

## Materials and Methods

### *Hair Analysis:*

Hair samples were sent to the Southwestern Institute of Forensic Sciences (Dallas, TX) for analysis. The samples were evaluated visually, stereoscopically, and by light microscopy to determine human or animal origin. Hairs that were classified as potential novel hominid were also evaluated for DNA typing potential by examining for root material. Only hairs that were not human in appearance and could not be identified as any other species were utilized in this study<sup>15-19</sup>.

Hairs were initially examined with an illuminated magnifying lens to observe gross characteristics such as coarseness, hair shaft profile and color. Select hairs were then measured to determine the approximate length. The selected hairs were examined using a Leica MZ 7.5 stereomicroscope at magnifications ranging from ~6.3×–75×. The stereoscopic examination was performed to examine the hair for roots and apparent tissue. Select hairs were then mounted in longitudinal whole mount microscopic preparations using Permount or xylene. Typically xylene was used for temporary mounts, but occasionally Permount was used for stiffer hairs. The whole mount slide preparations were examined microscopically at magnifications ranging from 50×–400× using a Leica CFM2 comparison microscope to determine morphological features of individual hairs. The hair maximum widths were measured using a calibrated reticule. The hairs were examined for a variety of microscopic features such as: medulla, pigmentation, cortical fusi, ovoid bodies, cuticle, and root and tip characteristics. The hair cuticle patterns of select hairs were cast by embedding the hair shafts in a thin layer of clear fingernail polish. The casts were examined microscopically. Photographs and photomicrographs were taken of select hairs using a Canon G3 digital camera and a camera adapter (model MM3XS) from Martin Microscope.

### *DNA Isolation*

Since the presence of normal human DNA contamination of submitted samples was a primary concern throughout this study, all samples were thoroughly cleaned in a manner consistent with forensic testing procedures. In order to further rule out contamination from human personnel and lab workers, samples from submitters and scientists working with the samples were collected for comparison with the results obtained in the various DNA tests.

Hair samples were then sorted into two groups for extraction at DNA Diagnostics. DNA from those samples containing 5-50 or more single hair roots were selected and the roots clipped into 1.5 mL microcentrifuge tubes. The hair roots were thoroughly cleansed with water and ethanol prior to extraction to remove any extraneous DNA.

Hair roots were placed in microcentrifuge tubes for DNA extraction and ATL buffer (Qiagen) was added. These samples were digested with proteinase K (PK, 20 mg/mL) and dithiothreitol (DTT, 1.0 M) at 56°C overnight, followed by a three-step organic extraction procedure using phenol:chloroform:isoamyl alcohol (25:24:1) with an additional PCI extraction. This process was followed by a butanol wash and buffer exchange/concentration into TE<sup>-4</sup> buffer (10 mM Tris,

0.1mM EDTA, pH 8.0) using Microcon<sup>®</sup>-100 ultrafiltration devices (Millipore, Billerica, MA)<sup>93-94</sup>.

The remaining unknown hairs with only 1-5 hair roots were sent to the North Louisiana Criminalistics Laboratory (NLCL, Shreveport, Louisiana) for DNA extraction and purification. The roots were cleaned with water prior to digestion. The cleaned roots were digested in ATL buffer (Qiagen), PK, and DTT at 56°C until completely dissolved, which generally was overnight. The DNA in this crude extract was purified using the EZ1<sup>®</sup> DNA Investigator Kit with cRNA (Qiagen) and eluted into TE<sup>-4</sup> on a BioRobot EZ1<sup>®</sup> (Qiagen).

Saliva swabs, blood swabs, and tissue cuttings (10 mg) were placed in microcentrifuge tubes for DNA extraction. The samples were extracted using the above mentioned organic method with the exception that DTT was not used during digestion.

Reference samples, in the form of buccal swabs from submitters who collected the unknown hair and tissue samples, were isolated using 50mM NaOH and heated to 100°C for 10 minutes followed by the addition of 1M Tris (pH 8.3)<sup>99</sup>. The DNA extracted at DNA Diagnostics was quantified using a Nanodrop<sup>™</sup> spectrophotometer (Thermo Scientific, Willington, DE). Hair samples sent to the NLCL were quantified by real time PCR using the Applied Biosystems Quantifiler<sup>®</sup> Human kit on an Applied Biosystems Prism<sup>®</sup> 7000 Sequence Detection System<sup>100</sup>. Samples that yielded DNA concentrations too low to use in standard testing had their DNA concentration augmented using multiple displacement amplification method per the manufacturer's instructions.

#### *Gender and Short Tandem Repeat (STR) Analysis*

Nuclear DNA from the selected samples, human references, and appropriate controls was amplified using the Promega PowerPlex<sup>®</sup> 16 kit<sup>20</sup> and an Applied Biosystems<sup>™</sup> 9600<sup>®</sup> Thermal Cycler according to manufacturer's suggested conditions. The PowerPlex<sup>®</sup> 16 is a STR multiplex system containing primers for the following loci: Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317, and D5S818. Amplified products were prepared and subjected to electrophoresis according to the PowerPlex<sup>®</sup> 16 kit instructions on an AB Prism<sup>®</sup> 310 Genetic Analyzer or an AB Prism<sup>®</sup> 377 DNA Sequencer. The samples were also amplified using AmpF $\ell$ STR<sup>®</sup> Yfiler<sup>®</sup> PCR Amplification Kit (Applied Biosystems by Life Technologies) that amplifies 17 Y-STR loci in a single PCR reaction using 5-dye chemistry. Amplified products were prepared and subjected to electrophoresis according to the AmpF $\ell$ STR<sup>®</sup> Yfiler<sup>®</sup> PCR Amplification Kit instructions on an ABI Prism<sup>®</sup> 310 Genetic Analyzer or an AB Prism<sup>®</sup> 377 DNA Sequencer.

After preliminary screening using PowerPlex<sup>®</sup> 16 and AmpF $\ell$ STR<sup>®</sup> Yfiler<sup>®</sup> at DNA Diagnostics Inc, aliquots of the extracted DNA samples were then sent to Family Tree DNA (Houston, TX) as blind samples labeled "unknown", where the amelogenin gene was amplified separately using proprietary primers. In addition, the PowerPlex<sup>®</sup> 16 amplification and the AmpF $\ell$ STR<sup>®</sup> Yfiler<sup>®</sup> PCR amplification were repeated using the extracted DNA across all samples. This was an effort to determine if the samples could be hominid or primate in origin as well as to rule out human contamination as would be indicated by a mixed profile.

### *Electron Microscopy:*

Electron microscopy was utilized on selected samples to visualize the DNA structure. Purified DNA from sample 26 (obtained from a tissue sample collected after the shooting of an unknown hominid) and a human control was sent to Texas A&M Microscopy and Imaging Center for structural analysis using electron microscopy. The DNA solution (estimated at 0.2 µg/mL) was mixed with ammonium acetate (0.25M) and a small amount of cytochrome C protein. The mix was placed as two droplets on Parafilm<sup>®</sup> and allowed to sit for 2 minutes and 4 minutes respectively. The surface film with the DNA was picked up with an EM grid and subjected to rotary shadowing with platinum<sup>70-74</sup>.

### *Mitochondrial DNA Sequencing:*

The samples were sequenced to determine their mitochondrial DNA haplotypes in order to identify the species of the samples and to delineate maternal origins<sup>21-28</sup>. Aliquots of purified DNA from all of the samples in this study, along with human controls to monitor for possible contamination, were shipped to Family Tree DNA. Proprietary methods were used to amplify the mitochondrial DNA genome. The DNA was amplified using 48 sets of human specific primer pairs that overlapped. Extra primers were developed and utilized in case of failure due to mutation. The amplicons were sequenced on an Applied Biosystems<sup>®</sup> 3130xl Genetic Analyzer. Those samples failing to amplify for HV1 were then screened for bacteria using proprietary procedures. Any of the unknown samples testing positive for bacterial DNA in the mitochondrial WGA amplicons were removed from the study.

### *Nuclear DNA Analysis:*

Initially amplification of the gender-specific locus, Amelogenin, a core forensic locus that is specific to apes and humans was performed<sup>95-98</sup>. Human and primate DNA typically have amplicons of 106 base pairs in length for the female determinate allele which is found in the homologous X region of the Y chromosome and 112 base pairs for the male determinate allele on the non-homologous X region of the Y chromosome. This easily distinguishes male and female DNA profiles. Non primate DNA, when amplified at this locus, may, at times, yield a peak at 103 bases (100 bases for capillary instruments) for some animals, (personal observation) which made it an excellent locus for this screening.

Aliquots of purified DNA from all samples were sent to SeqWright (Houston, TX), for nuclear DNA sequencing. This laboratory was not informed of the origins of the samples. The MC1R, amelogenin, TAP1 and MHY16 heavy chain loci examined in this project were sequenced. SeqWright also designed the majority of the proprietary primers used in this study so the actual sequence lengths and primer positions would be unknown to the submitting laboratories (DNA Diagnostics and NLCL). This was done in order to assure that the DNA had not been manipulated. The sequencing was performed on an Applied Biosystems 3130xl Genetic Analyzer using BigDye<sup>®</sup> technology from Applied Biosystems.

### *Whole Human Genome SNP analysis:*

Aliquots of purified DNA were sent to the Genomics Core Laboratory at the University of Southern California (Los Angeles, CA) for Illumina Bead Array<sup>®</sup> 2.5 million SNP array testing.

They were not informed of the origins of the samples. The testing was performed per the patented Illumina<sup>™</sup> protocol for Bead Array testing<sup>69</sup>.

*Whole Genome Sequencing:*

Three samples were selected (26, 31, 140) for whole genome sequencing. The samples were sent to UT Southwestern Medical Center in Dallas, TX for next generation whole genome sequencing. The sequencing was performed using the HiSeq 2000 by Illumina and visualized using Sequence Viewer Software by Illumina. CASAVA 1.8 was used for secondary analysis, such as generation of fastq files and alignments to reference genomes. Custom scripts were used to extract reads showing good alignment to specific chromosomes of the reference genome. Using CLC Bio Genomic Workbench version 5.1, the extracted reads were assembled to create a consensus sequence using the targeted chromosome as a reference; also, de novo assembly using the extracted reads was performed to yield contigs. Both the assembled reads and the de novo assembly contigs were fed to BLAST version 2.2.26 ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)) with the following settings: database Nucleotide collection (nr/nt), optimized for highly similar sequences (megablast). The BLAST tree view (Distance tree of results) was used to generate phylogeny trees (produced using BLAST pairwise alignments). The raw sequence data used in this study is provided in the supplemental information.