

Supplementary Information for:
Genetic diversity and low-level infection of Malawi
Polyomavirus in pediatric patients in Brazil

Thuany Giovana Pereira Daniel¹, Gabriel Montenegro de Campos^{1,2},
Tiago Souza Salles¹, Isabela Carvalho Brcko¹, James Siqueira
Pereira¹, Claudia Renata dos Santos Barros¹, Jardelina de Souza
Todao Bernardino¹, Joice Naiara Bertaglia Pereira¹, Mauricio
Cesar Sampaio Ando¹, Milene Silva Tino¹, Aryene Goes Trezena¹,
Vinícius Carius de Souza¹, Gabriela Marengone Altizani³, Lucio
Ayres Caldas⁴, Rodrigo Haddad⁵, Alex Ranieri Jerônimo Lima¹,
Marta Giovanetti^{6,7}, Sandra Coccuzzo Sampaio¹, Maria Carolina
Elias¹, Carlos Alberto Scrideli³, Svetoslav Naney Slavov^{1*}

¹Butantan Institute, São Paulo, Brazil.

²Postgraduation Program in Infectious Diseases and Global Health,
Faculty of Medicine, University of São Paulo (FMUSP), São Paulo, Brazil.

³Department of Puericulture and Pediatrics, Faculty of Medicine of
Ribeirão Preto (FMRP), University of São Paulo, Ribeirão Preto, Brazil.

⁴Precision Medicine Research Center, Institute of Biophysics, Federal
University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil.

⁵Campus of Ceilândia, University of Brasília (UnB), Brasília, Federal
District, Brazil.

⁶Department of Science and Bio-Technology, Università Campus
Bio-Medico di Roma, Rome, Italy.

⁷Rene Rachou Institute, Oswaldo Cruz Foundation (FIOCRUZ), Belo
Horizonte, Minas Gerais, Brazil.

Supplementary Figure 1

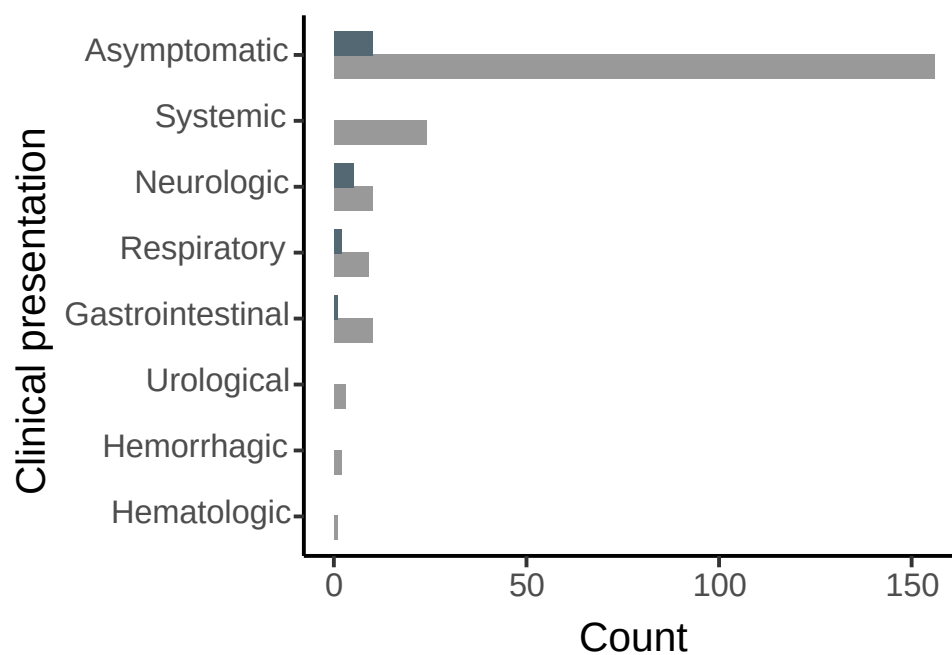


Fig. S1 Clinical manifestations of the positive and negative for MWPvV DNA pediatric patients. This figure represents solely descriptive distributions/frequencies of the clinical presentations of the positive and negative for MWPvV DNA pediatric patients. This figure is intended only to describe the characteristics of the studied population and do not represent statistical comparisons or hypothesis-testing analysis.

Supplementary Figure 2

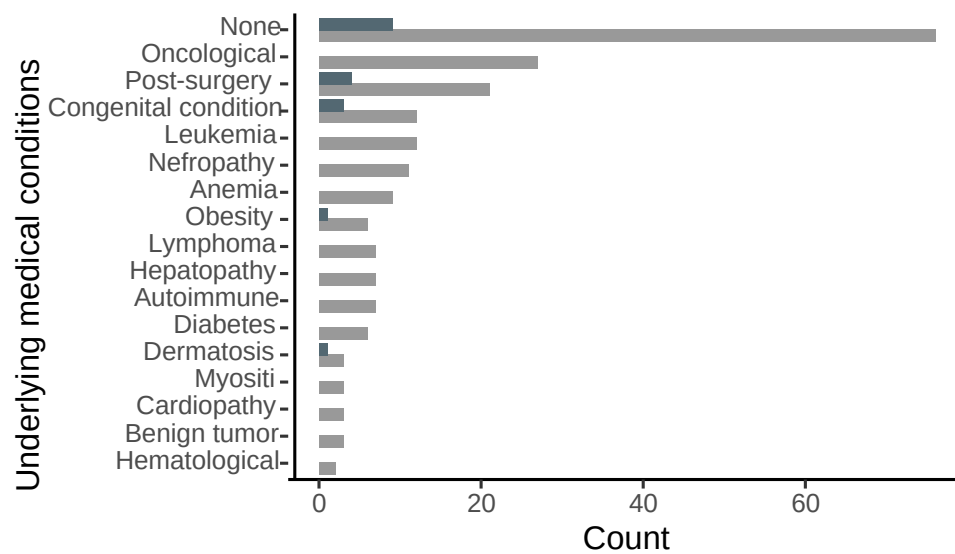


Fig. S2 Underlying clinical conditions of the positive and negative for MWPvV DNA pediatric patients. This figure represents solely descriptive distributions/frequencies of the underlying clinical conditions of the positive and negative for MWPvV DNA pediatric patients. This figure is intended only to describe the characteristics of the studied population and do not represent statistical comparisons or hypothesis-testing analysis.

SUPPLEMENTARY FIGURE 3

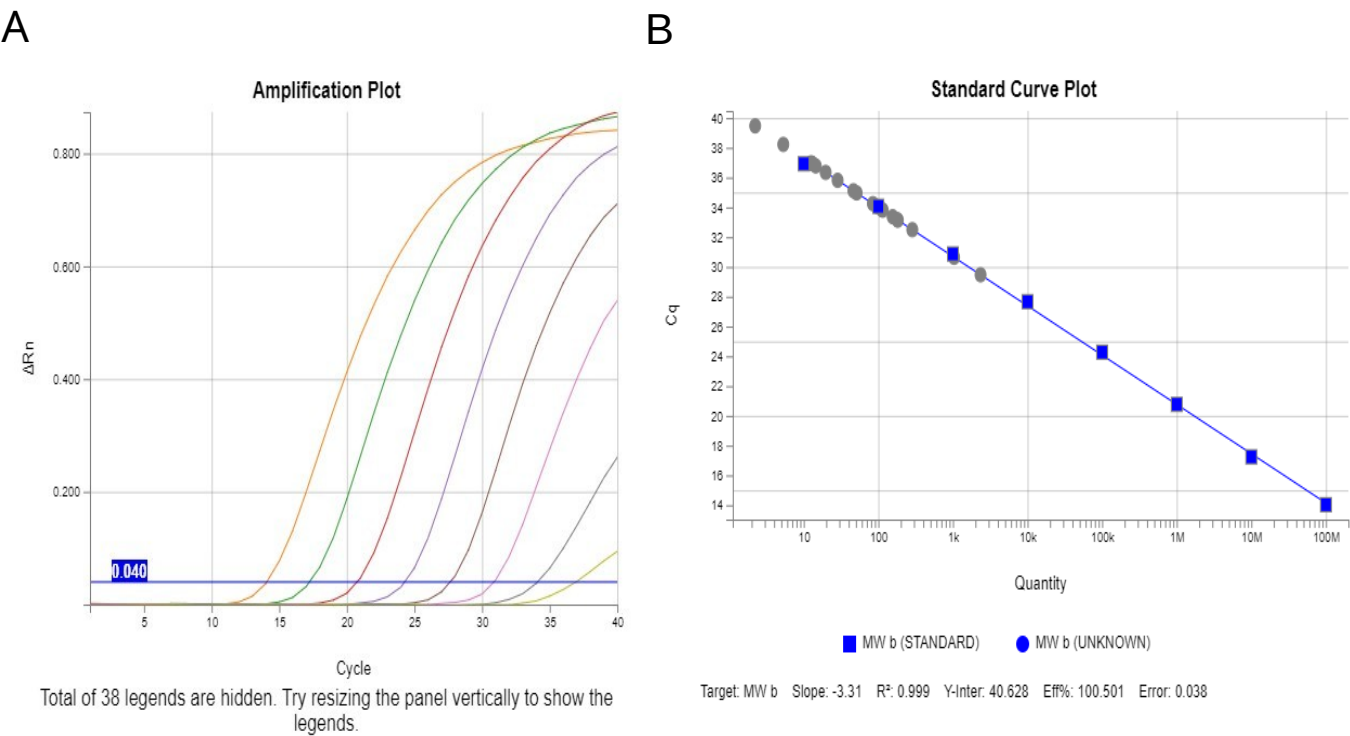


Fig. S3 Absolute quantification of MWPvV DNA in respiratory swab samples. **A.** Linear representation of the MWPvV standard curve generated from serial tenfold dilutions ranging from 108 to 10 copies per reaction, corresponding to Ct values of 14.0 and 36.92 respectively **B.** MWPvV standard curve illustrating the linear relationship between Ct values and viral copy number. MWPvV-positive samples are overlaid on the curve.

Supplementary Figure 4

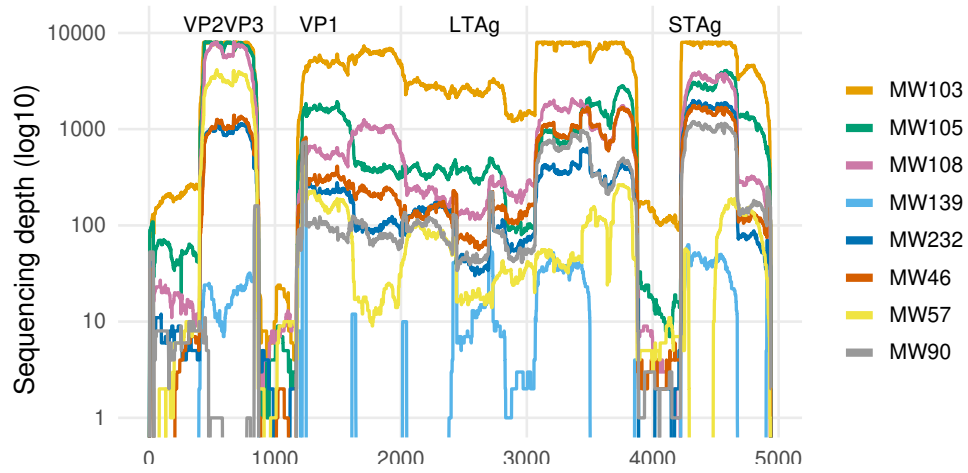


Fig. S4 Genome-wide sequencing depth of MWPyV-positive samples. Sequencing depth across the MWPyV reference genome is shown for each sample on log10 scale. Shaded regions represent annotated coding regions, including VP2, VP3, VP1, large T antigen (LTA), and small T antigen (STA). Samples retained for downstream analyses exhibited consistently high coverage across structural and regulatory regions, whereas in MWPyV39 and MW9PyV0 reflect incomplete genome recovery possibly due to the low viral load.

SUPPLEMENTARY FIGURE 5

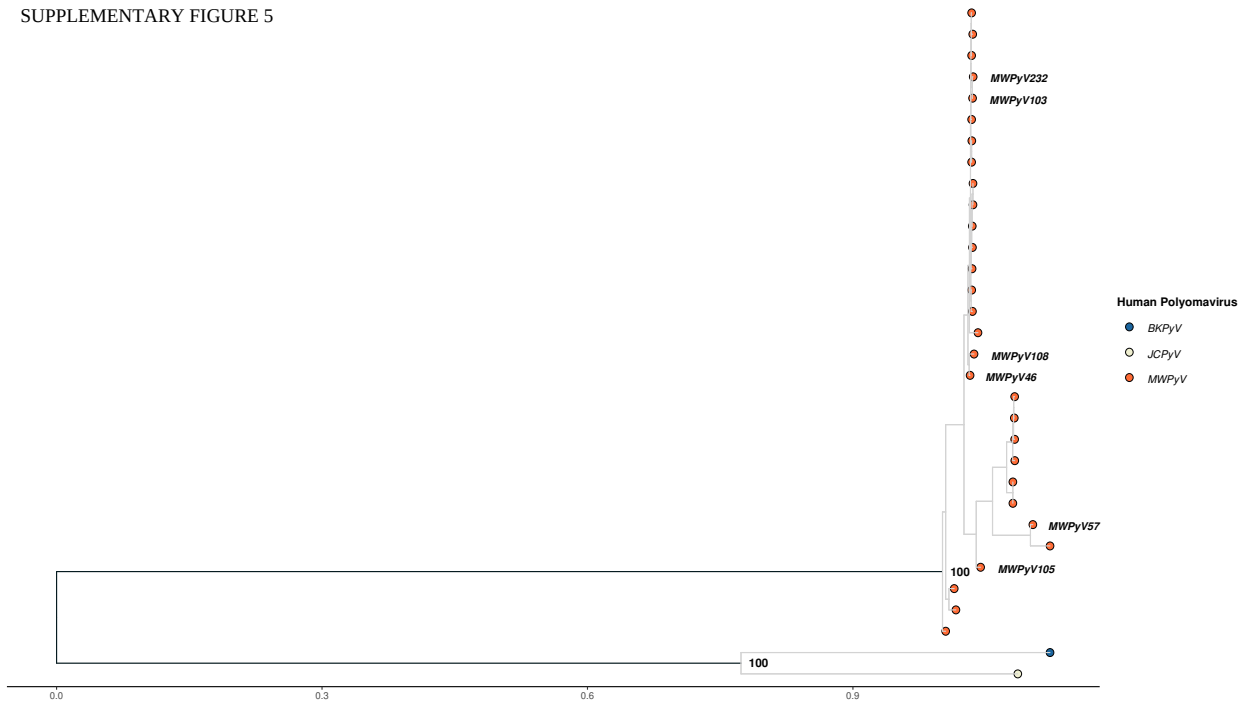


Fig. S5 Phylogenetic placement of MWPyV within the *Polyomaviridae* family, including BK polyomavirus (BKPyV) and JC polyomavirus (JCPyV) as outgroups. Branches are drawn to scale according to the number of substitutions per site. Ultrafast bootstrap support values are shown for key basal nodes. Tip colors indicate virus species, including MWPyV, BKPyV, and JCPyV.

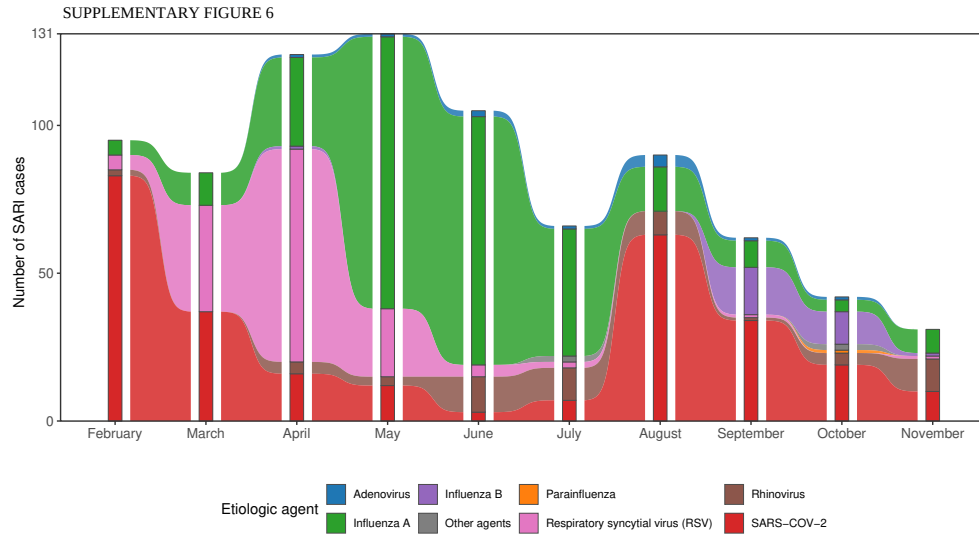


Fig. S6 Temporal distribution of severe acute respiratory infection (SARI) cases by etiological agents in the city of Ribeirão Preto, the monthly distribution of laboratory-confirmed respiratory pathogens among SARI cases from February to November 2024. Flow widths are proportional to the number of cases attributed to each virus per month. The visualization highlights temporal fluctuations and the changing relative contribution of each etiological agent over time, including SARS-CoV-2, influenza A and B viruses, respiratory syncytial virus (RSV), rhinovirus, adenovirus, parainfluenza viruses, and other respiratory pathogens.

SUPPLEMENTARY FIGURE 7

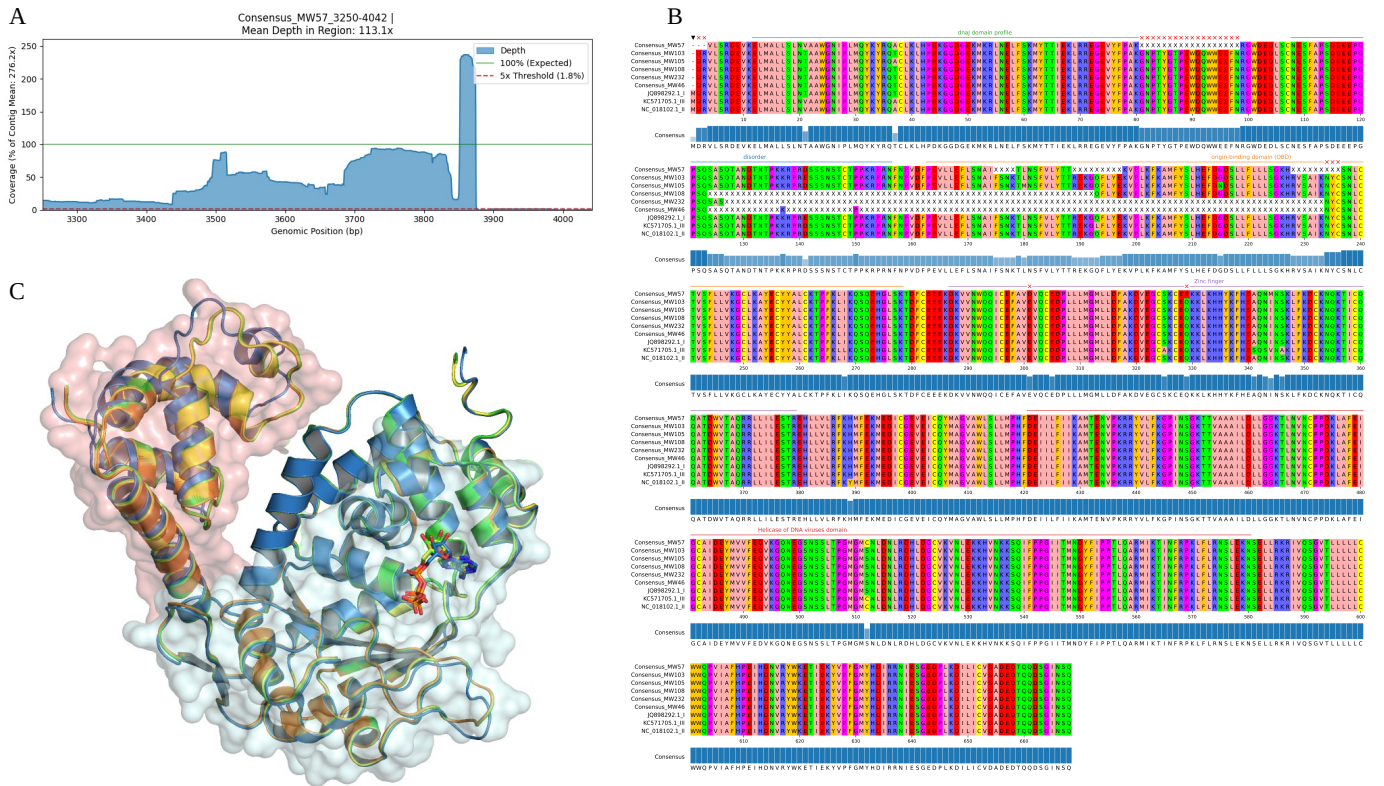


Fig. S7 Sequence and structural characterization of the recombinant Large T antigen from the MWPyV57 isolate. **A.** Normalized sequencing depth across the inferred recombinant region (positions 3450-3842) of the MWPyV57 genome. Sequencing reads were mapped to the assembled viral genome, and per-base coverage was calculated using SAMtools and the pysam Python library. Coverage values were normalized to the expected mean coverage of the corresponding contig, estimated from the number of mapped reads, average read length, and contig length. The green horizontal line indicates the expected normalized coverage (100%), whereas the red dashed line corresponds to the equivalent of 5 $\times$ sequencing depth. Continuous read support across the inferred recombination breakpoint supports the authenticity of the recombinant sequence and argues against an assembly artefact. **B.** Multiple amino acid sequence alignment of the Large T antigen from the recombinant MWPyV57 isolate and representative Malawi polyomavirus genotypes I (JQ898292.1), II (NC.018102.1), and III (KC571705.1). Coding sequences (CDSs) were translated and aligned using Clustal Omega. The recombinant region comprises exclusively synonymous nucleotide substitutions, resulting in complete conservation of the amino acid sequence, including the origin-binding, Zinc Finger, and Helicase domains. **C.** Structural comparison of the Zinc Finger and Helicase domains predicted using AlphaFold v3. Surface representation highlights the Zinc Finger domain in salmon and the Helicase domain in pale cyan. Cartoon representations are colored according to genotype: MW57 (green), genotype I (JQ898292.1; sky blue), genotype II (NC.018102.1; yellow), and genotype III (KC571705.1; orange). Structural superposition reveals highly conserved three-dimensional conformations among all sequences, indicating that the recombination event does not alter the predicted structural organization of these functional domains. Protein structure prediction and structural alignment were performed using a custom pipeline available at [github](https://github.com). Protein structure presentation was generated PyMOL (<https://www.pymol.org/>).

Table S1 Viral load (copies/reaction) of the positive for MWPyV DNA samples.

ID	Viral load (copies/reaction)
MWPyV11	13
MWPyV27	96.5
MWPyV46	110.9
MWPyV48	51.18
MWPyV57	179.5
MWPyV90	114.5
MWPyV99	13.2
MWPyV103	2346
MWPyV105	1037
MWPyV108	155
MWPyV114	19
MWPyV138	84
MWPyV139	182.5
MWPyV145	28
MWPyV208	12.5
MWPyV225	46
MWPyV228	14
MWPyV232	285

Table S2 Primers used for complete genome sequencing.

Name	Sequence (5' – 3')	Position
Left 1	GCTGAATTTGCAAGTAGTAAAAAGTTTGC	5 – 33
Right 1	ATCAAGAATAACAGAGATTACAGCACCC	4468 – 4495
Left 2	GTTAACACGAGATTTTCTTTGGAAATACTCC	396 – 426
Right 2	TCCCAACCAGAAAGAACATTAAAATAATATTC	4072 – 4104
Left 3	CTTTACAGCAATGGAGACCAGACTATTT	777 – 804
Right 3	GAGCTTCTCCTCTAAATATATAGTCTCCACT	3693 – 3723
Left 4	GTTCAAGTGGATAATTCTGAAAACTCAGTTT	1172 – 1202
Right 4	TGTTTGTTTTTACATCATTGCCCATCC	3302 – 3328
Left 5	CGGACACCACAATGACAGTTGAG	1573 – 1595
Right 5	CCATCTCTATCTAAAAATGGCTTTAAATGAGGC	2896 – 2927
Left 6	CAGTTTTGCAAGTTCAGGGGGAA	1973 – 1995
Right 6	ACACAGTTACATCCTGTACTTGATTATCAG	2498 – 2527
Left 7	ACAAATACCTAGCATTTGATGGACAGC	2375 – 2400
Right 7	TGAGTTGCTCAGAAAAAGAAATTGTTCAAAG	2098 – 2127
Left 8	CAAAGGGTACATATTTTTCAATTGTTTCTTTCC	2697 – 2729
Right 8	AGATAAGTTAGCTTTTGAAATTGGTTGTGC	1774 – 1803
Left 9	CATTCTGGAGTTAAAGATGAGTTAGAGC	3058 – 3086
Right 9	AAAACAATTTGTCAGCAAGCTACAGAC	1421 – 1447
Left 10	TGTTTAAACCTTAATACCAACAGATGTTCTCT	3419 – 3450
Right 10	AAATTATTGTTCAAATCTGTGTACAGTGAGC	1054 – 1084
Left 11	ACATTCATAGGCCTTTAAGCAACCTTT	3817 – 3843
Right 11	AGGATGGGATGAAGATTTAAGTTGTAATGAAT	652 – 683
Left 12	TTTGAGAAGGCCCCGGTTCTT	4215 – 4235
Right 12	CGATGTAGTAACCTTAGGAGATGTGCTA	258 – 285
Left 13	GCAAAAAGCATTTGTAGCAAAAACATTCA	4477 – 4505
Right 13	ATGGATAGAGTTCTTTCTAGAGATGAAGTAAAAG	1 – 34