Mitochondrial DNA Sequencing of the Control Region and Whole Genome

An assessment of library prep methods and data analysis as part of an integrated workflow

Introduction

The ForenSeq™ mtDNA Control Region Kit targets the hypervariable regions of the mitochondrial genome (mtGenome) as part of an integrated next-generation sequencing (NGS) workflow. The kit offers a flexible solution for mitochondrial (mtDNA) analysis that efficiently and comprehensively interrogates degraded and otherwise challenging samples. With the introduction of the ForenSeq mtDNA Whole Genome Kit, Verogen is empowering laboratories to explore the entire mtGenome for enhanced haplotype resolution and databasing capability.

Laboratories are frequently challenged with processing low-quantity, low-quality samples for forensic casework. Sequencing mtDNA, particularly in cases of missing persons and disaster victim identification (DVI), addresses this challenge with the transformative capability to generate larger, more informative data sets at a lower cost per nucleotide while extending the lower limits of sample quality. The ForenSeq mtDNA Whole Genome Kit and ForenSeq mtDNA Control Region Kit have the same core primer set, providing comparable control region analysis. The key difference is that for whole genome, the primer set is expanded for coverage across the entire molecule and maximum variant detection. This technical note describes mtDNA sequencing of the control region and whole genome and examines the workflows, specifications, study parameters, and performance data that comprise the ForenSeq mtDNA portfolio.

Complementary mtDNA Analysis Methods

In collaboration with mtDNA experts from around the world, Verogen developed each ForenSeq mtDNA kit as the library prep component of a fully supported, end-to-end workflow that includes sequencing on the MiSeq FGx® Sequencing System and data analysis in ForenSeq UAS v2.0 (Figure 1). The MiSeq FGx System efficiently clusters and sequences libraries while the software bundles rapid data analysis with robust visualization and reporting features for easy interpretation and export. Taken together, this portfolio provides clear results on an interface designed for mtDNA applications and enhanced for usability. Data from ForenSeq UAS v2.0 are compatible with the Integrative Genome Viewer (IGV), EMPOP, CODIS, and other downstream tools. Simply use the software reporting features to select applicable positions and convert sequencing data into a database-friendly file format, such as CMF.[1,2]

Given the relatively small size of the control region (1.2 kb) and high concentration of variation, sequencing this region has traditionally served as an efficient method of mtDNA analysis. However, the availability of the ForenSeq mtDNA Whole Genome Kit warrants a side-by-side look at both control region and whole genome sequencing. Both kits deliver robust, reliable data and streamlined NGS workflows, with whole genome benefitting laboratories looking to maximize mtDNA capabilities for population studies or severely degraded DNA. The whole genome kit also offers a cost advantage by generating more data from the same input.

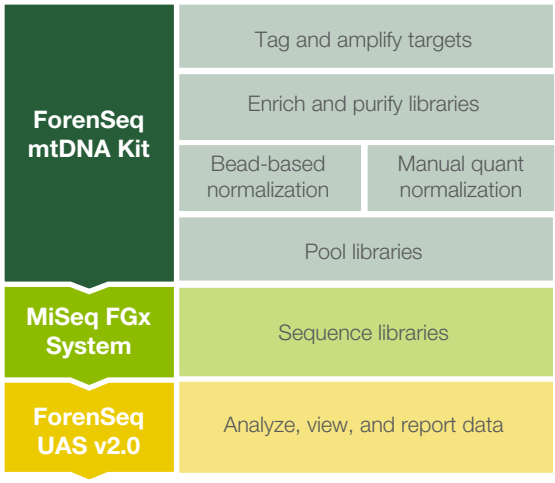

Figure 1: Integrated NGS workflow delivers reliable results—The ForenSeq mtDNA workflow includes library prep with a ForenSeq mtDNA library prep kit, sequencing on the MiSeq FGx System, and analysis in ForenSeq UAS v2.0.

The ForenSeq mtDNA Control Region Kit offers the convenience of a concise data set that can be directly uploaded to CODIS, which accepts only the hypervariable I (HVI) and hypervariable II (HVII) regions of the mtGenome. Laboratories requiring only HVI and HVII data can start using or continue to use the control region kit exclusively and with confidence. Alternatively, they can onboard the ForenSeq mtDNA Whole Genome Kit as a complementary tool and use the control region kit for CODIS uploads, or simply transition to whole genome.

Matched Amplicon Designs

To ensure sensitive and accurate analysis of mtDNA, both ForenSeq mtDNA kits leverage a tiled amplicon design that generates sequencing libraries and rich data from less than 100 pg genomic DNA (gDNA) input (Figure 2). Both preps are rapid, converting the input gDNA samples into sequence-ready libraries in a single day with fewer than 2 hours of hands-on time. To enhance flexibility for low-throughput or high-throughput requirements and automation, both kits have two options for library normalization: bead-based or manual quantification.

The ForenSeq mtDNA Control Region Kit uses over 120 primers designed against the most recent and well-curated mtDNA variant and frequency data to generate 18 primary amplicons that span the complete mtDNA control region. At less than 150 bp in length, all primary amplicons are small, a design feature that maximizes performance on degraded samples. A minimal overlap of at least 3 bp facilitates analysis and prevents sequence gaps due to bioinformatic trimming. Building on the tiled amplicon design of the control region kit, the whole genome kit is suited for analysis of the entire 16,569 bp human mtGenome. This fully kitted assay includes reagents to amplify approximately 500 small amplicons and cover the entire genome in two PCR reactions (Table 1).[3]

Table 1: Library Prep Kit Specifications

Specification	ForenSeq mtDNA Control Region Kit	ForenSeq mtDNA Whole Genome Kit
Target	1200 bp control region	16,569 bp mtGenome
Number of amplicon pools	2	2
Number of amplicons	18	245
Mean amplicon size	118 bp	131 bp
Minimum amplicon overlap	≥ 3 bp	≥ 3 bp
Number of primers	> 120	> 500
Input DNA recommendation	100 pg gDNA*	100 pg gDNA*
Multiplexing capacity	3–48 samples	16 samples
Reagent kit	MiSeq FGx Reagent Micro Kit	MiSeq FGx Reagent Kit
Protocol	ForenSeq library prep	ForenSeq library prep
Hands-on time	~90 minutes	~105 minutes

* Each 100 pg gDNA sample is divided into two 50 pg reactions.

Comparable Control Region Metrics

Verogen evaluated various performance metrics to assess control region data generated from both ForenSeq mtDNA kits. These metrics included PCR inhibitor resistance, concordance with control DNA, and coverage. Verogen extensively tested the more recent ForenSeq mtDNA Whole Genome Kit and compared the results to existing data for ForenSeq mtDNA Control Region Kit. For all metrics, the comparison demonstrated a high degree of equivalence between the control region and whole genome kits, establishing both as equally desirable choices depending on the intended application. Although the whole genome kit provides more data, both kits offer virtually identical data of exceptional quality.[4]

PCR Inhibitor Resistance

To evaluate PCR inhibitor resistance, Verogen scientists tested data recovery in the presence of calcium, humic acid, and *E. coli* DNA to 100 pg gDNA and confirmed that all bases in the control region were correctly called. Coverage (read counts) in samples containing inhibitors were compared to the control sample, which did not contain any inhibitors. For both the ForenSeq mtDNA Control Region Kit and the ForenSeq mtDNA Whole Genome Kit, results indicated similar resistance to PCR inhibitors across the control region (Figure 3). The study also used BWA alignment to analyze the sample containing *E. coli* for the presence of microbial DNA in sequencing data. These results indicated zero bacterial sequence. Verogen also compared the variant call rates for inhibited samples to the positive control. A > 95% call rate was observed at all conditions excepting 100 ng/µl humic acid, where calls rates were 62% for control region samples and 80% for whole genome samples (Figure 3).

Comparison of Observed Variants to Control DNA

To evaluate concordance, Verogen compared control region data from both kits to five control DNA samples: three Coriell samples and two NIST Standard Reference Material (SRM) samples. The NIST SRM samples are well-characterized and intended to provide quality control

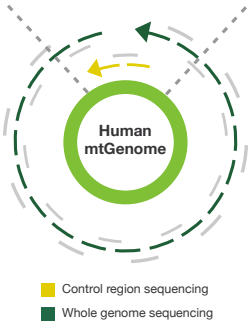

Figure 2: Tiled amplicon design improves data recovery—Short amplicons with minimal overlap span the control region or circle the entire mtGenome in a tiled format to achieve balanced coverage.

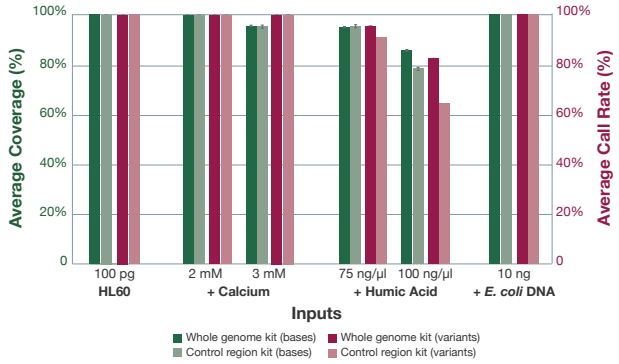

Figure 3: Control region coverage in the presence of PCR inhibitors—ForenSeq mtDNA kits have a common buffer system that combats inhibitors typical in forensic samples to provide a robust environment for successful amplicon generation. PCR inhibitors added to input DNA at concentrations indicated along the x-axis showed coverage and call rate.

when amplifying and sequencing human mtGenome sequences. Verogen generated and sequenced three replicates for each sample and prep kit on a MiSeq FGx System. Data were analyzed in ForenSeq UAS v2.0 with the Verogen mtDNA Control Region Analysis Method or Verogen mtDNA Whole Genome Analysis Method, as applicable, using the default analysis thresholds. Between the expected variants and the variants generated by both prep kits, 100% concordance was observed (Table 2).[5]

Optimized Workflow Integration and Support

As commercial-grade, all-inclusive reagent systems, the ForenSeq mtDNA kits offer significant advantages for performance, reliability, and flexibility compared to home-brew and third-party assays. Based on the simple and automation-friendly ForenSeq library prep workflow, the ForenSeq mtDNA kits offer an accessible entry point for laboratories that want to transition from custom, CE-based mtDNA protocols to NGS-based methods. Laboratories using third-party assays, which often employ piecemeal chemistry, can benefit from the fully kitted and tested format of Verogen library prep and sequencing reagents, optimizing batch-to-batch reliability and simplifying inventory management. Laboratories familiar with the ForenSeq DNA Signature Prep Kit can leverage existing workflows to streamline the introduction of mtDNA capabilities.

Table 2: Comparison of Two Library Prep Methods for Control Region Variant Concordance

			ForenSeq mtDNA Control Region Kit		ForenSeq mtDNA Whole Genome Kit	
Sample	Input (pg)	Expected Variants	Concordance (%)	Observed Variants	Concordance	Observed Variants
CHR	100	64Y 73G 195C 204C 207A 263G 309.1C 315.1C 16183C 16189C 16193.1c 16193.2c 16223T 16278T 16519C	100	64Y 73G 195C 204C 207A 263G 309.1C 315.1C 16183C 16189C 16193.1c 16193.2c 16223T 16278T 16519C	100	64Y 73G 195C 204C 207A 263G 309.1C 315.1C 16183C 16189C 16193.1c 16193.2c 16223T 16278T 16519C
9947A	100	93G 195C 214G 263G 309.1C 309.2C 315.1C 16311C 16519C	100	93G 195C 214G 263G 309.1C 309.2C 315.1C 16311C 16519C	100	93G 195C 214G 263G 309.1C 309.2C 315.1C 16311C 16519C
HL60	100	73G 150T 152C 263G 295T 315.1C 489C 16069T 16193T 16278T 16362C	100	73G 150T 152C 263G 295T 315.1C 489C 16069T 16193T 16278T 16362C	100	73G 150T 152C 263G 295T 315.1C 489C 16069T 16193T 16278T 16362C
GM03798	100	263G 315.1C 16357C 16519C	100	263G 315.1C 16357C 16519C	100	263G 315.1C 16357C 16519C
GM10472A	100	73G 185A 228A 263G 295T 315.1C 462T 482C 489C 16069T 16126C 16292T	100	73G 185A 228A 263G 295T 315.1C 462T 482C 489C 16069T 16126C 16292T	100	73G 185A 228A 263G 295T 315.1C 462T 482C 489C 16069T 16126C 16292T

Offering two modes of operation, Forensic Genomics and Research Use Only (RUO), the MiSeq FGx System sequences all mtDNA libraries, regardless of assay origin. Provided the libraries are sequenced on the MiSeq FGx System, ForenSeq UAS v2.0 can analyze the resulting data, facilitating comparative studies and simplifying the transition from alternative mtDNA assays to a ForenSeq mtDNA kit. The ForenSeq mtDNA kits, MiSeq FGx System, and ForenSeq UAS v2.0 constitute the fully integrated, fully supported Verogen mtDNA workflow.

Built for scalability to support all throughput requirements, each workflow component is designed, developed, and tested together to maximize performance on forensic samples and facilitate operations from library prep through data analysis and beyond. Verogen backs all components with validation and implementation services and exceptional support: advice, training, troubleshooting, and other technical support requirements. Reagent manufacturing is quality controlled for reproducible performance and the convenience of complete kitting. Leveraging an advanced assay design and gold-standard sequencing chemistry, the ForenSeq mtDNA kits are part of a curated portfolio that provide a range of application choices, freeing laboratories from quality and support limitations.

Conclusion

The ForenSeq mtDNA Control Region and ForenSeq mtDNA Whole Genome Kits produce highly concordant data across a spectrum of performance parameters. Data from the control region, which are common to both kits, are entirely comparable. By selectively interrogating the HVI and HVII regions, the control region kit offers a straightforward path from sample to sequencing to CODIS. Encompassing the rest of the mtGenome sequence, the whole genome kit grants additional power of discrimination while extending possibilities for severely degraded DNA and population studies. In either case, the ForenSeq mtDNA kits present an ideal solution for laboratories seeking to transform the most fragile samples into powerful results. Laboratories seeking to up-level or introduce mtDNA NGS capabilities will find that these ready-to-use kits consistently detect haplotype variants within mtDNA while saving the time and effort of identifying targets, designing primers, and optimizing other assays.

References

1. Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov, “Integrative genomics viewer," *Nature Biotechnology* 29 (January 2011): 24–26, https://doi.org/10.1038/nbt.1754.
2. Parson, Walther, and Arne Dür, “EMPOP—A forensic mtDNA database,” *Forensic Science International: Genetics* 1, no. 2 (June 2007): 88–92, https://doi.org/10.1016/j.fsigen.2007.01.018.
3. MITOMAP: A Human Mitochondrial Genome Database (accessed July 10, 2020), http://www.mitomap.org.
4. Walichiewicz, Paulina, Justin Eagles, Anthony Daulo, Meghan Didier, Chris Edwards, Keenan Fleming, Yonmee Han, et al., *Performance Evaluation of the ForenSeq mtDNA Control Region Solution*, https://verogen.com/wp-content/uploads/2019/11/Mito-ISHI-Poster_final.pdf.
5. Riman, Sarah, Kevin M Kiesler, Lisa A Borsuk, and Peter M Vallone, “Characterization of NIST human mitochondrial DNA SRM-2392 and SRM-2392-I standard reference materials by next generation sequencing,” *Forensic Science International: Genetics*, 29 (July 2017): 181–192, https://doi.org/10.1016/j.fsigen.2017.04.005.

Verogen • 1.833.837.6436 toll-free (North America) • +1.858.451.4187 tel • +44 (20) 399 28411 (United Kingdom) • info@verogen.com • www.verogen.com

Pub No. VD2020010 Current as of September 2020

VEROGEN