illumina®

Analysis of Skeletal Remains Using Deep mtDNA Sequencing

MiSeq® System aids in identification of skeletal remains by sequencing mtDNA hypervariable regions.

Introduction

Forensic researchers face significant challenges in identifying unknown remains found in mass graves, such as those located in the war-torn countries of Vietnam, Kosovo, and Somalia. Dental records and circumstantial evidence are often inconclusive. DNA collected from skeletal samples is typically highly degraded, contaminated, and present in small quantities. The poor quality of these samples limits the success of conventional short tandem repeat (STR) typing techniques. In contrast, mitochondrial DNA (mtDNA) sequencing is quite useful in analyzing these samples, offering the following advantages:

- It is abundant and present in hundreds of copies per cell compared to nuclear DNA.
- It has a low mutation rate.
- It is maternally inherited and does not undergo sexual recombination, enabling maternal relationships to be established across generations.

For years, forensic laboratories have used capillary electrophoresis (CE)-based Sanger sequencing to perform mtDNA analysis. While the full mtDNA genome is approximately 16 kb, the hypervariable I and II (HVI and HVII) regions are usually the targets of mtDNA sequencing for human identification. These regions are located within a 1.1 kb region called the "D-loop." Capable of providing deeper coverage of specific regions of interest, next-generation sequencing (NGS) systems enable mtDNA sequencing to be performed at a high throughput, using simpler workflows.

The Illumina MiSeq System is a fully integrated NGS instrument, providing a rapid workflow and streamlined library preparation that is ideal for sequencing small genomes and amplicon samples, such as the D-loop region of mtDNA. This application note describes the use of the MiSeq System for deep sequencing of the HVI and HVII regions in mtDNA and how this technique is being used to aid the identification of human skeletal remains found in mass graves in Vietnam (Figure 1).

mtDNA Sequencing of Bone Fragments

Forensic scientists used the MiSeq System to sequence bone fragments uncovered at a mass grave site in Vietnam after DNA analysis of the samples using CE-based sequencing failed. The process began with the extraction of total DNA from blood and bone samples using established methods. Four individual amplicons spanning the HVI and HVII regions were amplified from mtDNA using PCR primers specific to the D-loop/Control region[1,2] (Figure 2). Library preparation was performed using the Nextera® XT DNA Library Preparation Kit. The samples were then sequenced on the MiSeq

Figure 1: Recovery of Human Remains—MiSeq technology is helping to identify human skeletal remains such as these, recovered in Vietnam. There are an estimated 300,000 unidentified human skeletal remains lying in mass graves across Vietnam.

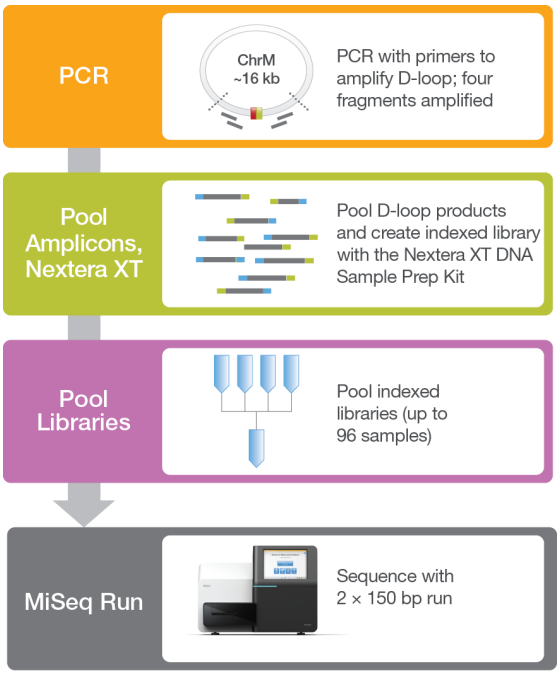


Figure 2: NGS mtDNA Sequencing Workflow—Specific amplicons spanning the HVI and HVII regions were sequenced on the MiSeq System using a 2 × 150 bp run protocol.

System using a 2 × 150 bp run protocol. Raw reads were aligned to the mtDNA reference using BWA,[3] and variants were screened against the mtDNA revised Cambridge reference sequence (rCRS)[4] and called using the Genome Analysis Toolkit (GATK).[5,6] The frequencies of each variant were calculated and compared to establish shared maternal lineage.

Analysis and Results

Concordant Sequencing Results of HVI/II Regions

Using this method, ultradeep sequencing coverage (>10,000 reads) was reproducibly obtained for all samples. All variants previously reported in control DNA 9947A (Promega) were identified at high frequencies (Figure 3). There was concordance in variant positions and frequencies between technical replicates. Results for blood samples were concordant with those previously analyzed using CE-based Sanger sequencing.

Human Identification from Buried Remains

Total DNA was extracted from a bone sample (995A) recovered from a grave at the Southern Battlefield Cemetery in Vietnam. A corresponding blood sample (995B) was also drawn from the presumptive brother of the interred individual and sequenced using Illumina sequencing technology. The variant positions and frequencies in the HVI/II regions were compared (Figure 4). Results indicated that the 2 samples may share maternal lineage, supporting the hypothesis of kinship.

MiSeq mtDNA Protocols and Analysis Tools

mtDNA library preparation, sequencing, and analysis workflows and tools are available for the MiSeq System. These include a protocol for clean, intact DNA samples,[7] as well as one for difficult samples,[8] such as bone, where sequencing of the mtDNA D-loop hypervariable region is critical. Both incorporate data analysis with MiSeq Reporter and variant calling with the mtDNA Variant Analyzer Software tool. These same analysis capabilities, along with several usability enhancements, are also available in BaseSpace®, the Illumina genomics computing platform.

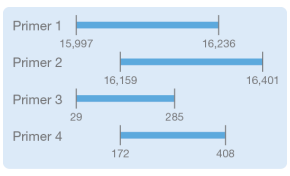
		Sample 917 (blood)		Sample 926 (blood)		Sample 945 (blood)		Sample 995 (bone #1)	Sample 995 (bone #2)	9947A (Control)
DLOOP	GEN	917B_1	917B_2	926B_1	926B_2	945B_1	945B_2	995A_2	995B_2	9947A_1
913	G->A	0.001	0.004	0.001	0.006	0.975	0.962	0.017	0.002	0.001
973	T->C	0.918	0.961	0.011	0.025	0.024	0.020	0.028	0.032	0.027
976	C->T	0.002	0.003	0.000	0.003	0.988	0.948	0.003	0.003	0.000
1007	C->T	0.948	0.938	0.987	0.983	0.979	0.982	0.937	0.976	0.014
1050	C->T	0.983	0.940	0.042	0.002	0.001	0.002	0.000	0.000	0.004
1053	A->G	0.002	0.004	0.002	0.003	0.003	0.003	0.966	0.992	0.003
1055	T->C	0.001	0.041	0.001	0.024	0.001	0.021	0.998	0.998	0.001
1062	C->T	0.001	0.005	0.967	0.972	0.006	0.001	0.001	0.000	0.004
1078	C->T	0.002	0.001	0.914	0.970	0.004	0.001	0.000	0.001	0.005
1081	T->C	0.002	0.008	0.002	0.003	0.994	0.990	0.004	0.006	0.002
1095	T->C	0.003	0.017	0.005	0.005	0.002	0.004	0.002	0.005	0.998
1146	T->C	0.995	0.979	0.007	0.009	0.002	0.007	0.008	0.011	0.001
1426	A->G	0.980	0.991	0.987	0.995	0.953	0.998	0.998	0.995	0.006
1446	A->G	0.006	0.001	0.004	0.003	0.006	0.003	0.002	0.004	0.991
1503	C->T	0.995	0.996	0.992	0.990	0.993	0.992	0.990	0.991	0.000
1548	T->C	0.017	0.013	0.009	0.003	0.002	0.003	0.004	0.001	0.992
1552	T->C	0.002	0.002	0.003	0.003	0.991	0.992	0.004	0.002	0.001
1557	T->C	0.004	0.004	0.011	0.025	0.003	0.004	0.994	0.998	0.004
1567	A->G	0.007	0.003	0.015	0.002	0.006	0.001	0.004	0.003	0.974
1616	A->G	0.877	0.986	0.980	0.993	0.833	0.994	0.992	0.854	0.836

Figure 3: SNP Results from MiSeq System Show Strong Concordance— Raw frequencies are shown for each SNP across the D-loop. Blood and control results were concordant with samples previously sequenced with Sanger sequencing.

Conclusions

The Nextera XT DNA Library Preparation Kit and MiSeq System provide robust data for human identification. Sequence derived from the HVI and HVII regions of the mtDNA D-loop region was used to establish that samples from 2 individuals shared a maternal lineage. The Illumina MiSeq System provides a reliable, reproducible method for determining human relationships from degraded samples that are not amenable to analysis by other methods.

DLOOP	GEN	995A_2	995B_2	9947A_1
1007	C->T	0.937	0.976	0.014
1053	A->G	0.966	0.992	0.003
1055	T->C	0.998	0.998	0.001
1095	T->C	0.002	0.005	0.998
1426	A->G	0.998	0.995	0.006
1446	A->G	0.002	0.004	0.991
1503	C->T	0.990	0.991	0.000
1548	T->C	0.004	0.001	0.992
1557	T->C	0.994	0.998	0.004
1567	A->G	0.004	0.003	0.974
1616	A->G	0.992	0.854	0.836



Note: Two pairs of amplicons cover each of the HV regions, with overlap of approximately 240 bp within HVI and 110 bp overlap within HVII.

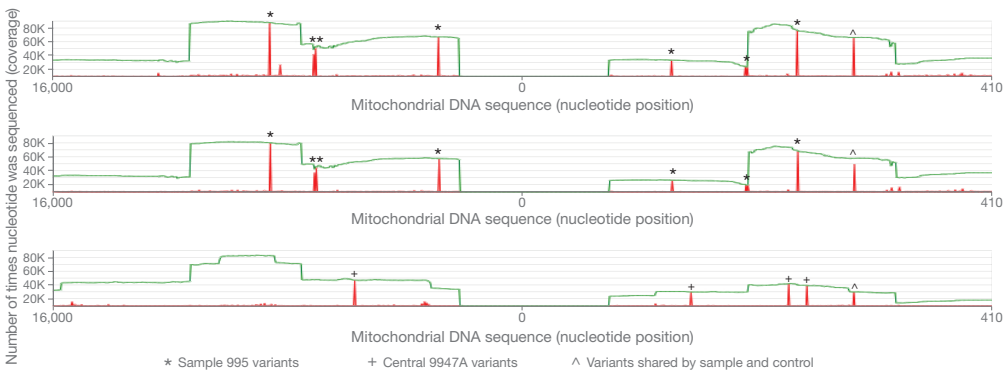


Figure 4: Illumina Sequencing of Bone Samples Confirms Putative Relationships— DNA from a bone sample recovered from a grave (995A) and a blood sample (995B) were compared. The high degree of concordance in identified SNPs and profile characteristics suggests the 2 individuals are brothers.

Learn More

To learn more about the MiSeq System, visit: www.illumina.com/miseq

For more on Illumina sequencing technology for forensic applications, visit: www.illumina.com/areas-of-interest/forensic-genomics.html.

To access mtDNA Processor and mtDNA Variant Analyzer in BaseSpace, log in to BaseSpace at www.Illumina.com, go to the Apps tab, and search for mtDNA Processor and mtDNA Variant Analyzer.

References

1. Vigilant L, Pennington R, Harpending H, Kocher T, and Wilson A. Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci.* 1989; 86: 9350–9354.
2. Wilson M, DiZinno J, Polanskey D, Replogle J, and Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med.* 1995;108: 68–74.
3. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010; 26: 589–595.
4. Andrews R, Kubacka I, Chinnery P, Lightowlers R, Turnbull D, and Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999; 23: 147.
5. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20: 1297–303.
6. DePristo M, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43: 491–498.
7. Human mtDNA Genome Protocol (support.illumina.com/downloads/human_mtdna_genome_guide_15037958.ilmn). Accessed 26 January 2016.
8. Human mtDNA D-loop Hypervariable Region Protocol, (support.illumina.com/downloads/human_mtdna_d_loop_hypervariable_region_guide_15034858.ilmn). Accessed 26 January 2016.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

illumina®