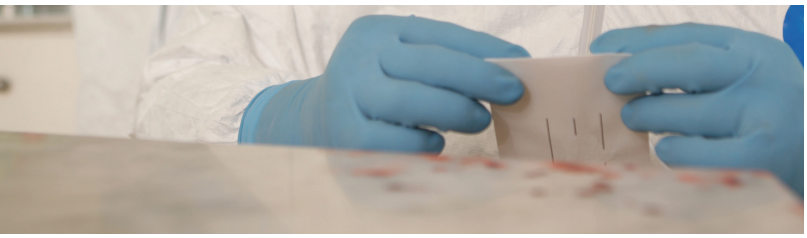
applied biosystems



Solutions for mtDNA analysis

NDIS Approved

Identify missing persons or analyze remains

In forensic casework, the high copy number per cell of mitochondrial DNA (mtDNA) is useful in the context of challenging samples that fail to produce an autosomal STR profile. When analyzing the mtDNA control region or whole genome from unidentified remains (bones and teeth) or crime scene samples (hair roots or shafts), full amplicon coverage and uniform variant calling are key factors that contribute to high-quality genetic results for successful human identification.

The Applied Biosystems™ Precision ID mtDNA Whole Genome Panel is an innovative approach to mtDNA sequencing, specifically developed for forensic applications. This mtDNA tiling approach, using amplicons that are only 163 bp in average length, assists with obtaining optimal mitochondrial genome (mtGenome) coverage from highly compromised, degraded samples such as hair shafts, teeth, and bones (Figure 1). The Precision ID mtDNA Whole Genome Panel, using the control region data for analysis, is approved for inclusion in the National DNA Index System (NDIS) database.

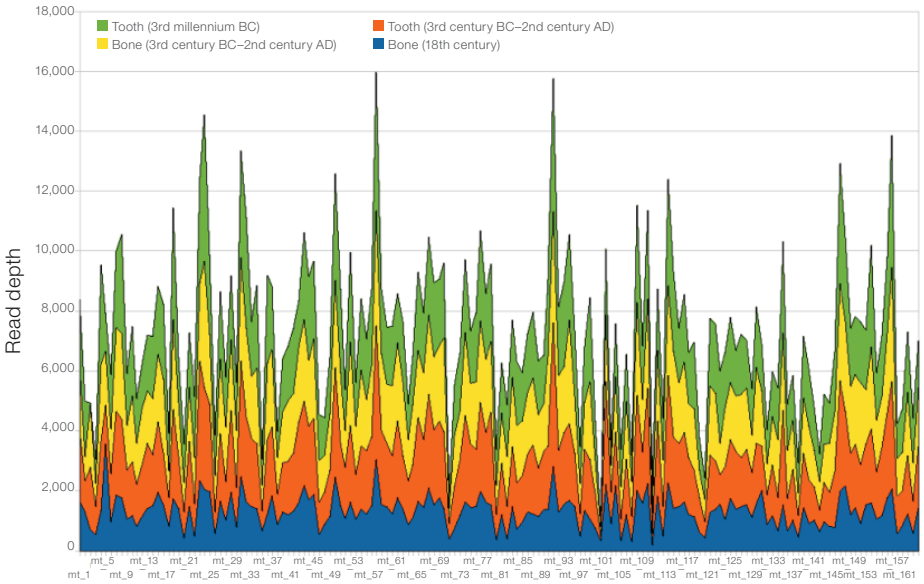


Figure 1. Sequence coverage of degraded samples. Whole genome coverage of 4 samples of ancient tooth and bone was generated using 1 ng of input DNA, automated library preparation using the Ion Chef™ System, and sequencing on the Ion S5™ System. The small amplicon design and the inclusion of degenerate primers in the Precision ID mtDNA Whole Genome Panel increase success with degraded and low amounts of template DNA.

ThermoFisher SCIENTIFIC

The Applied Biosystems™ Precision ID mtDNA Control Region Panel is based on the same tiling approach used in the Precision ID mtDNA Whole Genome Panel. This targeted panel spans the entire 1.2 kb control region, which encompasses HV-I, HV-II, and HV-III, with the same optimal small amplicon design ideal for performance with degraded forensic samples. In addition, both panel designs leverage primer degeneracy in known variable regions to help ensure robust performance across diverse population samples.

The Applied Biosystems™ Precision ID NGS System for human identification can help you solve tough cases by getting more information from your challenging samples. By combining the Ion Chef™ Instrument, Ion GeneStudio™ S5 systems, Precision ID mtDNA panels, and Applied Biosystems™ Converge™ Software, adopting next-generation sequencing (NGS) of mtDNA in your laboratory is now simpler than ever. Precision ID technology applies the simplicity of PCR to targeted sequencing, enabling tens to thousands of genomic regions to be amplified simultaneously in a single PCR tube. With as little as 2 pg of DNA input, you can quickly and easily analyze complete mitochondrial profiles from these challenging samples with as few as 5 pipetting steps and 45 minutes of hands-on time.

Table 1. Precision ID mtDNA panel specifications.

	Precision ID mtDNA Whole Genome Panel	Precision ID mtDNA Control Region Panel
Target	16.569 kb mtGenome	1.2 kb (control region 16,024–576)
Amplicon length	Average length of 163 bp with amplicon overlap of 11 bp	Average length of 153 bp with amplicon overlap of 18 bp
Primer pool size	2 primer pools of 81 primer pairs each Primer pool 1 has 119 additional degenerate primers* Primer pool 2 has 164 additional degenerate primers*	2 primer pools of 7 primer pairs each Primer pool 1 has 45 additional degenerate primers* Primer pool 2 has 68 additional degenerate primers*

* Degenerate primers were designed for primers that contained a single-nucleotide polymorphism (SNP) with a global or specific population frequency of ≥10% from the 1000 Genomes Project data set, or had a MITOMAP observed count of 700 or greater.

Performance

- Small amplicons enable robust amplification with degraded samples [1] (Table 1)
- Whole genome haplotypes generated from hair shafts, hair roots [2], and blood stains [3]
- Complete mtGenome coverage with as little as 2 pg of input (Figure 2)
- >99.99% specificity and accuracy compared to Sanger sequencing (Figure 3)
- Heteroplasmy detection down to ~10%

Simplicity

- Automated library and template preparation using the Ion Chef System
- A single assay provides both whole genome and control region coverage
- Automated mtDNA alignment and variant calling, and haplogroup analysis, with Converge Software
- Export compatibility for EMPOP and CODIS databases

Sample	Input (pg)	Uniformity (%)	Average amplicon coverage	Number of variants detected
HG1389	100	91.6	2,008	**90.8 (98.6%)**
	10	89.6	2,413	**91.5 (99.5%)**
	5	89.1	2,293	**91.5 (99.5%)**
	2	91.2	2,318	**90.8 (98.6%)**
HG1260	100	93.4	1,935	**89 (100%)**
	10	92.0	2,528	**89 (100%)**
	5	93.1	2,243	**89 (100%)**
	2	91.1	2,430	**89 (100%)**

Figure 2. Whole genome mitochondrial panel sensitivity. Coverage across the genome and uniformity (percentage of amplicons less than 0.2x mean amplicon coverage) remained stable across the input amounts (average of 4 runs), as did the ability to accurately call known variants.

Sensitivity	Specificity	Accuracy
98.5%* (n = 132)	99.997%** (n = 66,144)	99.994% (n = 66,276)

* 2 false negatives: 1 missed 309.2 and reported 2841T>A in GM10742.
** 2 extra variants: 2 different samples with 309del at about 50% frequency.

Figure 3. Concordance with Sanger sequencing. Compared to Sanger sequencing, variant calling with SRM 2392 samples showed little discordance and was limited primarily to the 309 position. Samples include HL-60, CHR, GM009947A, and GM10742 (n is the total number analyzed).

Scalability

- Multiple chip formats to meet your throughput needs
- Multiplex up to 32 samples for whole genome analysis or up to 56 for control region analysis (Table 2)

Speed

- Less than 45 minutes of hands-on time for a DNA-to-data targeted sequencing workflow (Figure 4)
- Sample to data in less than 2 days

Library preparation

Genomic DNA (as little as 2 pg) isolated from a sample is converted to a sequencing library by targeted amplification of regions of interest. The Precision ID library preparation workflow may be performed manually or can be completely automated for up to 8 samples per run for 1- or 2-pool panel designs on the Ion Chef System. Requiring less than 2 pipetting steps per sample, library preparation on the Ion Chef System reduces hands-on time and variation often seen with manual library preparation, generating pooled libraries ready for downstream template preparation.

Template preparation

Libraries that are prepared manually or by automation are clonally amplified on the Ion Chef System by emulsion PCR of library molecules captured on beads. The Ion Chef System automates all template preparation steps, including creating the emulsion mixture, performing the PCR, carrying out the post-PCR purifications, and finally loading the purified templated beads onto the Ion S5™ chips. The prepared chips are ready for sequencing on the Ion GeneStudio S5 systems.

Sequencing

A sequencing run on the Ion GeneStudio S5 systems is initiated by loading a reagent cartridge, buffer, cleaning solution, and waste container. The Ion S5 chip is then loaded and the run is started. The addition of nucleotides by the DNA polymerase results in the production of hydrogen ions; the change in pH is converted to sequencing signals through ion-sensitive wells that hold the templated beads.

Table 2. Sample throughput for Precision ID mtDNA panels.

		Ion GeneStudio S5 series systems		
		Ion 510 Chip	Ion 520 Chip	Ion 530 Chip
Reads		2–3 million	4–6 million	15–20 million
Samples per chip*	Precision ID mtDNA Whole Genome Panel	–	25	32
	Precision ID mtDNA Control Region Panel	37	56	–

* Recommendations are based on in-house determination of the number of samples that can be multiplexed while still achieving a minimum coverage of 100x of all mtDNA amplicons. Individual lab results may vary depending on workflow used and customer requirements. For automated library preparation, 32 is the maximum number of samples that can be run on a chip, which is limited by the number of barcodes available in the DL8 configuration.

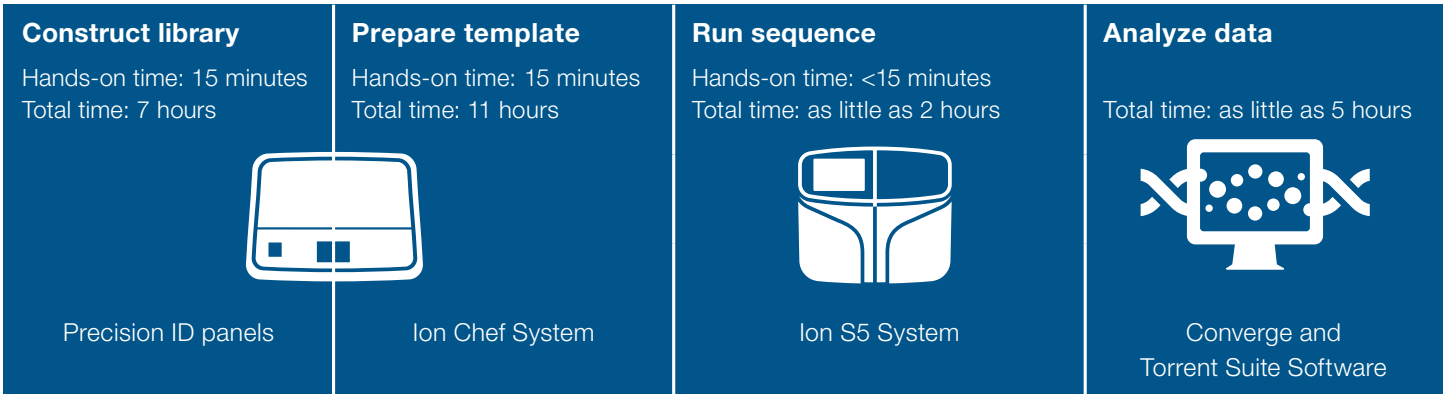


Figure 4. DNA-to-data workflow using Precision ID panels.

Data analysis

Analysis of the mtGenome can be challenging due to complex alignments, the presence of mtDNA heteroplasmy, and insertions and deletions present throughout the genome that may impact the accuracy of variant calling. The Converge NGS Data Analysis module automates mtDNA analysis, leveraging optimized base calling, alignment, and quality filtering algorithms specific for the Precision ID mtDNA panels. These solutions give forensic DNA laboratories the flexibility to detect variation within noncoding control sequences using the Precision ID mtDNA Control Region Panel, or to take advantage of the genetic diversity of full mtGenome sequence data using the Precision ID mtDNA Whole Genome Panel.

Phylogenetically guided sequence alignment is reported to produce more accurate variant calling for mitochondrial analysis. NGS reads from the BAM files are first mapped to nodes in PhyloTree [4] and then realigned using a custom Smith-Waterman alignment algorithm that integrates PhyloTree and EMPOP [5] information into the scoring function. Variants are called with reference to the revised Cambridge Reference Sequence (rCRS). The closest haplogroup is calculated, and variants are evaluated based on their occurrence in the haplogroup as well as other general metrics, including frequency, strand bias, and coverage. Variants and coverage can be viewed in a grid format (Figure 5), linear plot with read pileup (Figure 6), or circular plot (Figure 7).

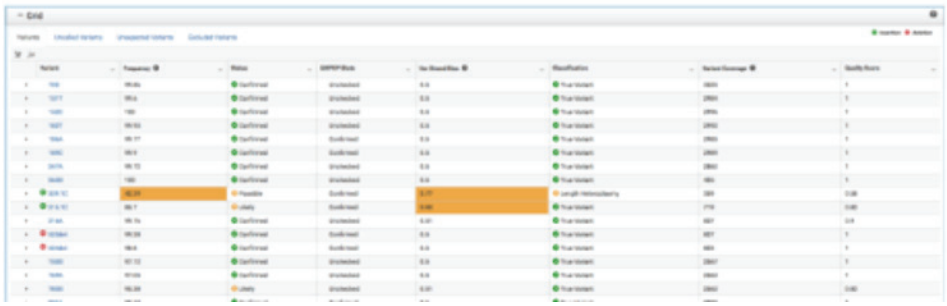


Figure 5. Grid view of variants in Converge Software. Variant information including variant frequency, likelihood status (confirmed, likely, possible, unclear, and unlikely), EMPOP state (whether the variant is expected in the closest haplogroup), strand bias, classification (true variant, point heteroplasmy, length heteroplasmy, nuclear mitochondrial DNA segment (NUMT), or artifact), coverage, and quality score is reported.

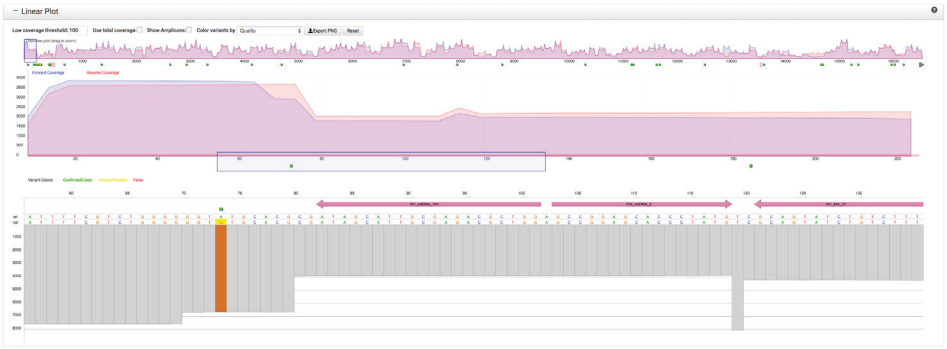


Figure 6. Linear coverage plot and read pileup in Converge Software. Forward (blue) and reverse (red) coverage is shown for the entire mtGenome (top panel), with the ability to zoom into selected regions (middle panel). Variants plotted below the coverage diagram are colored by their status (green, yellow, or red). The linear plot displays read pileups analogous to IGV views (bottom panel) for manual inspection of mtDNA sequence data.

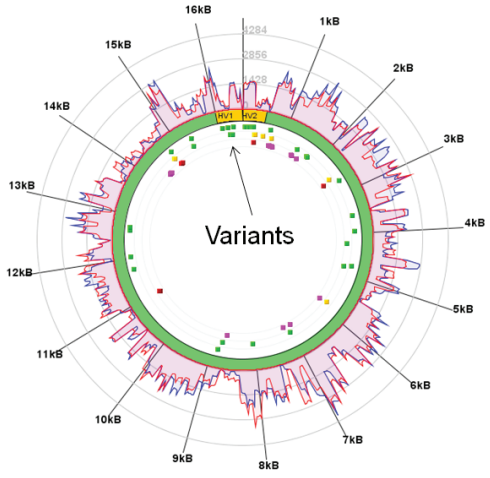


Figure 7. Circular plot in Converge Software. This plot type is implemented both in Integrative Genomics Viewer (IGV), which is incorporated to support visualization, and in Converge Software. Clicking on a variant in Converge Software allows a detailed alignment of reads to be examined in IGV.

Heteroplasmy, the occurrence of more than one mtDNA profile within a sample, can add to the difficulty of interpreting evidence. Heteroplasmy may exist as a difference in length (length heteroplasmy, LHP) or as an SNP (point heteroplasmy, PHP). The Converge NGS Data Analysis module can accurately detect heteroplasmic positions down to 10% (Figure 8), assuming a minimal sequencing coverage of >100x.

NUMTs are insertions of mtDNA sequences into the nuclear genome and can be nonspecifically (or unintentionally) amplified with the mtDNA genome, presenting a potential source of contamination. The Converge NGS Data Analysis module contains NUMT statistics and can detect and filter this type of contamination (Figure 9).

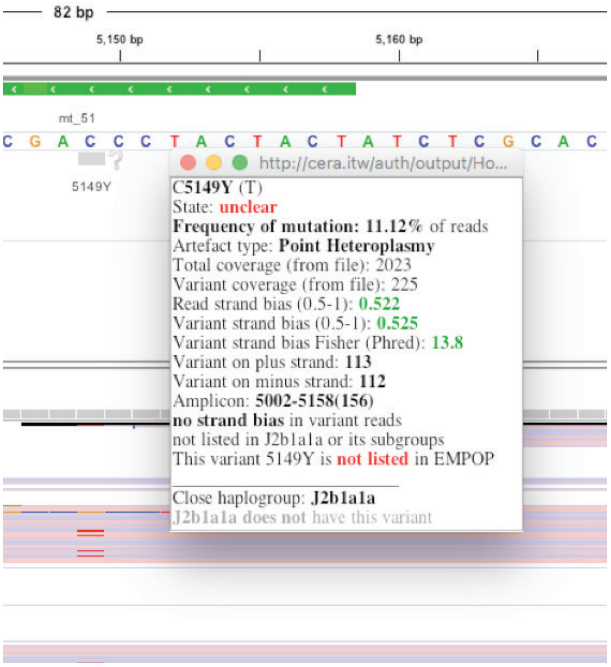


Figure 8. Point heteroplasmy in IGV. A variant was correctly classified as a PHP by Converge Software at position 5,149 bp.

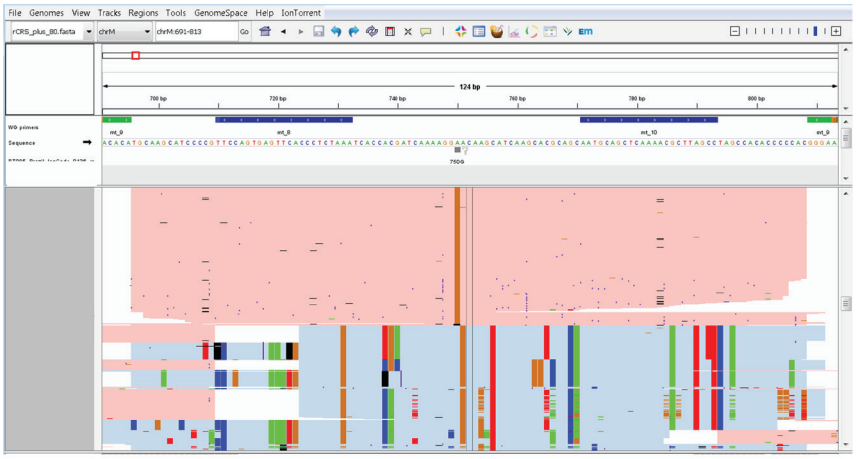


Figure 9. Example of NUMT contamination. The pink reads in the upper section are from the sample, whereas the reads in the bottom section are from nuclear contamination.

The EMPOP mtDNA database is used to collect and provide search capability for mtDNA haplotypes globally. Forensic investigators can upload and evaluate phylogenetic trees, both PHP and LHP, and haplogroup estimation by geography and population affiliation. Converge Software provides exported sequence data in a format compatible with EMPOP submission guidelines. Converge Software also allows comparison of sample haplotypes (Figure 10). Finally, Converge Software allows for data exports in a CODIS-compatible format for streamlined analysis.

Figure 10. Haplotype comparisons. Heat map comparison of multiple samples with each other.

applied
biosystems

Rely on our experienced team

Reach out to our Human Identification Professional Services (HPS) team. Since 2007, we have completed over 400 successful validation projects worldwide. Our team consists of more than 20 technical support specialists, each averaging 8 years of real-world forensic experience, providing customers with in-depth training and support on our instruments, chemistries, and software.

References

1. Wai KT, Barash M, Gunn P (2018) Performance of the Early Access AmpliSeq Mitochondrial Panel with degraded DNA samples using the Ion Torrent platform. *Electrophoresis* 39:2776–2784.
2. Strobl C, Eduardoff M, Bus MM et al. (2018) Evaluation of the Precision ID whole mtDNA genome panel for forensic analyses. *Forensic Sci Int Genet* 35:21–25.
3. Yao L, Xu Z, Zhao H et al. (2018) Concordance of mitochondrial DNA sequencing methods on bloodstains using Ion PGM. *Leg Med* 32:27–30.
4. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394.
5. Parson W, Dur A (2007) EMPOP—a forensic mtDNA database. *Forensic Sci Int Genet* 1:88–92.

Ordering information

Product	Quantity	Cat. No.
Precision ID mtDNA Whole Genome Panel	96 reactions manual, 48 reactions automated	A30938
Precision ID mtDNA Control Region Panel	96 reactions manual, 48 reactions automated	A31443
Precision ID Library Kit	96 reactions	A26435
	384 reactions	A30941
Precision ID DL8 Kit	32 reactions	A33212
IonCode Barcode Adapters 1-384 Kit	3,840 reactions	A29751
Ion S5 Precision ID Chef & Sequencing Kit (2 runs per initialization)	8 reactions	A33208
Ion 530 Chip Kit	8 chips	A27764
Ion 520 Chip Kit	8 chips	A27762
Ion 510 Chip Kit	8 chips	A34292
Converge Software and server	1 each	A35131
Case Management and NGS Data Analysis License	1 user, 3-year license	A35987
	5 users, 3-year license	A36237
HID Ion Chef System	1 each	A30070
HID Ion GeneStudio S5 System	1 each	A41431
HID Ion GeneStudio S5 Plus System	1 each	A41432
HID Ion GeneStudio Prime System	1 each	A41433

Find out more at **thermofisher.com/hid-ngs**

ThermoFisher SCIENTIFIC