Supplementary File S1. Technical Appendix for manuscript:

Investigation of US *Cyclospora cayetanensis* outbreaks in 2019 and evaluation of an improved *Cyclospora* genotyping system against 2019 cyclosporiasis outbreak clusters

Joel Barratt[1,2]*, Katelyn Houghton[1], Travis Richins[1,2], Anne Straily[1], Ryan Threlke1[1], Betelehem Bera[3], Jayne Kenneally[3], Brooke Clemons[3], Susan Madison-Antenucci[3], Elizabeth Cebelinski[4], Brooke M. Whitney[5], Katherine R. Kreil[5], Vitaliano Cama[1], Michael J. Arrowood[6], Yvonne Qvarnstrom[1]

[1] Parasitic Diseases Branch, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

[2] Oak Ridge Institute for Science and Education, Oak ridge, TN, USA

[3]Parasitology Laboratory, Wadsworth Center, New York State Department of Health, Albany, NY USA

[4] Minnesota Department of Health, St. Paul, MN, USA

[5] Coordinated Outbreak Response and Evaluation, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD, USA

[6] Waterborne Disease Prevention Branch, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

*To whom correspondence should be addressed. Email: jbarratt@cdc.gov Alternative email: nsk9@cdc.gov

Table of Contents

Supplementary Methods – Library Preparation for Illumina Sequencing

DNA extractions protocols used at participating NY and MN state laboratories

In the Parasitology Laboratory at the Wadsworth Center, New York (NY), the QIAamp PowerFecal DNA protocol was used with the following modifications. Aliquots of stool (1.0 ml) were transferred to a 2 mL conical tube, and centrifuged for 3 minutes at 20,000 x g. The supernatant was removed and the pellet was resuspended in 750 µL PowerBead Solution and 60 µL Solution C1 before transfer to a bead tube. After heating and vortexing the bead tube was centrifuged at 20,000 x g for 5 minutes. Post centrifugation, DNA was extracted from 450 µL of the supernatant with 10 µL of added plasmid extraction control. The standard QIAcube protocol was also modified for no tip reuse to eliminate the potential of cross contamination. Extracts were stored at -20°C.

At the Minnesota (MN) Department of Health Infectious Disease Laboratory, aliquots of 0.25-1 mL of stool were transferred into a 2 mL tube. The samples were washed twice by adding 1 mL of water, followed by thorough vortexing and then centrifugation at ≥20,000 x *g* for 3 minutes. After discarding the supernatant, the roughly 0.25 mL size stool pellets were re-suspended in water and delipidized with Citrus Clearing Solvent. The resulting pellets were washed with water and then resuspended in DNAzol (MRC, Inc., OH, USA), transferred to a Matrix C tube (MP Biomedicals) and boiled for 15 minutes. The tubes were placed on a FastPrep instrument (MP Biomedicals) and homogenized for 1 minute at speed setting 5.5 followed by centrifugation at 15,000 x *g* for 3 minutes. Clarified supernatants were ethanol precipitated, the pellets were washed twice and resuspended in 100 µL of water. The resulting solutions were purified further using the Qiagen QIAamp PCR Purification kit with a final elution volume of 50 µL.

Methods for deep amplicon sequencing used at the CDC laboratory

Due to differences in PCR sensitivity each amplicon was purified and normalized separately using a SequalPrep Normalization Kit (Thermo Fisher Scientific, MA, USA) with an elution volume of 20 µL. After purification and normalization, equal volumes of each amplicon were pooled and the resulting normalized pool of amplicons was subjected to library preparation using Nextera XT DNA Library Prep Kit in accordance with the manufacturer's instructions (Illumina, CA, USA). After amplicon pooling and library preparation, DNA concentration and DNA molecule size were assessed using a Qubit dsDNA HS Assay Kit (Invitrogen) and a High Sensitivity D1000 ScreenTape on the 2200 TapeStation (Agilent, CA, USA), respectively. The pooled amplicon libraries were diluted to 10 -15 pM and sequenced on the MiSeq Platform using either the MiSeq Reagent Kits V2 (500) or Nano V2 (500) (Illumina).

Methods for PCR and deep amplicon sequencing used at the MN laboratory

Thirty-two stool specimens containing *Cyclospora cayetanensis* as confirmed at a clinical lab (using either the BioFire FilmArray platform or microscopic examination of modified acid-fast stained fecal smears), were subjected to amplification of the eight genotyping markers, using the amplification and reaction conditions previously described. The thermal cyclers used included either a DNA Engine, iCycler or C1000 (Bio-Rad, CA, USA). The amplicons were individually purified using a Monarch purification kit (New England Biolabs) and amplicons where non-specific bands were apparent following gel electrophoresis, had the band of interest excised from the gel and purified using a MinElute Gel Extraction Kit (Qiagen). The purified amplicons were quantified, diluted, and pooled per specimen and subjected to the Nextera XT DNA Library Prep

method described above (in the CDC protocol) which was modified from a previously described protocol[1].

Methods for PCR and deep amplicon sequencing used at the NY laboratory

Ninety-one stool specimens positive for *C. cayetanensis* based on a routinely used real-time PCR and with a Ct value < 35, were subjected to genotyping. The PCR conditions were performed as described above for targets one to seven (in the CDC method), with the modification that 2 µL of template DNA was used instead of 1 µL. Target eight was amplified using HotStarTaq Master Mix (Qiagen), 400 nM of each primer, and 2 µL of DNA in a 25 µL reaction. The PCR conditions consisted of 95°C for 15 min, 35 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 45 sec, and final extension at 72°C for 10 min. This was performed on a SimpliAmp Thermal Cycler (Thermo Fisher Scientific). The amplicons were visualized on 1.5% agarose gel stained with ethidium bromide (Thermo Fisher Scientific) and visualized under UV light using Bio-Rad Universal Hood II Gel Doc System (Bio-Rad, CA, US). Sequencing was also attempted on amplicons that were negative by gel electrophoresis.

All the amplicons were purified using AMPure XP beads clean up (Beckman Coulter Life Sciences, IN, USA). Six microliters of each amplicon was transferred to a 96 well plate, and 10.8 µL of AMPure XP beads were added to each sample, capped and mixed by vortexing briefly. After five minutes of incubation at room temperature (RT) the plate was placed on a magnetic stand for two minutes and supernatant was removed and discarded without disturbing the beads. The beads were washed twice with 200 µL of 80% ethanol. After ethanol was removed, the plate was taken off the magnetic stand and 40 µL of nuclease-free water was added to each well. Wells were capped

[1] Modified from PulseNet: https://www.cdc.gov/pulsenet/pathogens/protocols.html

and the plate was briefly vortexed. After two minutes of incubation at RT the plate was placed back on the magnet stand and 30 µL of eluted amplicon was transferred to a new plate. Ten microliters of each amplicon from a single sample was pooled into a single tube. The pool was quantified using Qubit (Invitrogen), diluted, and subjected to library preparation as previously described (the CDC method above). The pooled libraries were diluted to 7-8 pM and loaded onto MiSeq using a MiSeq Nano, Micro or standard flow cells (depending on need).

Methods for PCR and deep amplicon sequencing used at the TX laboratory

PCR was performed using 200 nM of primers for targets one to four and eight, in a 25µL reaction using Platinum PCR Super Mix High Fidelity (Invitrogen) and 1 µL of DNA. The PCR conditions for targets one to four were 94°C for 2 min followed by 35 cycles of 94°C for 15 sec, 55°C for 15 sec, 68°C for 15 sec, and a final extension of 68°C for 5 min. For target eight, 40 seconds of extension was used following the same conditions for targets one to four. Targets five to seven PCRs were performed as previously described (CDC). The thermal cyclers used were a Verity (Applied Biosystems) and Master Cycler Nexus Gradient (Eppendorf, Germany).

The amplicons were visualized on 1.5% agarose gel stained with Ethidium Bromide (Invitrogen) and visualized under UV light using Biorad Universal Hood II Gel Doc System (Bio-Rad). All the amplicons were pooled, per specimen, and then purified using Monarch (New England Biolabs) on a final elution of 20 µL. After quantification using a Qubit, and dilution to 0.2 ng/µL, the pooled amplicons were subjected to library prep using Nextera XT (Illumina) following manufacturer's instructions. The pooled libraries were again quantified using a Qubit, and 10-20 pM were loaded onto a MiSeq Nano V2 (500) kit, and sequenced on a MiSeq Platform (Illumina). Sequencing was attempted on amplicons that were negative by gel electrophoresis.

Supplementary Methods – Description of Modules 1 through 3

The following sections provide a brief overview of the three major modules of the CDCs *C. cayetanensis* genotyping system. These descriptions should provide a rudimentary understanding of the functions performed by each module. For the precise details on how to run and install this software, and for more detailed information on how the workflows function, please refer to the following GitHub repository: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow

Haplotype calling (Module 1)

We previously utilized a haplotype calling workflow developed using the proprietary software *Geneious*[2]. This workflow performed Illumina read quality control and then identified the haplotypes present in each specimen by employing a map-to-reference strategy that required an exhaustive database of know haplotypes. That workflow was unable to detect novel haplotypes *de novo* by nature of this map-to-reference strategy; a major shortcoming of that workflow[2]. Here, we introduce a pipeline that detects novel haplotypes *de novo*; this pipeline constitutes *Module 1* of the CDCs genotyping system. *Module 1* first performs Illumina read (fastq.gz) quality control (QC) and then searches for novel haplotypes within the current batch of Illumina reads submitted to the module by the user. *Module 1* automatically characterizes novel haplotypes detected by assigning them a name, recording the specimen from which they were derived, and by recording their novel sequence to a reference database that is updated each time the module is run. After screening specimens for novel sequences, *Module 1* executes a routine that confirms the

[2] Nascimento FS, Barratt J, Houghton K, Plucinski M, Kelley J, Casillas S, et al. Evaluation of an ensemble-based distance statistic for clustering MLST datasets using epidemiologically defined clusters of cyclosporiasis. Epidemiology and Infection. 2020:1-30.

haplotypes present in the Illumina reads generated for each specimen (i.e., generates a list of haplotypes for each specimen) by comparing the reads to the now-up-to-date reference database. *Module 1* includes a separate routine that performs haplotype discovery specifically for the mitochondrial junction repeat, which requires a separate (and more complex) routine due to its repetitive nature. Each time *Module 1* is initiated, these steps are performed on the fastq.gz supplied in that specific run. If new haplotypes are discovered in that run, the reference database of known haplotypes is expanded to include these new haplotypes; the previous database is never over-written. The user has the option to supply a range of arguments to the *Module 1* scripts, including setting the amount of RAM allocated for clustering, or the number of threads to be used when certain functions are performed. At the final stages of the *Module 1* workflow, a separate text file is printed for each specimen containing a list of the haplotypes detected in that specimen. The contents of each specimens' haplotype list are examined by the *Module 1* code and the data in these lists is compiled into a haplotype data sheet which is used as the direct input for *Module 2*. For further details on these aspects of *Module 1*, please refer to the following GitHub repository: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow.

Calculation of a distance matrix (Module 2)

Module 2 comprises the 'Eukaryotyping' procedure that calculates the ensemble-based distance statistic previously described[2]. An advantage of the 'Barratt-Plucinski' ensemble procedure is that specimens with sequence data missing for some loci may still be retained for analysis, though it is paramount that reasonable minimum data availability requirements are set by the user. Here, we use the same minimum data requirements set previously: [1] specimens must have sequence data available for any three of markers 5, 6, 7 and 8 plus at least one additional

marker, and/or [2] specimens must have sequence data available for at least any five markers. Specimens that fail to meet at least one of these two criteria are excluded from further analysis. A pairwise distance matrix is then generated using the R scripts and directions provided here: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow, which was modified from the original 'Eukaryotyping' software available at the following GitHub repository: https://github.com/Joel-Barratt/Eukaryotyping.

Genetic Cluster Delineation (Module 3)

Previously, we employed a bootstrapping strategy where epidemiologic data were used to infer the most appropriate number of genetic clusters from a hierarchically clustered Barratt-Plucinski ensemble matrix[2]. This entailed selecting one of three empirically-selected cluster numbers (5, 10 and 20 clusters) that: (a) minimized assigning specimens associated with different epidemiologic clusters to the same genetic cluster, and (b), minimized separation of specimens from epidemiologically-linked case-patients across multiple genetic clusters. However, the purpose of molecular surveillance is to inform epidemiologic investigations in the absence of epidemiologic information, so this strategy cannot be used in the context of routine molecular surveillance. That was the impetus for the development of *Module 3*.

Module 3 requires the user to first compile a list of reference specimens fulfilling each of two criteria: (a) the list must only contain specimens possessing known epidemiologic linkage to other (multiple) specimens in that same list, and (b), the list must contain specimens that had been placed correctly into a genetic cluster analogous with its epidemiologic linkage previously using the CDCs *C. cayetanensis* genotyping system. For this study, a subset of specimens linked to the Vendor A (n = 99) and Vendor B (n = 104) outbreaks of 2018 that previously obtained a true positive classification were included in this list (Supplementary File S2, Tab B). The paired-end

fastq.qz files generated for the 2018 specimens in this list were subjected to the *Module 1* workflow (see GitHub repository: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow), along with the paired-end reads generated from the 1,110 fecal specimens genotyped during 2019, to generate a haplotype data sheet including both the 2019 specimens and the 2018 reference population. A distance matrix is next calculated from this data sheet using *Module 2* of CDCs genotyping system and this matrix is required as one of the inputs of *Module 3*.

Module 3 requires five user inputs: (a) the list of the reference specimens described above - formatted as described in GitHub repository: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow (b) a stringency level (*S*) supplied as a percentage -- a number between 1 and 100 – to the source code as an argument, (c) a minimum number of clusters to test (also an argument), (d) a maximum number of clusters to test (also an argument), (e) a single distance matrix calculated using *Module 2*. This matrix must have been calculated using specimens in the reference list in addition to the population under investigation; in this case, the *C. cayetanensis* population causing illness in the US in 2019. Instructions on how to supply these files and arguments to *Module 3* are provided in this GitHub repository: https://github.com/Joel-Barratt/CYCLONE-CDC-Complete-Cyclospora-typing-workflow.

Module 3 first calculates the average distance (from the 'Barratt-Plucinski' matrix supplied) between all specimens possessing the same genetic and epidemiologic linkage; this linkage information is provided in the formatted reference list previously compiled. Three standard deviations are added to that average distance to determine a cutoff distance (*D*). The module then cycles through a loop that tests each cluster number from the supplied minimum to the supplied maximum, to determine the smallest cluster number where *S* percent of all within-cluster-distances

are less than or equal to *D*. If a range of cluster numbers is supplied that never satisfies this criterion, the module will not generate an output. If this criterion is satisfied, the most appropriate cluster number based on these criteria will be printed on the screen and a list of cluster memberships for each specimen will be printed to a text file. Here, we used a stringency level of 95% and investigated cluster numbers from 5 to 50. The matrix and reference list compiled are provided in this GitHub repository: https://github.com/Joel-Barratt/CDC-Complete-Cyclospora-typing-workflow-ALPHA-TEST.

Genotyping results obtained for epidemiologic clusters assigned to category 2

For two category 2 epi-clusters, only 2 specimens were genotyped and when clustered genetically they obtained 100% concordance with their epidemiologic linkage. For one category 2 epi-cluster, 4 case-specimens were genotyped and these clustered genetically also with 100% concordance. For another category 2 epi-cluster, 3 specimens were genotyped and each was assigned to a separate genetic cluster. A specific food vehicle could not be confidently identified for this epidemiologic cluster; basil or onions were suspected equally. Two additional epidemiologic clusters had 3 genotyped case-specimens, where 2 specimens were assigned to one genetic cluster and 1 was assigned to a different genetic cluster from its epidemiologically linked partners. One category 2 epi-cluster had 2 specimens genotyped, each falling into two separate genetic clusters. Finally, for 1 epidemiologic cluster, only 1 specimen was genotyped.

Table S2. Explanation of specimen counts and case-patient counts at different points in this study

Definition	Number	Notes and explanations
Number of **specimens** from cyclosporiasis cases for which genotyping was attempted	1,110	This includes specimens sent to the four participating laboratories for genotyping: CDC (n=430), TX (n=267), NY (n=381) and MN (n=32).
Number of **specimens** that were retained for classification based on the minimum data requirements for inclusion in the analysis	875	As described in the methods within the main manuscript text, these are the specimens that were retained for analysis. This includes specimens that (1) have sequence data available for any three of markers 5, 6, 7 and 8 (Table 1) plus one additional marker, and/or (2) have sequence data available for any 5 markers. There were 875 specimens collected during 2019 that met these criteria.
Number of "true-positive" **specimens** collected from 2018 that were analyzed alongside the 2019 specimens as part of this study.	203	This includes 99 specimens assigned to the Vendor A outbreak of 2018 and 104 specimens assigned to the Vendor B outbreak of 2018, that clustered correctly as described in the study by Nascimento et al. (2020)[α]
Total number of **specimens** analyzed in this study	1078	This includes 875 specimens from 2019 in addition to the 203 specimens from 2018 that were added to the dataset.
Number of domestically acquired **laboratory-confirmed cyclosporiasis cases** reported in 2019.	2,408	This is the number of cases that were laboratory confirmed in the USA by the 13th of November 2019. Only a subset of specimens from these case-patients (n=875) were sent to CDC for genotyping.
Number of **genotyped specimens** from this study with associated epidemiologic data	390	Includes the 203 specimens from 2018, alongside 187 specimens possessing associated epidemiologic links that were collected during 2019. Refer to Table 2 for further details.
Number of genotyped **specimens** from this study with **unknown** epidemiologic linkage	688	Of these 688 specimens, 16 were assigned to genetic cluster 1 containing many specimens linked to Distributor A. Cluster 3 was also associated with Distributor A (basil) and 147 specimens assigned to this cluster had no associated epidemiologic data. Cluster 17 (also associated with Distributor A) had 84 specimens without epidemiologic links assigned to this cluster. Cluster 18 (also associated with Distributor A) had 44 specimens without epidemiologic links assigned to this cluster. This is a total of 291 specimens without epidemiologic links, that were genetically linked to the four major genetic clusters associated with Basil (16 + 147 + 84 + 44 = 291) supplied by distributor A. The remaining 397 specimens without associated epidemiologic data were distributed among the other genetic clusters.
Number of **genotyped specimens** that are members of epidemiologically-defined clusters with less than six specimens genotyped.	20	The number of genotyped specimens associated with an epidemiologically-defined cluster where fewer than six specimens associated with that epidemiologically-defined cluster were genotyped.
Number of **genotyped specimens** with associated epidemiologic information that were used to assess genotyping performance	364	The number of specimens in this study with associated epidemiologic data (n=390) minus the number of specimens from epi-clusters where fewer than 6 specimens were genotyped (n=20), and minus specimens associated with Restaurant E (n = 6).

CDC = Centers for Disease Control and Prevention, NY = New York, TX = Texas

[α] http://dx.doi.org/10.1017/S0950268820001

Appendix A – Original reference haplotypes

```
>Nu_CDS4_PART_B_Hap_1
GTGGAGACGCCAGAAATCTCCGCCGCTCCTGGCACGGCCTCAGCATCCGTAGCCACAGCCTCGTCCACT
>Nu_CDS4_PART_B_Hap_2
GTGGAGACGCCAGAAATCTCTGCCGCTCCTGGCACGGCCTCAGCATCCGTAGCCACAGCCTCGTCCACT
>Nu_CDS4_PART_A_Hap_1
TCCGACGCTTCTGATGCCGCAGCTTCCATTCAGATCGACGACGTCGAGTCATCTGGTGAAGATGACTCG
>Nu_CDS4_PART_A_Hap_2
TCCGACGCTTCTGAGACCGCAGCTTCCATTCAGATCGACGACGTCGAGTCATCTGGTGAAGATGACTCG
>Nu_CDS3_PART_B_Hap_1
CATCTACAAGTGCACCAGCAGAACTAACCCTGTCACGAAAGGCGGCACGAGTGCTGCGCTAGAGGCTTCTCGGTTTGTTCAGCCGCCGC
>Nu_CDS3_PART_A_Hap_1
GTTGCTTTATCTTTGGCAGTTAGACAGGGAGGGATTCCTACATTTGCAGGTGATGCTGCTACCTCCATGCGCAGCTGCAACACCAGCAA
>Nu_CDS3_PART_A_Hap_2
GTTGCTTTATCTTTGGCAGTTAGACAGGGAGGGATTCCTACATTTGCAGGTGATGCTGCTACCTCCGTGCGCAGCTGCAACACCAGCAA
>Nu_CDS3_PART_A_Hap_3
GTTGCTTTATCTTTGGCAGTTAGACAGGGAGGGATTCCTACATTTGCAGGTGATGCTGATACCTCCGTGCGCAGCTGCAACACCAGCAA
>Nu_CDS2_PART_B_Hap_1
GCTGCCGCAGCGGCAGCAGCGGAAGTTCTGCGAACCGCCTCCCGTTCGTCCCTCGTGTCTACGCGAAGCGCCAAGAAGTTGCAGCATAC
AAAGAAGCAGCTTG
>Nu_CDS2_PART_A_Hap_1
CCAGATAGGTGGAGGCGCAGATGGCCTTAAGAGCAAGAATGCCGGACTGGTGATGCTGGCGAGGGGTGATGTTGCCGCGGCGACTGCTG
CTAGATCCCCTGCT
>Nu_CDS2_PART_A_Hap_2
CCAGATAGGTGGAGGCGCAGATGGCCTTAAGAGCAAGAATGCCGGACTGGTGATGCTGGCGAGGGGCGATGTTGCCGCGGCGACTGCTG
CTAGATCCCCTGCT
>Nu_CDS1_PART_B_Hap_1
TGTGCTTGCATGCATCCGTAGATGCATGTCACCTCTAGCCCTTTGTTGCTTGAGGCGCTGCTTATCCG
>Nu_CDS1_PART_B_Hap_2
TGTGCTTGCATGCATCCGTAGACGCATGTCACCTCTAGCCCTTTGTTGCTTGAGGCGCTGCTTATCCG
>Nu_CDS1_PART_A_Hap_1
GGATCACTCTCCAGGCGACACCCTCCTCGCGCTCTGTCCACCGTTTCTCTACGCTTCCCCATGTCTC
>Nu_CDS1_PART_A_Hap_2
GGATCACTCTCCAGGCGCTGCCCTCCTCGCACTGTGTCCACCCTTTCTCTACGCTTCCCCATGTCTC
>Nu_378_PART_D_Hap_1
CGATGTGCGTCCATAGAACGTTCCTTGCGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
GTCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_2
GAATTTGCATCCAAAGAACGTTCCTCGTGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
ATCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_3
CGATGTGCATCCATAGAACGTTCCTTGTGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
GTCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_4
GAATTTGCATCCATAGAACGTTCCTTGCGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
GTCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_5
GAATTTGCATCCACAGAACGTTCCTCGTGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTC
ATCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_6
GAATTTGCATCCATAGAACGTTCCTCGTGTCTGCATATTTTCCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAATTGGTTGCACTTT
GTCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_7
GAATTTGCATCCAAAGAACGTTCCTTGCGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
ATCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_C_Hap_1
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCGTTCGACAGGATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
GCTCTGCAGAG
>Nu_378_PART_C_Hap_2
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCGTTCGACAAAATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
GCTCTGCAGAG
>Nu_378_PART_B_Hap_1
```

AAGGGACCGAAAGATGCGTCCACTCCTTGTACTCTCGTTGCAGCGGCTCCAGCAGTTGCAGCGAGAGTCTGGATCAGAGGAGAAGGTGA
CGCGGCTGGTC
>Nu_378_PART_B_Hap_2
AAGAGACCGAAAGATGCGTCCACTCCTTGTACTCTCGTTGCAGCGGCTCCAGCAGTTGCAGCGAGAGTCTGGATCAGAGGAGAAGGTGA
CGCGGCTGGTC
>Nu_378_PART_A_Hap_1
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTTGGTCGATTGTCGCACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_378_PART_A_Hap_2
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTTGGTCGATTGTCGTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_378_PART_A_Hap_3
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTTGGTCGATTGTCTTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_360i2_PART_F_Hap_2
CTGCAGCGCACGAACAAAGTGCTGTACGCTCTCTTACAACGTTCCGATTAAAGATTTAGGGGCAGGCCATGAAGGCGCCAATGCAGCCG
TTTTAACAGACCTATGAAATT
>Nu_360i2_PART_F_Hap_1
CTGCAGCGCACGAACAAAGTGCTGTACGCTCTCTTAGAGCGTTCCGATTAAAGATTTAGGGGCAGGCCATGAAGGCGCCAATGCAGCCG
TTTTAACAGACCTATGAAATT
>Nu_360i2_PART_E_Hap_1
TGATTCCTGCATGCAAAACGCCTAATCCTAGGAACACACCTTTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCCCTGCTTC
TCCCGCAACAG
>Nu_360i2_PART_E_Hap_2
TGATTTCTGCATGCAAAACGCCTAATCCCAGGAACACGCCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCCCGCAACAG
>Nu_360i2_PART_E_Hap_3
TGATTCCTGCATGCAAAACGCCTTATCCCAGGAACACGCCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCCCGCAACAG
>Nu_360i2_PART_E_Hap_4
TGATTCCTGCATGCAAAACGCCTAATCCCAGGAACACGCCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCCCGCAACAG
>Nu_360i2_PART_E_Hap_5
TGATTTCTGCATGCAAAACGCCTAATCCCAGGAACACACCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCCCGCAACAG
>Nu_360i2_PART_D_Hap_1
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGGGCCTCAATTTTTT
CGAGTCTGGCT
>Nu_360i2_PART_D_Hap_2
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGAGCCTCAATTTTTC
TGAGTCTGGCT
>Nu_360i2_PART_D_Hap_3
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGGGCCTCAATTTTTC
TGAGTCTGGCT
>Nu_360i2_PART_D_Hap_4
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGAGCCTCAATTTTTC
TGAGTCTTGCT
>Nu_360i2_PART_C_Hap_1
GGGCTTCTTGCTGATGTATCGTCGTCTTCTGCCCCATCCACAGATTGCATCTTCTCGGATTAGTAAGTTAGGAGAGCAACGATCCGAGC
ACATTTAGGAG
>Nu_360i2_PART_C_Hap_2
GGGCTTCTTGCTGACGTATCGTCGTCTTCTGCCCCATCCACAGATTGCATCTTCTCGGATTAGTAAGTTAGGAGAGCAACGATCCGAGC
ACATTTAGGAG
>Nu_360i2_PART_C_Hap_3
GGGCTTCTTGCTGATGTATCGTCGTCTTCTGCCCCATCCACGGATTGCATCTTCTCAGATAAGTAAGTTAGGAGAGCAACGATCCGAGC
ACATTTAGGAG
>Nu_360i2_PART_C_Hap_4
GGGCTTCTTGCTGATGTATCGTCGTCTTCTGCCCCATCCACAGATTGCATCTTCTCGGATAAGTAAGTTAGGAGAGCAACGATCCGAGC
ACATTTAGGAG
>Nu_360i2_PART_B_Hap_1
GGGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCAGGAAGGGATTAGGCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGCT
>Nu_360i2_PART_A_Hap_2

TTCTTTTCGCACTCTGGCTGCCGCTGCTGGGTCCGCTCAAGCAAAGAGAATGGCACGCAGTCATTTTGGTGTATTTTTCATAATCCGTA
GCTTGTGGGGA
>Nu_360i2_PART_A_Hap_1
TTCTTTTCGCACTCTGGCTGCCGCTGCTGGTTCTGCTCAAGCAAAGAGAATGGCACGCGGTCATTTTGGTGTATTTTTCATAATCCGTA
GCCTGTGGGGA
>Mt_MSR_PART_F_Hap_1
CTACCTGACGCTTAGTAACGATTCCGTCTTCCAGCTGCCAATTTTTATTTCAAATGTTCCACTACCAAATTATATCCATATGTTCCAAA
TGTTCAGGAGTGTATGTAATAAGTTTTTGTTTT
>Mt_MSR_PART_F_Hap_2
CTACCTGACGCTTAGTAACGATTCCGTCTTCCAGCTGCCAATTTTTATTTCAAATGTTCCACTACCAAATTATATCCATCTGTTCCAAA
TGTTCAGGAGTGTATGTAATAAGTTTTTGTTTT
>Mt_MSR_PART_F_Hap_3
CTACCTGACGCTTAGTAACGATTCCGTCTTCCAGCTGCCAATTTATATTTCAAATGTTCCACTACCAAATTATATCCATCTGTTCCAAA
TGTTCAGGAGTGTATGTAATAAGTTTTTGTTTT
>Mt_MSR_PART_E_Hap_1
CAGTTTTTGGCTAATATTTGACTTTTTGAACCAGCCTGGGATCATAAAATTCAGCCAGGGACCGTTCCATTGGGATCACGCATGGATTC
AGTGTCCAGGA
>Mt_MSR_PART_D_Hap_1
GCTGAGTGCTTGTGCGATAATTATATCATATTTTTTAGACGCTAACTTCCCGGCTAAACTTCCCTTATTTTTTGTCTAAGAAACACACT
TCCCTTCTCGC
>Mt_MSR_PART_C_Hap_1
AAGACACTGCTAACTTTGATGAGGAGGTATAGTTTATGACATCTGCCCGGCACCCTAACAAAAAACCGGTTTTTAACACGTCAATACAT
TCTCTTAACTT
>Mt_MSR_PART_B_Hap_1
ACCAGCGCTTCCATTAATAAGAAAGTGAGAAGATTGTTTGAATAGTTTGGCATTTCGTAGTGTAGTTAAAAAACGGAAAGCGTCTGTTG
GATTCAAACGT
>Mt_MSR_PART_B_Hap_2
ACCAGCGCTTCCATTAATAAGAAAGTGATAAGATTGTTTGAATAGTTTGGCATTTCGTAGTGTAGTTAAAAAACGGAAAGCGTCTGTTG
GATTCAAACGT
>Mt_MSR_PART_A_Hap_1
GTTGTTTCCATCTCGACTCTACAGTTTAACTTCTGGAGTCCCTTTTGGTAAAAAACAGTACAAAGTTGGACTGAGGCTGAGTCTCTATG
CCGAATGAGGAGCACTGGATTGGATACCCAGGT
>Mt_MSR_PART_A_Hap_2
GTTGTTTCCATCTCGACTCTACAGTTTAACTTATGGAGTCCCTTTTGGTAAAAAACAGTACAAAGTTGGACTGAGGCTGAGTCTCTATG
CCGAATTAGGAGCACTGGATTGGATACCCAGGT
>Mt_MSR_PART_A_Hap_3
GTTGTTTCCATCTCGACTCTACAGTTTAACTTATGGAGTCCCTTTTGGTAAAAAACAGTACAAAGTTGGACTGAGGCTGAGTCTCTATG
CCGAATGAGGAGCACTGGATTGGATACCCAGGT
>Mt_Cmt109.A_Junction_Hap_1
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt127.A_Junction_Hap_2
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt154.A_Junction_Hap_3
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTACTATTATAAATA
TAAAAATTTAGTACACCTAGCC
>Mt_Cmt154.B_Junction_Hap_4
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTACTATTTTTAATA
TAAAAATTTAGTACACCTAGCC
>Mt_Cmt154.C_Junction_Hap_5
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTATTATTTTTAATA
TAAAAATTTAGTACACCTAGCC
>Mt_Cmt154.D_Junction_Hap_6
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTACTATTTTAAATA
TAAAAATTTAGTACACCTAGCC
>Mt_Cmt154.E_Junction_Hap_7
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTTAATAGTATTATTTTTAATAGTACTATTTTTAATA
TAAAAATTTAGTACACCTAGCC
>Mt_Cmt169.A_Junction_Hap_8
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTATTATTTTTAATA
GTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt169.B_Junction_Hap_9
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTTTAATA
GTACTATTTTTAATATAAAAATTTAGTACACCTAGCC

>Mt_Cmt169.C_Junction_Hap_10
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTTAATAGTATTATTTTTAATAGTATTATTTTTAATA
GTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.A_Junction_Hap_11
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.B_Junction_Hap_12
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.C_Junction_Hap_13
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTATAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.D_Junction_Hap_14
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTTTAATAGTATTATTTATAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.E_Junction_Hap_15
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTTAATAGTATTATTTATAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt184.F_Junction_Hap_16
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt199.A_Junction_Hap_17
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt199.B_Junction_Hap_18
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTTTAATAGTATTATTTATAATAGTATTATTTTTAATA
GTATTATTTTTAATAGTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt199.C_Junction_Hap_19
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTATAATA
GTATTATTTTTAATAGTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC
>Mt_Cmt214.A_Junction_Hap_20
TGCGGAAACTGTATTTTTATTATTTAATTTTACTATTTTAAATAGTATTATTTATAATAGTATTATTTATAATAGTATTATTTATAATA
GTATTATTTTTAATAGTATTATTTTTAATAGTATTATTTTTAATAGTACTATTTTTAATATAAAAATTTAGTACACCTAGCC

>Mt_Cmt214.X_Junction_Hap_21
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtattatttataata
gtattatttttaatagtattatttataatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Mt_MSR_PART_E_Hap_2
CAGTTTTTGGCTAATATTTGACTTTTAGAACCAGCCTGGGATCATAAAATTCAGCCAGGGACCGTTCCATTGGGATCACGCATGGATTC
AGTGTCCAGGA
>Mt_Cmt214.X_Junction_Hap_28
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtattatttttaata
gtattatttttaatagtattatttttaatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_360i2_PART_C_Hap_5
GGGCTTCTTGCTGATGTATCGTCGTCTTCTGCCCCATCCACAGATTGCATCTTCTCGGATTAGTAAGTTAGTAGAGCAACGATCCGAGC
ACATTTAGGAG
>Nu_360i2_PART_A_Hap_4
TTCTTTTCGCACTGTGGCTGCCGCTGCTGGGTCCGCTCAAGCAAAGAGAATGGCACGCAGTCATTTTGGTGTATTTTTCATAATACGTA
GCTTGTGGGGA
>Mt_Cmt199.X_Junction_Hap_29
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttaaatagtattatttataatagtattatttataata
gtattatttataatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_CDS3_PART_A_Hap_6
GTTGCTTTATCTTTGGCAGTTAGACAGGTAGGGATTCCTACATTTGCAGGTGATGCTGCTACCTCCATGCGCAGCTGCAACACCAGCAA
>Nu_360i2_PART_B_Hap_4
GTGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCAGGAAGGGATTAGGCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGCT
>Nu_378_PART_C_Hap_3
ATCTTTTGCCTTTAGTTCTTTTTGAGGTCAGTCGTTCGACAAAATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
GCTCTGCAGAG
>Nu_360i2_PART_E_Hap_7
TGATTCCTGCATGCAAAACGCCTTATCCCAGGAACACGCCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCACGCAACAG
>Nu_378_PART_A_Hap_7
CTTATGGTAATTTGCTTTGTGCATCTGCTTCCCTGTTGTTCGATTGTCGTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_378_PART_D_Hap_10
GAATTTGCATCCAAAGAACGTTCCTCGTGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTT
ATCGTGCTGTCTGTGCAATGGCTCA
>Nu_360i2_PART_D_Hap_6
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTTCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGGGCCTCAATTTTTT
CGAGTCTGGCT
>Mt_Cmt199.X_Junction_Hap_30
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtattatttttaata
gtattatttttaatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_CDS3_PART_A_Hap_4
GTTGCTTTATCTTTGGCAGTTAGACAGGGAGGGATTCCTACATTTGCAGGTGATGCTGCTACCTCCATGCGCAGCTGCAACACTAGCAA
>Nu_360i2_PART_B_Hap_5
GAGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCAGGAAGGGATTAGGCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGCT
>Mt_Cmt154.X_Junction_Hap_31
tgcggaaactgtatttttattattcaattttactattttaaatagtattatttataatagtattatttttaatagtactattataaata
taaaaatttagtacacctagcc
>Nu_378_PART_D_Hap_11
GAATTTGCATCCACAGAACGTTCCTCGTGTCTGCATATTTTCCGGTGGTAGATGCAGGAGCAAGAGTAGGAGTGAGTTGGTTGCACTTC
ATCGTGCTGTCTGTGCAATGGCGCA
>Nu_CDS1_PART_B_Hap_3
TGTGCTTGCATGCATCCGTAGACGCATGTCACCTCTAGCCCTTTGTTGCTTGAGGTGCTGCTTATCCG
>Mt_Cmt199.X_Junction_Hap_32
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttaaatagtattatttataatagtattatttttaata
gtattatttttaatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_360i2_PART_F_Hap_3
CTGCAGCGCACGAACAAAGTGCTGTACGCTCTCTTACAACGTTCCGATTAAAGATTTAGGGGCAGGCCATGAAGACACCAATGCAGCCG
TTTCAACAGACCTATAAAATT

>Mt_MSR_PART_B_Hap_3
ACCAGCGCTTCCATTAATAAGAAAGTGAGAAGATTGTTTGAATAGTTTGGCATTTCGTAGTGTAGTTAAAAAACGGAAAGCGTCTGTTG
GATTCAAACAT
>Nu_360i2_PART_E_Hap_6
TGATTCCTGCATGCAAAACGCCTTATCCCAGGAACACACCTGTTCTCATCAAAACTCAACGGTTGCGTGACATCAGATAGCGCTGCTTC
TCCCGCAACAG
>Nu_378_PART_A_Hap_6
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTGGGTCGATTGTCTTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_378_PART_D_Hap_8
CGATGTGCATCCATAGAACGTTCCTTGTGTCTGCATATTTTTCGGTGGTAGATGCAGAAGCAAGAGTAGGAGTGAATTGGTTGCACTTC
ATCGTGCTGTCTGTGCAATGGCGCA
>Nu_378_PART_D_Hap_9
GAATTTGCATCCAAAGAACGTTCCTTGCGTCTGCATATTTTTCGGTGGTAGATGCAGGAGCAAGAGTAGTAGTGAGTTGGTTGCACTTT
ATCGTGCTGTCTGTGCAATGGCGCA
>Mt_Cmt199.X_Junction_Hap_24
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtattatttttaata
gtattatttttaatagtattatttttaatagtacttttttttaatataaaaatttagtacacctagcc
>Nu_CDS1_PART_A_Hap_3
GGATCACTCTCCAGGCGCTGCCCTCCTCGCACTGTGTCCACCCTTTCTCTACTCTTCCCCATGTCTC
>Mt_Cmt214.X_Junction_Hap_25
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttataatagtattatttttaata
gtattatttttaatagtattatttttaatagtattatttttaatagtactattttaatataaaaatttagtacacctagcc
>Mt_Cmt139.X_Junction_Hap_26
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatataaaaatttagtaca
cctagcc
>Nu_CDS3_PART_B_Hap_2
CATCTACAAGTGCACCAGCAGAACTAACCCTGTCACGAAAGGCGGCACGAGTGCTGCGCTAGAGGCTTCTCGGTTAGTTCAGCCGCCGC
>Nu_360i2_PART_A_Hap_3
TTCTTTTCGCACTGTGGCTGCCGCTGCTGGGTCCGCTCAAGCAAAGAGAATGGCACGCAGTCATTTTGGTGTATTTTTCATAATCCGTA
GCTTGTGGGGA
>Nu_378_PART_A_Hap_4
CTTATGGTCATTTGCTTTGTGTATCTGCTTCCCTGTGGGTCGATTGTCGCACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_360i2_PART_B_Hap_2
GGGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCATGAAGGGATTAGGCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGCT
>Nu_378_PART_A_Hap_5
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTTGTTCGATTGTCGTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Mt_Cmt199.X_Junction_Hap_22
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttaaatagtattatttataata
gtattatttttaatagtattatttttaatagtactattttaatataaaaatttagtacacctagcc
>Nu_360i2_PART_D_Hap_5
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATTCCAAGAGGGCCTCAATTTTTT
CGAGTCTGGCT
>Mt_Cmt199.X_Junction_Hap_23
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtattatttttaata
gtactatttttaatagtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_360i2_PART_B_Hap_3
GGGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCAGGAAGGGATTAGGCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGAT
>Nu_CDS3_PART_B_Hap_3
CATCTACAAGTGCACCAGCAGAACTAACCCTGTCACGAAAGGCGGCACGAGTGCTGCGCTATAGGCTTCTCGGTTTGTTCAGCCGCCGC
>Mt_Cmt139.X_Junction_Hap_27
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttttaatagtactatttttaatataaaaatttagtaca
cctagcc
>Mt_Cmt184.X_Junction_Hap_33
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttttaatagtattatttttaatagtattatttttaata
gtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Mt_Cmt184.X_Junction_Hap_34
tgcggaaactgtatttttattatttaattttacttttttaaatagtattatttataatagtattatttttaatagtattatttttaata
gtattatttttaatagtactatttttaatataaaaatttagtacacctagcc

>Mt_Cmt184.X_Junction_Hap_35
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttaaatagtattatttataatagtattatttataata
gtattatttttaatagtactatttttaatataaaaatttagtacacctagcc
>Nu_378_PART_A_Hap_8
CTTATGGTCATTTGCTTTGTGCATCTGCTTCCCTGTGGGTCGATTGTCGTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_378_PART_C_Hap_4
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCGTTCGACAGGATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
TCTCTGCAGAG
>Nu_CDS3_PART_A_Hap_5
GTTGCTTTATCTTTGGTAGTTAGACAGGGAGGGATTCCTACATTTGCAGGTGATGCTGCTACCTCCATGCGCAGCTGCAACACCAGCAA
>Mt_Cmt169.X_Junction_Hap_36
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttaaatagtattatttataatagtattatttttaata
gtactatttttaatataaaaatttagtacacctagcc
>Mt_Cmt169.X_Junction_Hap_37
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttataatagtattatttttaatagtactatttttaata
gtactattataaatataaaaatttagtacacctagcc
>Nu_360i2_PART_B_Hap_6
GGGAATGTCACCTAATGTTGCGTCACTAGCATCGGATGGCAGGAAGGGATTAGTCTCAGATTCGGGCAGCATAAGCTGCAGGAGTTGTA
AGGCGGCGGCT
>Nu_360i2_PART_C_Hap_6
GGGCTTCTTGCTGATGTATAGTCGTCTTCTGCCCCATCCACAGATTGCATCTTCTCGGATTAGTAAGTTAGGAGAGCAACGATCCGAGC
ACATTTAGGAG
>Mt_Cmt139.X_Junction_Hap_38
tgcggaaactgtatttttattatttaattttactattttaaatagtattatttttaatagtactattataaatataaaaatttagtaca
cctagcc
>Nu_360i2_PART_D_Hap_8
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCATAAAAGCAAAGGGTATGCCAAGAGGGCCTCAATTTTTT
CGAGTCTGGCT
>Nu_378_PART_C_Hap_5
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCGTTCGACAAAATAGGGCGATTGGGGAGGCTGGTACATGGACACCAGTGTTGCAAGG
GCTCTGCAGAG
>Nu_378_PART_A_Hap_9
CTTATGGTCATTTTCTTTGTGCATCTGCTTCCCTGTTGGTCGATTGTCGTACCAGGTGTTGGAGTCCGAGGTGTATTCTGAACTGAGGA
ATGTAGGTGCATGTTACTGGCACGG
>Nu_CDS1_PART_B_Hap_4
TGTGCTTGCATGCATCCGTAGACGCATGTCACCTCTATCCCTTTGTTGCTTGAGGCGCTGCTTATCCG
>Nu_CDS2_PART_B_Hap_2
GCTGCCGCAGCGGCAGCAGCGGAAGTTCTGCGAACCGCCTCCCGTTCGTCCCTCGTGTCTACGCCAAGCGCCAAGAAGTTGCAGCATAC
AAAGAAGCAGCTTG
>Nu_378_PART_C_Hap_6
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCGTTCGACAAAATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
GATCTGCAGAG
>Nu_CDS1_PART_B_Hap_5
TGTGCTTGCATGCATTCGTAGACGCATGTCACCTCTAGCCCTTTGTTGCTTGAGGCGCTGCTTATCCG
>Nu_378_PART_C_Hap_7
ATCTTTTGCCTTGAGTTCTTTTTGAGGTCAGTCATTCGACAAAATAGGGCGATTGGGGAGGCTGGTACATGGGCACCAGTGTTGCAAGG
GCTCTGCAGAG
>Nu_CDS3_PART_B_Hap_4
CATCTACAAGTGCACCAGCAGAACTAACCCTGTCACGAAAGGCGGCACGAGTGCTGCGCTAGAGGCTTCTCGGTTTGTTCAGACGCCGC
>Mt_MSR_PART_E_Hap_3
CAGTTTTTGGCTAATATTTGACTTTGTGAACCAGCCTGGGATCATAAAATTCAGCCAGGGACCGTTCCATTGGGATCACGCATGGATTC
AGTGTCCAGGA
>Nu_360i2_PART_D_Hap_7
TAGCTTTCTGGTAGTGCGTCGCAAGAATTTTGCACCATCCTTGCGGGTGCGTAAAAGCAAAGGGTATGCCAAGAGGGCATCAATTTTTT
CGAGTCTGGCT
>Nu_378_PART_B_Hap_3
AAGAGACCGAAAGATTCGTCCACTCCTTGTACTCTCGTTGCAGCGGCTCCAGCAGTTGCAGCGAGAGTCTGGATCAGAGGAGAAGGTGA
CGCGGCTGGTC
>Nu_378_PART_B_Hap_4
AAGGGACCGAAAGATGCGTCCAATCCTTGTACTCTCGTTGCAGCGGCTCCAGCAGTTGCAGCGAGAGTCTGGATCAGAGGAGAAGGTGA
CGCGGCTGGTC

Appendix C – Seq IDs for each specimen shown in Figure 1