conserved hypothetical protein	*
	PGN_0651	conserved hypothetical protein	
	PGN_0652	conserved hypothetical protein	*
	PGN_0653	conserved hypothetical protein	*
	PGN_0654	conserved hypothetical protein	*
	PGN_0657	conserved hypothetical protein	*
	PGN_0681	hypothetical protein	
	PGN_0682	hypothetical protein	
	PGN_0683	TonB-linked receptor Tlr	*
	PGN_0702	conserved hypothetical protein	*
	PGN_0703	hypothetical protein	
	PGN_0712	conserved hypothetical protein	
	PGN_0717	conserved hypothetical protein	*
	PGN_0719	probable ABC transporter permease protein	*
	PGN_0730	hypothetical protein	
	PGN_0731	hypothetical protein	
	PGN_0734	conserved hypothetical protein	*
	PGN_0748	hypothetical protein	
	PGN_0752	hypothetical protein	*
	PGN_0755	hypothetical protein	*
	PGN_0783	putative DNA-binding protein, histone-like family	*
	PGN_0784	hypothetical protein	
	PGN_0785	hypothetical protein	
	PGN_0793	4-alpha-glucanotransferase	*
	PGN_0795	conserved hypothetical protein	*
	PGN_0798	conserved hypothetical protein	*
	PGN_0820	hypothetical protein	
	PGN_0821	hypothetical protein	
	PGN_0835	conserved hypothetical protein	
	PGN_0840	conserved hypothetical protein	*
	PGN_0843	conserved hypothetical protein	*
	PGN_0849	conserved hypothetical protein	
	PGN_0850	conserved hypothetical protein	
	PGN_0851	conserved hypothetical protein	
	PGN_0854	hypothetical protein	
	PGN_0855	hypothetical protein	
	PGN_0862	Type III restriction enzyme, res subunit	
	PGN_0863	DNA methylase N-4/N-6	
	PGN_0873	conserved hypothetical protein	*
	PGN_0877	SNF2-related helicase	*
	PGN_0878	hypothetical protein	
	PGN_0881	conserved hypothetical protein	*
	PGN_0892	hypothetical protein	
	PGN_0899	conserved hypothetical protein	
	PGN_0905	putative dihydroorotate dehydrogenase	*
	PGN_0908	hypothetical protein	
CTnPg3	PGN_0917	tyrosine type site-specific recombinase	*
	PGN_0919	hypothetical protein	
	PGN_0920	putative partial excisionase	
	PGN_0921	hypothetical protein	
	PGN_0922	conserved hypothetical protein	
	PGN_0923	putative DNA primase	*
	PGN_0924	mobilization protein	*
	PGN_0925	putative mobilization protein	*
	PGN_0926	conserved hypothetical protein	
	PGN_0927	conserved hypothetical protein	*
	PGN_0928	conserved hypothetical protein	
	PGN_0929	conserved hypothetical protein	
	PGN_0930	hypothetical protein	

PGN_0931	conserved hypothetical protein	
PGN_0932	probable GCN5-related N-acetyltransferase	
PGN_0933	probable transcriptional regulator	*

PGN_0937	hypothetical protein	
PGN_0939	conserved hypothetical protein	*
PGN_0940	partial transposase in IS*Pg2*	*
PGN_0951	conserved hypothetical protein	*
PGN_0954	partial transposase in IS*Pg6*	*
PGN_0956	hypothetical protein	*
PGN_0959	probable transcriptional regulator	*
PGN_0972	TPR domain protein	*
PGN_0978	hypothetical protein	
PGN_0979	hypothetical protein	
PGN_0983	hypothetical protein	
PGN_0991	putative ribonuclease BN	*
PGN_0994	conserved hypothetical protein	*
PGN_0995	TPR domain protein	*
PGN_0996	TPR domain protein	*
PGN_1009	calcium-transporting ATPase	*
PGN_1016	conserved hypothetical protein	*
PGN_1018	conserved hypothetical protein	*
PGN_1028	conserved hypothetical protein	*
PGN_1030	hypothetical protein	
PGN_1031	conserved hypothetical protein	*
PGN_1048	hypothetical protein	
PGN_1051	hypothetical protein	
PGN_1058	putative bacterioferritin comigratory protein	*
PGN_1059	conserved hypothetical protein	*
PGN_1061	conserved hypothetical protein	*
PGN_1063	partial transposase Orf2 in IS*Pg5*	*
PGN_1067	hypothetical protein	
PGN_1068	conserved hypothetical protein	*
PGN_1069	hypothetical protein	
PGN_1070	hypothetical protein	
PGN_1071	hypothetical protein	
PGN_1072	hypothetical protein	
PGN_1073	hypothetical protein	
PGN_1074	hypothetical protein	
PGN_1075	hypothetical protein	*
PGN_1076	putative DNA methylase	*
PGN_1082	hypothetical protein	
PGN_1083	hypothetical protein	
PGN_1084	hypothetical protein	
PGN_1086	hypothetical protein	
PGN_1099	probable phosphoesterase	*
PGN_1101	conserved hypothetical protein	
PGN_1102	hypothetical protein	
PGN_1109	hypothetical protein	
PGN_1114	conserved hypothetical protein	*
PGN_1122	NADPH-NAD transhydrogenase beta subunit	*
PGN_1129	conserved hypothetical protein	*
PGN_1133	hypothetical protein	
PGN_1143	conserved hypothetical protein	*
PGN_1144	hypothetical protein	
PGN_1167	conserved hypothetical protein	*
PGN_1180	conserved hypothetical protein	*
PGN_1183	hypothetical protein	
PGN_1192	DNA-binding protein, histone-like family	*
PGN_1199	DNA-binding protein, histone-like family	*
PGN_1210	hypothetical protein	

	PGN_1211	hypothetical protein	
	PGN_1212	hypothetical protein	
	PGN_1213	putative ATP-binding protein	
	PGN_1214	hypothetical protein	
	PGN_1215	hypothetical protein	
	PGN_1217	hypothetical protein	
	PGN_1227	TPR domain protein	*
	PGN_1228	hypothetical protein	
	PGN_1231	hypothetical protein	
	PGN_1237	hypothetical protein	
	PGN_1241	hypothetical protein	
	PGN_1242	conserved hypothetical protein	*
	PGN_1246	hypothetical protein	
	PGN_1247	hypothetical protein	
	PGN_1248	hypothetical protein	
	PGN_1249	conserved hypothetical protein	
	PGN_1250	hypothetical protein	
	PGN_1253	hypothetical protein	
	PGN_1255	putative heptosyltransferase	*
	PGN_1257	hypothetical protein	
	PGN_1265	hypothetical protein	
	PGN_1266	conserved hypothetical protein	*
	PGN_1268	putative oxidoreductase	*
	PGN_1274	conserved hypothetical protein	*
	PGN_1277	conserved hypothetical protein	*
CTnPg1-b	PGN_1281	putative conserved protein found in conjugative transposon, TraM	*
	PGN_1282	conserved protein found in conjugative transposon, TraN	*
	PGN_1283	conserved protein found in conjugative transposon, TraO	*
	PGN_1284	putative DNA primase involved in conjugative transposon, TraP	*
	PGN_1285	conserved protein found in conjugative transposon, TraQ	*
	PGN_1286	probable lysozyme	*
	PGN_1287	conserved hypothetical protein	
	PGN_1288	conserved hypothetical protein	
	PGN_1289	conserved hypothetical protein	
	PGN_1290	conserved hypothetical protein	
	PGN_1291	conserved hypothetical protein related to phage	
	PGN_1292	probable anti-restriction protein	
	PGN_1293	conserved hypothetical protein	
	PGN_1294	conserved hypothetical protein	
	PGN_1295	hypothetical protein	
	PGN_1299	hypothetical protein	
	PGN_1306	hypothetical protein	
	PGN_1307	hypothetical protein	
	PGN_1320	conserved hypothetical protein	*
	PGN_1323	TPR domain protein	*
	PGN_1324	probable ABC transporter membrane protein	
	PGN_1325	probable ABC transporter membrane protein	
	PGN_1326	hypothetical protein	
	PGN_1327	hypothetical protein	*
	PGN_1328	hypothetical protein	
	PGN_1329	hypothetical protein	*
	PGN_1330	probable branched chain amino acid ABC transporter ATP-binding protein	*
	PGN_1331	conserved hypothetical protein	*
	PGN_1337	conserved hypothetical protein	*
	PGN_1340	conserved hypothetical protein	*
	PGN_1351	conserved hypothetical protein	*
	PGN_1353	conserved hypothetical protein	*
	PGN_1354	conserved hypothetical protein	*
	PGN_1357	conserved hypothetical protein	*
	PGN_1365	conserved hypothetical protein	*

PGN_1371	hypothetical protein	
PGN_1379	hypothetical protein	
PGN_1380	hypothetical protein	
PGN_1384	probable 1-acyl-sn-glycerol-3-phosphate acetyltransferase	*
PGN_1385	hypothetical protein	
PGN_1386	conserved hypothetical protein	*
PGN_1406	hypothetical protein	
PGN_1411	putative N-ethylammeline chlorohydrolase	*
PGN_1413	conserved hypothetical protein	*
PGN_1414	hypothetical protein	
PGN_1421	hypothetical protein	
PGN_1422	putative partial DNA-binding protein, histone-like family	*
PGN_1423	putative partial DNA-binding protein, histone-like family	*
PGN_1424	putative partial Type II restriction enzyme	
PGN_1425	putative partial Type II restriction enzyme	
PGN_1426	adenine-specific methyltransferase	
PGN_1427	hypothetical protein	
PGN_1438	hypothetical protein	
PGN_1439	hypothetical protein	
PGN_1465	hypothetical protein	
PGN_1473	hypothetical protein	
PGN_1477	conserved hypothetical protein	*
PGN_1483	conserved hypothetical protein	*
PGN_1502	conserved hypothetical protein	*
PGN_1507	hypothetical protein	
PGN_1508	hypothetical protein	
PGN_1515	conserved hypothetical protein	*
PGN_1523	putative polysaccharide export outer membrane protein	*
PGN_1524	conserved hypothetical protein	*
PGN_1531	hypothetical protein	
PGN_1532	conserved hypothetical protein	*
PGN_1533	putative carbonic anhydrase	
PGN_1534	hypothetical protein	
PGN_1535	hypothetical protein	
PGN_1536	hypothetical protein	
PGN_1537	probable cation efflux system protein	*
PGN_1538	putative cation efflux system	*
PGN_1539	putative ABC transport system exported protein	*
PGN_1540	putative ABC transport membrane protein	*
PGN_1541	hypothetical protein	
PGN_1544	conserved hypothetical protein	*
PGN_1545	hypothetical protein	
PGN_1558	conserved hypothetical protein	*
PGN_1559	hypothetical protein	
PGN_1560	conserved hypothetical protein	*
PGN_1561	conserved hypothetical protein	*
PGN_1563	hypothetical protein	
PGN_1577	putative preprotein translocase SecE subunit	*
PGN_1583	hypothetical protein	
PGN_1586	hypothetical protein	
PGN_1592	conserved hypothetical protein	*
PGN_1607	hypothetical protein	
PGN_1609	hypothetical protein	
PGN_1620	hypothetical protein	
PGN_1621	conserved hypothetical protein	*
PGN_1627	probable 4-amino-4-deoxy-L-arabinose transferase	
PGN_1629	conserved hypothetical protein with integral membrane domain DUF6	*
PGN_1632	hypothetical protein	
PGN_1643	conserved hypothetical protein	*
PGN_1648	putative 50S ribosomal protein L21	*

PGN_1660	conserved hypothetical protein	*
PGN_1663	hypothetical protein	
PGN_1664	conserved hypothetical protein	*
PGN_1665	hypothetical protein	
PGN_1675	conserved hypothetical protein	*
PGN_1684	hypothetical protein	
PGN_1686	hypothetical protein	
PGN_1687	conserved hypothetical protein	*
PGN_1720	hypothetical protein	
PGN_1728	lysine-specific cysteine proteinase Kgp	*
PGN_1732	hypothetical protein	*
PGN_1733	hemagglutinin protein HagA	*
PGN_1740	putative RNA polymerase ECF-type sigma factor	*
PGN_1746	cytochrome c nitrite reductase catalytic subunit NrfA	*
PGN_1747	conserved hypothetical protein	*
PGN_1754	hypothetical protein	
PGN_1765	hypothetical protein	
PGN_1766	conserved hypothetical protein	*
PGN_1769	hypothetical protein	
PGN_1787	probable 5-formyltetrahydrofolate cyclo-ligase	*
PGN_1795	conserved hypothetical protein	*
PGN_1796	conserved hypothetical protein	*
PGN_1799	hypothetical protein	*
PGN_1809	hypothetical protein	
PGN_1810	hypothetical protein	
PGN_1816	conserved hypothetical protein	*
PGN_1817	conserved hypothetical protein	*
PGN_1818	conserved hypothetical protein	*
PGN_1819	hypothetical protein	
PGN_1820	conserved hypothetical protein	*
PGN_1821	conserved hypothetical protein	*
PGN_1822	hypothetical protein	
PGN_1824	hypothetical protein	*
PGN_1825	hypothetical protein	*
PGN_1863	50S ribosomal protein L22	*
PGN_1879	hypothetical protein	
PGN_1894	conserved hypothetical protein	*
PGN_1897	putative transport related membrane protein	*
PGN_1898	probable transport protein	*
PGN_1900	hypothetical protein	
PGN_1902	conserved hypothetical protein	
PGN_1903	putative adenine-specific DNA methyltransferase	
PGN_1908	hypothetical protein	
PGN_1909	conserved hypothetical protein	*
PGN_1912	partial transposase in IS*Pg6*	*
PGN_1915	conserved hypothetical protein	*
PGN_1923	hypothetical protein	
PGN_1925	conserved hypothetical protein	*
PGN_1927	conserved hypothetical protein	*
PGN_1928	conserved hypothetical protein	*
PGN_1929	conserved hypothetical protein	*
PGN_1931	conserved hypothetical protein	*
PGN_1932	conserved hypothetical protein	*
PGN_1933	conserved hypothetical protein	*
PGN_1934	conserved hypothetical protein	*
PGN_1943	putative polyprenyl synthetase	*
PGN_1950	hypothetical protein	
PGN_1952	hypothetical protein	
PGN_1956	hypothetical protein	
PGN_1962	conserved hypothetical protein	

PGN_1963	conserved hypothetical protein	
PGN_1964	probable CRISPR-associated helicase Cas3 core	*
PGN_1965	hypothetical protein	
PGN_1971	hypothetical protein	
PGN_1972	conserved hypothetical protein	
PGN_1978	conserved hypothetical protein	*
PGN_1985	probable N-acetylmuramoyl-L-alanine amidase	*
PGN_1988	conserved hypothetical protein	*
PGN_2000	hypothetical protein	
PGN_2002	conserved hypothetical protein	*
PGN_2014	cation efflux system protein	*
PGN_2027	conserved hypothetical protein	*
PGN_2032	putative alpha-galactosidase	*
PGN_2036	hypothetical protein	
PGN_2046	conserved hypothetical protein	*
PGN_2047	conserved hypothetical protein	*
PGN_2063	hypothetical protein	
PGN_2084	conserved hypothetical protein	*

* these CDSs show significant sequence similarities to CDSs on the W83 chromosome, but do not fulfill the criterion for the ‘conserved CDSs’ that we defined in the present study (see the Materials and Methods section in the main text).

Supplementary Table 6. W83-specific CDSs

	PG No.	product	
	PG0007	hypothetical protein	
	PG0008	IS*Pg5* transposase Orf2	*
	PG0009	IS*Pg5* transposase Orf1	
	PG0019	IS*Pg4* transposase	*
	PGa0003	hypothetical protein	*
	PG0040	IS*Pg5* transposase Orf1	
	PG0041	IS*Pg5* transposase Orf2	*
	PG0050	IS*Pg4*, transposase	*
	PG0053	hypothetical protein	*
	PG0065	efflux transporter, RND family, MFP subunit	*
	PG0068	hypothetical protein	
	PG0080	hypothetical protein	*
	PG0085	alpha-galactosidase	*
	PGa0007	hypothetical protein similar to PG2076	*
	PG0101	hypothetical protein	*
	PGa0008	hypothetical protein similar to PG1787	*
GP1 locus	PG0109	hypothetical protein	
	PG0110	glycosyl transferase, group 1 family protein	
	PG0111	capsular polysaccharide biosynthesis gene, putative	*
	PGa0009	conserved hypothetical protein	*
	PG0113	conserved domain protein	
	PG0114	hypothetical protein	
	PG0116	conserved hypothetical protein	*
	PG0117	polysaccharide transport protein, putative	*
	PG0118	glycosyl transferase, group 2 family protein	*
	PG0123	hypothetical protein	*
	PG0132	hypothetical protein	
	PG0145	hypothetical protein	
	PG0161	hypothetical protein	*
	PGa0012	hypothetical protein	*
	PG0174	pyridine nucleotide-disulphide oxidoreductase family protein	*
	PGa0013	partial Mfa1 fimbrilin	*
	PG0177	IS*Pg4*, transposase	*
	PGa0014	partial Mfa1 fimbrilin	*
	PG0183	lipoprotein, putative	*
	PG0185	RagA protein	*
	PG0186	lipoprotein RagB	*
	PGa0015	IS*Pg6*, transposase, partial	*
	PGa0016	IS*Pg6*, transposase, partial	*
	PG0188	lipoprotein, putative	*
	PG0197	hypothetical protein	
	PGa0017	hypothetical protein similar to PG1426	*
	PG0198	conserved hypothetical protein	*
	PG0218	hypothetical protein	*
	PG0219	conserved domain protein	
	PG0221	hypothetical protein	
	PG0225	IS*Pg4*, transposase	*
	PG0240	hydrolase, haloacid dehalogenase-like family	*
	PG0246	hypothetical protein	*
	PG0274	hypothetical protein	
	PG0278	hypothetical protein	*
	PG0289	hypothetical protein	*

PG0297	hypothetical protein	*
PGa0021	hypothetical protein similar to PG0297	*
PG0302	hypothetical protein	*
PG0314	ribosomal protein L21	*
PG0319	hypothetical protein	*
PG0333	membrane protein, putative	*
PG0340	hypothetical protein	*
PG0354	hypothetical protein	
PG0371	hypothetical protein	
PG0374	hypothetical protein	
PG0382	conserved hypothetical protein	
PGa0023	putative preprotein translocase SecE subunit	*
PG0404	hypothetical protein	*
PG0408	hypothetical protein	*
PG0410	hypothetical protein	
PG0422	hypothetical protein	
PG0423	hypothetical protein	*
PG0426	IS*Pg5*, transposase Orf2	*
PG0427	IS*Pg5*, transposase Orf1	
PG0428	hypothetical protein	
PG0431	hypothetical protein	
PG0436	capsular polysaccharide transport protein, putative	*
PG0437	polysaccharide export protein, BexD/CtrA/VexA family	*
PG0442	hypothetical protein	
PG0447	conserved hypothetical protein	*
PG0456	PHP N-terminal domain protein	
PG0457	hypothetical protein	
PG0458	IS*Pg5*, transposase Orf2	*
PG0459	IS*Pg5*, transposase Orf1	*
PG0461	IS*Pg7*, transposase	
PG0466	hypothetical protein	*
PG0487	IS*Pg4*, transposase	*
PG0492	hypothetical protein	
PG0493	hypothetical protein	*
PG0494	hypothetical protein	*
PG0496	hypothetical protein	
PG0499	hypothetical protein	*
PG0507	hypothetical protein	
PG0536	hypothetical protein	
PG0542	hypothetical protein	
PG0543	transcriptional regulator, putative	*
PG0544	type I restriction-modification system, M subunit, putative	
PG0545	hypothetical protein	
PG0546	conserved domain protein	*
PG0554	hypothetical protein	
PG0556	hypothetical protein	
PG0557	hypothetical protein	*
PG0559	chlorohydrolase family protein	*
PG0563	hypothetical protein	
PG0564	hypothetical protein	
PG0565	hypothetical protein	
PG0590	IS*Pg5*, transposase Orf1	*
PG0591	IS*Pg5*, transposase Orf2	*
PG0605	hypothetical protein	*
PG0608	hypothetical protein	

	PG0609	hypothetical protein	
	PG0610	hypothetical protein	*
	PG0611	lipoprotein, putative	*
	PG0614	hypothetical protein	*
	PG0617	hypothetical protein	*
	PG0626	hypothetical protein	
	PGa0028	tonB-linked receptor Tlr, frame-shifted with PGN_0683	*
	PGa0029	tonB-linked receptor Tlr, frame-shifted with PGN_0683	*
	PG0655	hypothetical protein	
	PG0661	hypothetical protein	
	PGa0030	conserved hypothetical protein frame-shifted with PGN_0702	*
	PGa0031	conserved hypothetical protein with a phosphotransferase enzyme family domain, frame-shifted with PGN_0702	*
	PG0681	hypothetical protein	*
	PG0683	ABC transporter, permease protein, putative	*
	PG0700	hypothetical protein	*
	PG0717	lipoprotein, putative	*
	PG0718	conserved hypothetical protein	
	PG0719	sensor histidine kinase	
	PG0722	hypothetical protein	*
	PG0723	hypothetical protein	
	PG0727	hypothetical protein	
	PG0741	conserved hypothetical protein	
	PG0742	antigen PgaA	
	PG0767	4-alpha-glucanotransferase	*
	PG0769	fibronectin type III domain protein	*
	PG0770	hypothetical protein	*
	PG0771	hypothetical protein	
	PG0774	hypothetical protein	*
	PG0786	hypothetical protein	
	PG0810	hypothetical protein	*
large mobile element region I	PG0814	hypothetical protein	
	PG0815	hypothetical protein	
	PG0816	hypothetical protein	*
	PG0819	integrase	*
	PG0820	integrase	*
	PG0821	lipoprotein, putative	*
	PG0822	hypothetical protein	
	PG0823	conserved hypothetical protein	
	PG0824	conserved hypothetical protein	
	PG0826	transcriptional regulator, AraC family	*
	PG0829	hypothetical protein	*
	PG0831	hypothetical protein	*
	PG0832	hypothetical protein	
	PG0833	conserved hypothetical protein	
	PG0834	hypothetical protein	
	PG0835	hypothetical protein	
	PGa0043	integrase	*
	PG0838	integrase	*
	PG0839	conserved hypothetical protein	
	PG0840	hypothetical protein	
	PG0841	mobilizable transposon, excision protein, putative	*
	PG0842	mobilizable transposon, hypothetical protein, putative	*
	PG0843	hypothetical protein	
	PG0844	hypothetical protein	

PGa0046	probable integrase	*
PG0847	conserved domain protein	*
PG0848	hypothetical protein	
PG0849	hypothetical protein	
PG0850	DNA binding protein, excisionase family, putative	
PG0851	conserved hypothetical protein	*
PG0853	DNA-binding protein, histone-like family	*
PG0854	hypothetical protein	
PG0855	hypothetical protein	
PG0856	hypothetical protein	
PG0857	transcriptional regulator, putative	
PG0858	conserved hypothetical protein	
PG0859	conserved hypothetical protein	
PG0860	transcriptional regulator, putative	
PG0861	helicase, SNF2/RAD54 family	*
PG0862	type IIS restriction endonuclease, putative	
PG0864	site-specific recombinase, resolvase family	
PG0866	hypothetical protein	
PG0867	hypothetical protein	
PG0868	mobilization protein	*
PG0869	mobilization protein	*
PG0870	conserved hypothetical protein	
PG0871	hypothetical protein	
PG0872	mobilizable transposon, Xis protein	
PG0873	mobilizable transposon, TnpC protein	
PG0874	mobilizable transposon, Int protein	*
PG0875	mobilizable transposon, TnpA protein	
PG0877	hypothetical protein	*
PG0879	hypothetical protein	*
PG0880	bacterioferritin comigratory protein	*
PG0888	hypothetical protein	
PG0915	conserved hypothetical protein	*
PG0918	hypothetical protein	*
PG0929	hypothetical protein	
PGa0049	hypothetical protein	*
PG0938	calcium-transporting ATPase	*
PG0942	IS*Pg5*, transposase Orf1	*
PG0943	IS*Pg5*, transposase Orf2	*
PG0954	TPR domain protein	*
PG0955	hypothetical protein	*
PG0958	ribonuclease BN, putative	*
PG0970	IS*Pg4*, transposase	*
PG0971	McrBC restriction endonuclease system, McrB subunit, putative	
PG0972	conserved hypothetical protein	
PG0979	hypothetical protein	
PG0982	TPR domain protein	*
PG0984	hypothetical protein	
PG0994	hypothetical protein	
PG0997	transcriptional regulator, putative	*
PG1000	hypothetical protein	*
PG1002	hypothetical protein	*
PG1014	TPR domain protein	*
PG1018	hypothetical protein	*
PG1021	hypothetical protein	
PG1024	hypothetical protein	

PG1025	conserved domain protein	
PG1026	hypothetical protein	
PG1027	hypothetical protein	
PG1028	TPR domain protein	*
PG1033	conserved hypothetical protein	*
PG1055	thiol protease	
PG1059	hypothetical protein	*
PG1061	IS*Pg6*, transposase	*
PG1065	dihydroorotate dehydrogenase	*
PG1071	conserved hypothetical protein	*
PG1083	hypothetical protein	*
PG1102	hypothetical protein	*
PG1107	hypothetical protein	
PG1108	hypothetical protein	
PG1109	mobilization protein	
PG1110	hypothetical protein	
PGa0058	conserved hypothetical protein	
PGa0059	conserved hypothetical protein	
PG1112	hypothetical protein	
PG1113	integrase	*
PG1130	TPR domain protein	*
PG1142	exopolysaccharide synthesis-related protein	*
PG1148	hypothetical protein	
PG1150	hypothetical protein	
PG1152	hypothetical protein	
PG1153	hypothetical protein	
PG1154	hypothetical protein	
PG1155	ADP-heptose--LPS heptosyltransferase, putative	*
PG1166	hypothetical protein	
PG1167	hypothetical protein	
PG1169	hypothetical protein	*
PG1171	oxidoreductase, putative	*
PG1196	hypothetical protein	*
PG1198	hypothetical protein	*
PG1199	hypothetical protein	*
PG1200	hypothetical protein	
PGa0065	putative transcriptional regulator	*
PG1202	hypothetical protein	
PG1203	transcriptional regulator, putative	*
PG1205	DNA-binding protein, histone-like family	*
PG1206	hypothetical protein	
PG1207	hypothetical protein	
PG1215	lipoprotein protein, putative	*
PG1218	hypothetical protein	*
PG1222	hypothetical protein	
PG1229	hypothetical protein	*
PG1249	1-acyl-sn-glycerol-3-phosphate acetyltransferase, putative	*
PG1251	hypothetical protein	*
PG1250	hypothetical protein	
PG1261	IS*Pg4*, transposase	*
PG1266	hypothetical protein	*
PG1268	hypothetical protein	
PG1311	conserved hypothetical protein	*
PG1325	hypothetical protein	*
PG1332	NAD(P) transhydrogenase, beta subunit	*

	PG1341	hypothetical protein	*
	PG1356	hypothetical protein	*
	PG1363	hypothetical protein	*
	PG1373	hypothetical protein	
	PG1388	hypothetical protein	*
	PG1398	hypothetical protein	*
	PG1402	AP endonuclease domain protein	*
	PG1419	IS*Pg5*, transposase Orf1	*
	PG1420	IS*Pg5*, transposase Orf2	*
	PG1429	hypothetical protein	
	PG1433	hydrolase	*
large mobile element region II	PG1435	integrase	*
	PG1436	ATPase, putative	
	PG1439	hypothetical protein	
	PG1440	hypothetical protein	
	PG1441	lysozyme-related protein	*
	PG1442	hypothetical protein	*
	PG1444	hypothetical protein	*
	PG1447	transcriptional regulator, AraC family	*
	PG1449	conserved hypothetical protein	
	PG1450	conserved hypothetical protein	
	PG1451	conserved hypothetical protein	
	PG1452	lipoprotein, putative	*
	PG1453	integrase	*
	PG1454	integrase	*
	PG1457	hypothetical protein	*
	PG1459	hypothetical protein	
	PG1465	hypothetical protein	
	PG1466	hypothetical protein	
	PG1467	methlytransferase, UbiE/COQ5 family	
	PG1469	type I restriction-modification system, M subunit, putative	
	PG1470	hypothetical protein	*
	PG1471	conserved hypothetical protein	
	PG1472	hypothetical protein	
	PG1473	conjugative transposon protein TraQ	*
	PG1474	conjugative transposon protein TraO	*
	PG1475	conjugative transposon protein TraN	*
	PG1476	conjugative transposon protein TraM	*
	PG1477	hypothetical protein	
	PG1478	conjugative transposon protein TraK	*
	PG1479	conjugative transposon protein TraJ	*
	PG1480	conjugative transposon protein TraI	*
	PG1481	conjugative transposon protein TraG	*
	PG1482	conjugative transposon protein TraF	*
	PG1483	conjugative transposon protein TraE	*
	PG1484	hypothetical protein	
	PG1485	conjugative transposon protein TraC	
	PG1486	conjugative transposon protein TraA	*
	PG1487	hypothetical protein	
	PG1488	hypothetical protein	*
	PG1489	conserved hypothetical protein	*
	PG1490	TraG family protein	*
	PG1494	conserved hypothetical protein	*
	PG1495	DNA topoisomerase III	*
	PG1496	hypothetical protein	*

PG1497	DNA-binding protein, histone-like family	*
PG1498	hypothetical protein	
PG1499	hypothetical protein	
PG1500	conserved domain protein	*
PG1501	transcriptional regulator, TetR family	
PG1503	LytB-related protein	
PG1504	NAD dependent protein	*
PG1505	radical SAM domain protein	
PG1507	hypothetical protein	
PG1508	hypothetical protein	
PG1509	HAD-superfamily hydrolase, subfamily IA, variant 1 family protein	
PG1510	hypothetical protein	
PG1511	hypothetical protein	
PG1512	conserved domain protein	*
PG1513	phosphoribosyltransferase, putative/phosphoglycerate mutase family protein	
PG1514	glycerol dehydrogenase-related protein	
PG1515	ribulose bisphosphate carboxylase-related protein	
PG1516	hypothetical protein	
PGa0078	IS*Pg3*, transposase, partial	*
PG1519	hypothetical protein	*
PG1521	O-succinylbenzoic acid--CoA ligase	
PG1522	mandelate racemase/muconate lactonizing enzyme family protein	
PG1523	naphthoate synthase	*
PG1524	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase	
PG1525	isochorismate synthase, putative	*
PG1526	conserved hypothetical protein	*
PG1527	hypothetical protein	
PG1528	conserved hypothetical protein	
PG1529	hypothetical protein	
PG1530	conserved domain protein	
PG1532	hypothetical protein	
PG1533	Toprim domain protein	*
PG1534	conserved domain protein	

PG1537	conserved hypothetical protein	*
PG1546	hypothetical protein	
PG1551	HmuY protein	*
PG1573	transcriptional regulator, Crp family	*
PG1574	hypothetical protein	*
PG1591	conserved hypothetical protein	
PG1617	hypothetical protein	*
PG1630	hypothetical protein	*
PG1638	thioredoxin family protein	*
PG1644	IS*Pg5*, transposase Orf2	*
PG1645	IS*Pg5*, transposase Orf1	*
PG1649	hypothetical protein	
PG1655	hypothetical protein	*
PG1658	IS*Pg4*, transposase	*
PG1661	hypothetical protein	
PG1662	hypothetical protein	*
PG1671	hypothetical protein	*
PG1673	IS*Pg4*, transposase	*
PGa0084	probable partial hemagglutinin-related protein	*
PG1675	hypothetical protein	

PGa0085	putative ABC-type transporter ATP-binding protein	*
PG1685	hypothetical protein	
PG1695	hypothetical protein	*
PG1709	IS*Pg5*, transposase Orf1	*
PG1710	IS*Pg5*, transposase Orf2	*
PGa0087	putative partial transcriptional repressor	*
PGa0088	putative partial transcriptional repressor	*
PG1738	hypothetical protein	*
PG1739	conserved domain protein	*
PG1742	hypothetical protein	
PG1786	hypothetical protein	*
PG1787	hypothetical protein	*
PG1799	hypothetical protein	*
PG1811	hypothetical protein	
PG1818	hypothetical protein	*
PG1819	hypothetical protein	*
PG1820	cytochrome c nitrite reductase, catalytic subunit NrfA	*
PG1827	RNA polymerase sigma-70 factor, ECF subfamily	*
PG1837	hemagglutinin protein HagA	*
PGa0090	lysine-specific cysteine proteinase	*
PG1854	5-formyltetrahydrofolate cyclo-ligase family protein	*
PG1862	hypothetical protein	*
PG1863	hypothetical protein	*
PG1864	leucine-rich protein	*
PG1866	hypothetical protein	
PG1869	hypothetical protein	
PG1871	hypothetical protein	*
PG1889	hypothetical protein	*
PG1890	lipoprotein, putative	*
PG1891	hypothetical protein	*
PGa0093	hypothetical protein	*
PG1892	hypothetical protein	*
PG1894	hypothetical protein	
PG1933	ribosomal protein L22	*
PGa0101	partial chloride channel protein	*
PG1966	conserved hypothetical protein	*
PG1968	hypothetical protein	
PG1969	hypothetical protein	*
PG1970	hypothetical protein	
PG1978	hypothetical protein	
PG1979	hypothetical protein	*
PG1982	CRISPR-associated protein Cas1	*
PG1983	CRISPR-associated protein, TM1791 family	*
PG1984	hypothetical protein	*
PG1986	CRISPR-associated protein, TM1793 family	*
PG1987	CRISPR-associated protein, TM1811 family	*
PG1988	hypothetical protein	*
PG1989	hypothetical protein	*
PG1998	polyprenyl synthetase	*
PG2016	CRISPR-associated helicase Cas3	*
PG2017	hypothetical protein	
PG2018	hypothetical protein	
PG2019	hypothetical protein	
PG2031	hypothetical protein	*
PG2038	N-acetylmuramoyl-L-alanine amidase, putative	*

PG2042	thioredoxin family protein	*
PGa0105	IS*Pg2*, transposase, partial	*
PG2057	IS*Pg5*, transposase Orf2	*
PG2058	IS*Pg5*, transposase Orf1	*
PG2075	hypothetical protein	*
PG2076	hypothetical protein	*
PG2098	hypothetical protein	
PG2100	immunoreactive 63 kDa antigen PG102	*
PG2104	hypothetical protein	
PG2105	lipoprotein, putative	*
PG2111	ThiS protein	*
PG2112	hypothetical protein	*
PGa0108	hypothetical protein	*
PG2114	hypothetical protein	*
PG2128	IS*Pg5*, transposase Orf2	*
PG2129	IS*Pg5*, transposase Orf1	*
PG2132	fimbrilin	*
PG2133	lipoprotein, putative	*
PG2134	lipoprotein, putative	*
PG2135	lipoprotein, putative	*
PG2136	hypothetical protein	*
PG2139	hypothetical protein	*
PG2152	DNA-binding protein, histone-like family	*
PG2166	hypothetical protein	
PG2186	transcriptional regulator, putative	*
PG2194	IS*Pg4*, transposase	*
PG2208	hypothetical protein	*
PG2220	hypothetical protein	*

* these CDSs show significant sequence similarities to CDSs on the ATCC 33277 chromosome, but do not fulfill the criterion for the ‘conserved CDSs’ that we defined in the present study (see the Materials and Methods section in the main text).

Supplementary Table 7. IS and MITEs of W83 genome in this study

insertion sequence	start	stop	strand	annotation[a]
IS*Pg1*_01 (partial)	62287	62556	+	
IS*Pg1*_02 (partial)	62562	63281	+	IS1.1A
IS*Pg1*_03	216437	217774	+	IS1.2
IS*Pg1*_04 (partial)	298868	299174	-	
IS*Pg1*_05	501972	503309	+	IS1.3
IS*Pg1*_06	605930	607267	-	IS1.4
IS*Pg1*_07 (partial)	811059	811778	-	IS1.5
IS*Pg1*_08 (partial)	811784	812245	-	IS1.5
IS*Pg1*_09 (partial)	814725	815186	+	IS1.6
IS*Pg1*_10 (partial)	815192	815911	+	IS1.6
IS*Pg1*_11 (partial)	873736	873915	-	
IS*Pg1*_12 (partial)	873938	874162	-	
IS*Pg1*_13 (partial)	874251	874751	-	
IS*Pg1*_14 (partial)	874839	875693	-	
IS*Pg1*_15	882006	883343	-	IS1.7
IS*Pg1*_16 (partial)	905767	906634	-	IS1.8
IS*Pg1*_17 (partial)	906640	907101	-	IS1.8
IS*Pg1*_18 (partial)	912908	913369	+	IS1.9
IS*Pg1*_19 (partial)	913375	914094	+	IS1.9
IS*Pg1*_20 (partial)	1000353	1001219	-	IS1.10
IS*Pg1*_21 (partial)	1001225	1001686	-	IS1.10
IS*Pg1*_22 (partial)	1004290	1005057	-	
IS*Pg1*_23 (partial)	1049243	1049668	-	
IS*Pg1*_24 (partial)	1049757	1050557	-	
IS*Pg1*_25	1093024	1094361	-	IS1.11
IS*Pg1*_26	1257737	1259074	-	IS1.12
IS*Pg1*_27	1275418	1276755	-	IS1.13
IS*Pg1*_28 (partial)	1304317	1305183	-	IS1.14
IS*Pg1*_29 (partial)	1305189	1305650	-	IS1.14
IS*Pg1*_30	1319569	1320906	+	IS1.15
IS*Pg1*_31 (partial)	1398787	1399674	+	
IS*Pg1*_32 (partial)	1429734	1430600	-	IS1.16A
IS*Pg1*_33 (partial)	1430606	1430956	-	IS1.16A
IS*Pg1*_34 (partial)	1463773	1464234	+	IS1.17
IS*Pg1*_35 (partial)	1464240	1464959	+	IS1.17
IS*Pg1*_36 (partial)	1481478	1482344	-	IS1.18
IS*Pg1*_37 (partial)	1482350	1482811	-	IS1.18
IS*Pg1*_38	1532319	1533656	+	IS1.19
IS*Pg1*_39	1706131	1707468	-	IS1.20
IS*Pg1*_40 (partial)	1945179	1945640	+	
IS*Pg1*_41 (partial)	1945646	1945915	+	
IS*Pg1*_42	2005442	2006779	+	IS1.21
IS*Pg1*_43 (partial)	2009079	2010265	-	IS1.22
IS*Pg1*_44 (partial)	2101270	2101989	-	IS1.23
IS*Pg1*_45 (partial)	2101995	2102456	-	IS1.23
IS*Pg1*_46 (partial)	2156466	2156783	+	
IS*Pg1*_47 (partial)	2156789	2157508	+	
IS*Pg1*_48 (partial)	2274557	2274886	+	IS1.24A
IS*Pg1*_49 (partial)	2275269	2275988	+	IS1.24B
IS*Pg1*_50 (partial)	2299564	2300430	-	IS1.25
IS*Pg1*_51 (partial)	2300436	2300897	-	IS1.25
IS*Pg2*_01	308933	310139	+	IS2.1

IS*Pg2*_02	927843	929049	-	IS2.2
IS*Pg2*_03 (partial)	1269659	1269922	-	
IS*Pg2*_04	1430943	1432149	+	IS2.3
IS*Pg2*_05	1831177	1832383	-	IS2.4
IS*Pg2*_06 (partial)	2153955	2154260	-	
IS*Pg2*_07 (partial)	2229726	2230037	-	
IS*Pg2*_08	2283292	2284498	-	IS2.5
IS*Pg3*_01	230914	231983	+	IS3.1
IS*Pg3*_02	293939	295008	-	IS3.2
IS*Pg3*_03 (partial)	336207	336545	-	
IS*Pg3*_04	852831	853900	+	
IS*Pg3*_05 (partial)	1094473	1095388	+	IS3.4B
IS*Pg3*_06	1339809	1340878	-	IS3.5
IS*Pg3*_07 (partial)	1594936	1595556	-	
IS*Pg3*_08 (partial)	2004988	2005261	-	IS3.6A
IS*Pg3*_09 (partial)	2006695	2007400	-	IS3.6B
IS*Pg4*_01	22050	23443	+	IS4.1
IS*Pg4*_02	60926	62319	+	IS4.2
IS*Pg4*_03	199096	200490	+	IS4.3
IS*Pg4*_04	258469	259863	-	IS4.4
IS*Pg4*_05	526260	527654	-	IS4.5
IS*Pg4*_06	1029615	1031009	+	IS4.6
IS*Pg4*_07	1338381	1339774	+	IS4.7
IS*Pg4*_08	1741731	1743125	-	IS4.8
IS*Pg4*_09	1758398	1759791	+	IS4.9
IS*Pg4*_10	2300946	2302339	-	IS4.10
IS*Pg5*_01	7070	8583	-	IS5.1
IS*Pg5*_02	48047	49560	+	IS5.2
IS*Pg5*_03	464144	465658	-	IS5.3
IS*Pg5*_04 (partial)	500509	501984	-	IS5.4A
IS*Pg5*_05	648155	649666	+	IS5.5
IS*Pg5*_06	1002184	1003695	+	IS5.6
IS*Pg5*_07	1505117	1506630	+	IS5.7
IS*Pg5*_08	1724666	1726179	-	IS5.8
IS*Pg5*_09	1795028	1796541	+	IS5.9
IS*Pg5*_10	2154747	2156507	-	IS5.10
IS*Pg5*_11	2233852	2235365	-	IS5.11
IS*Pg6*_01 (partial)	223284	223466	-	
IS*Pg6*_02 (partial)	223576	224003	-	
IS*Pg6*_03	1128807	1129945	-	IS6.1
IS*Pg7*_01	503521	504948	-	IS7.1

MITEs	start	stop	strand	annotation[a]
MITE239_01	36482	36720	+	MITE 239.1
MITE239_02	1481023	1481261	-	MITE 239.2
MITE239_03	2154829	2155067	+	MITE 239.3
MITE239_04	2224966	2225204	-	MITE 239.4
MITE239_05	2274870	2275108	-	MITE 239.5
MITE*PgRS*_01 (partial)	105331	105419	+	
MITE*PgRS*_02	242476	243166	+	MITE 464.1

MITE*PgRS*_03	270233	270729	+	MITE 464.2
MITE*PgRS*_04 (partial)	389531	389640	+	
MITE*PgRS*_05	607401	608292	+	MITE 464.3
MITE*PgRS*_06 (partial)	627762	627886	-	
MITE*PgRS*_07	657919	658626	-	MITE 464.4
MITE*PgRS*_08	693460	694530	-	MITE 464.5
MITE*PgRS*_09	795445	795794	+	
MITE*PgRS*_10	1036691	1037352	+	MITE 464.6
MITE*PgRS*_11 (partial)	1237433	1238047	+	MITE 464.7
MITE*PgRS*_12	1264613	1265338	-	MITE 464.8
MITE*PgRS*_13 (partial)	1327845	1327901	+	
MITE*PgRS*_14	1371317	1372126	+	MITE 464.9
MITE*PgRS*_15	1391676	1392139	-	MITE 464.13
MITE*PgRS*_16	1419684	1420309	+	MITE 464.10
MITE*PgRS*_17 (partial)	1454999	1455105	+	
MITE*PgRS*_18 (partial)	1511717	1511774	+	
MITE*PgRS*_19	1710996	1711798	-	MITE 464.11
MITE*PgRS*20	1771774	1772019	+	
MITE*PgRS*_21	1785243	1785964	-	MITE 464.12
MITE700_01	275682	276358	+	MITE 700.1
MITE700_02	953354	954047	+	MITE 700.2
MITE700_03	1253850	1254255	+	
MITE700_04	1273283	1274000	-	MITE 700.3
MITE700_05 (partial)	1377896	1378504	+	MITE 700.4
MITE700_06	1751167	1751840	+	MITE 700.5
MITE700_07 (partial)	1936901	1937224	-	MITE 700.6
MITE700_08	1978319	1979006	+	MITE 700.7
MITE700_09	2045771	2046225	+	MITE 700.8

a, annotation in the genome sequence of W83 deposited in the GenBank database (accession number AE015924)

Supplementary Table 8. Quantification of the expression level of *kgp* and *glk* by real-time PCR

strain	*kgp**	*glk**
ATCC 33277	2.14E-12	1.66E-13
W83	1.11E-12	3.33E-13
TDC60	1.53E-12	2.26E-13
TDC117	2.51E-12	2.05E-13
TDC275	1.48E-12	2.17E-13
SU63	1.24E-12	2.64E-13
GAI7802	1.17E-12	2.37E-13

*, amount of mRNA relative to the 16S rRNA.