Converge Software

A comprehensive forensic analysis platform that integrates answers for you

Key features and benefits

- Analyze NGS data from the Applied Biosystems™ Precision ID NGS System for mtDNA variation, SNPs, and STRs
- Demonstrate concordance between CE and NGS STR profiles
- Streamline kinship and paternity analysis
- Centralize forensic data, forensic analysis, and case information in one place to help increase lab efficiency
- Customize information architecture, data inputs, and parameters to meet the needs of your laboratory

Introduction

For forensic laboratories, data management and analysis are daunting, requiring numerous procedural steps and complex decision-making. Many laboratories therefore rely heavily on their software systems to effectively manage their day-to-day operations.

Applied Biosystems™ Converge™ Software, an all-in-one modular enterprise platform from Thermo Fisher Scientific, integrates forensic DNA data management and analysis into a single software package designed to increase the efficiency of forensic and relationship DNA-testing laboratories. The system is highly configurable to fit specific laboratory workflows, not only for analysis parameters but also for incorporation of data fields according to a laboratory's standard operating procedures (SOPs). Working with our technical support team, laboratories can customize data transfer processes between Converge Software and laboratory information management systems (LIMS).

Converge Software offers streamlined solutions for next-generation sequencing (NGS) analysis of the mitochondrial genome (both whole genome and control regions), single nucleotide polymorphism (SNP) markers (both ancestry and identity markers), and short tandem repeat (STR) markers. This comprehensive platform also includes solutions for capillary electrophoresis (CE)-to-NGS comparisons of STR profiles, kinship and paternity testing, and case management. Additionally, full auditing functionality is included for chain-of-custody requirements.

ThermoFisher SCIENTIFIC

NGS analysis

With recent advances in NGS, crime laboratories are now able to analyze the mitochondrial genome (mtGenome) to identify remains when there is only poor-quality or no autosomal DNA available for analysis, targeted and forensically relevant SNP markers to generate investigative leads, and STR markers to help determine the number of contributors in a mixture analysis.

Mitochondrial DNA analysis

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. Traditional Sanger sequencing by CE is generally limited to sequencing of the hypervariable region, as sequencing of the whole mtGenome is time consuming and cost prohibitive. NGS now makes it possible to sequence the mtGenome, which increases discrimination and sensitivity.

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 now gives forensic DNA laboratories the flexibility to detect variation within noncoding control sequences using the Applied Biosystems™ Precision ID mtDNA Control Region Panel, or to take advantage of the genetic diversity of full mtGenome sequence data using the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel. The automated mtDNA NGS analysis pipeline integrates various sources of knowledge for mtDNA variation, including PhyloTree [1] and EMPOP [2], to provide fine-tuned alignments and variant calls that avoid the pitfalls of standard algorithms. Laboratories interested in more advanced analysis and reporting of mtDNA types can obtain haplotype and haplogroup information, along with quantitative assessments for point and length heteroplasmies.

NGS reads from the BAM files are first mapped to nodes in PhyloTree and then realigned using a custom Smith-Waterman alignment algorithm that integrates PhyloTree and EMPOP information into the scoring function. Variants are called with reference to the revised Cambridge Reference Sequence (rCRS). Additionally, 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 can be viewed in a grid format, circular plot, and linear view (Figure 1).

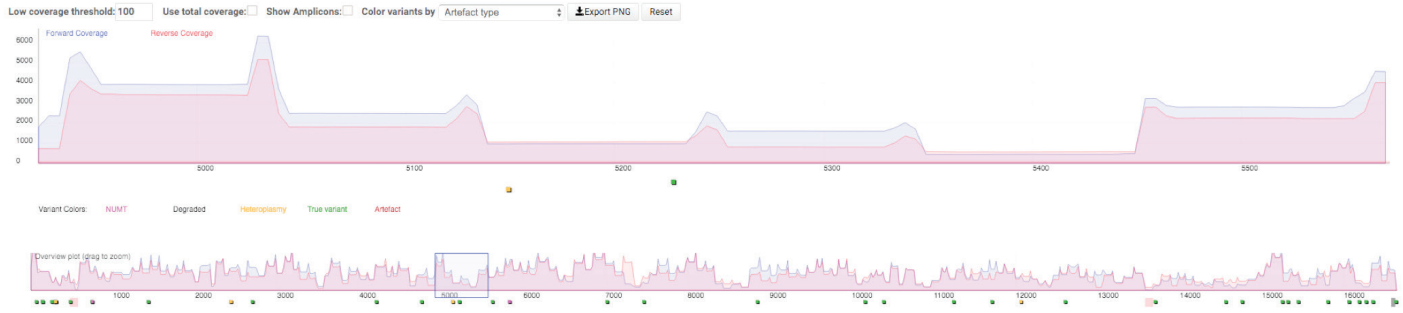

Figure 1. Linear coverage plot in Converge Software. Forward (blue) and reverse (red) coverage is shown across the entire mtGenome (bottom panel), with the ability to zoom into selected regions demonstrated in the top panel. Variants plotted below the coverage diagram are colored by their status (green: confirmed; yellow: possible, needs review; red: indicative of low-coverage regions).

Heteroplasmy, the occurrence of more than one mtDNA profile within a sample, adds to the difficulty of interpreting evidence. Heteroplasmy may exist as a difference in length (length heteroplasmy, LHP) or as a SNP (point heteroplasmy, PHP). The Converge NGS Data Analysis module can accurately detect heteroplasmic positions down to ~10% (Figure 2), assuming a minimal sequencing coverage of >100x.

Nuclear mitochondrial DNA segments (NUMTs) are insertions of mtDNA sequence into the nuclear genome, and they 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 3).

SNP analysis

Forensic samples may contain DNA that is too degraded to generate an STR profile. Or there may be an STR profile that does not generate a potential match in a national DNA databank. These samples may be analyzed with SNP markers, either ancestry-informative markers (AIMs) to generate an investigative lead, or identity markers to associate a degraded crime scene sample to a known reference. The Converge NGS Data Analysis module contains analysis parameters for generating reports from the Applied Biosystems™ Precision ID Ancestry Panel and the Applied Biosystems™ Precision ID Identity Panel.

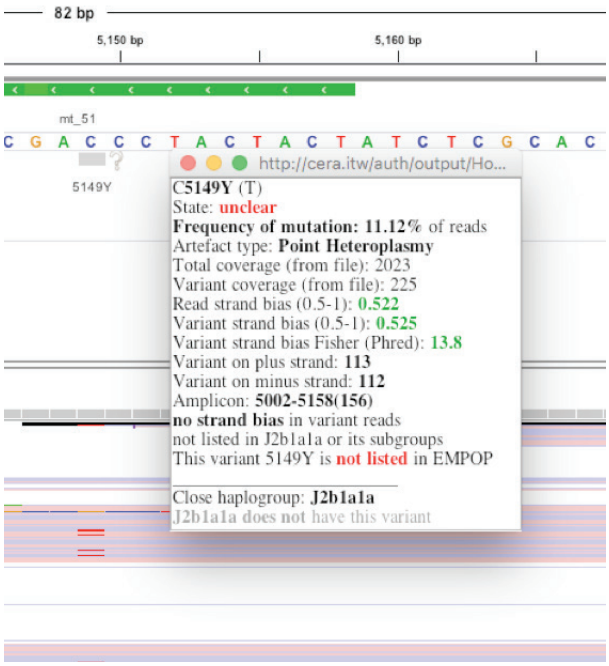

Figure 2. A variant was correctly classified as a PHP by Converge Software at position 5,149 bp.

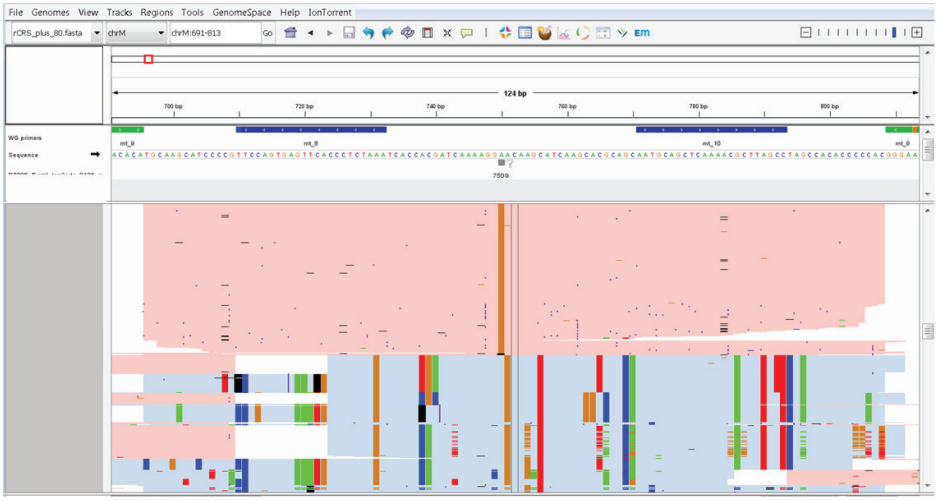

Figure 3. Example of NUMT contamination. The pink reads on top are from the sample, whereas the reads at the bottom are nuclear contamination.

The Converge SNP Analysis module provides a variety of metrics to monitor sequencing quality, including coverage of aligned reads to a hotspot, strand bias, number of reads containing each base at the hotspot, genotype call, and quality (Figure 4). Tertiary ancestry analysis consists of generating an estimation of admixture prediction (Figure 5) from seven geographical regions, including Africa, America, East Asia, Europe, Oceania, South Asia, and Southwest Asia, as well as population likelihoods (based on algorithms from Kenneth Kidd).

Identity analysis consists of calculating random match probability (RMP) based on genotype frequencies generated from 1000 Genomes Project data as well as a Y-haplogroup prediction (Figure 6).

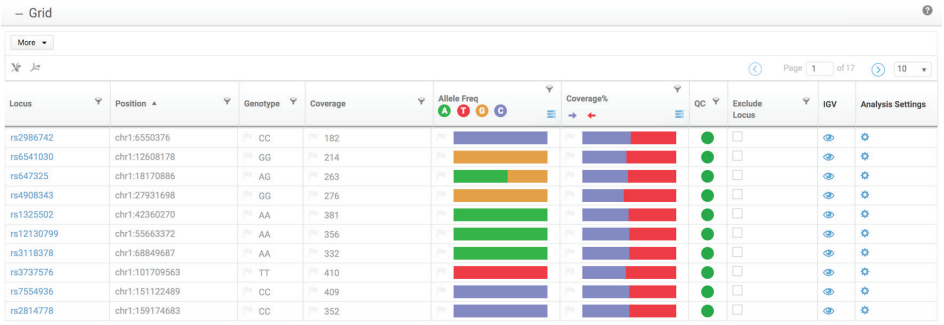

Figure 4. Grid view of Converge SNP analysis metrics. Attributes are genotype, sequencing read coverage, allele frequency, strand bias, and QC. A link to Integrative Genome Viewer (IGV) is also included.

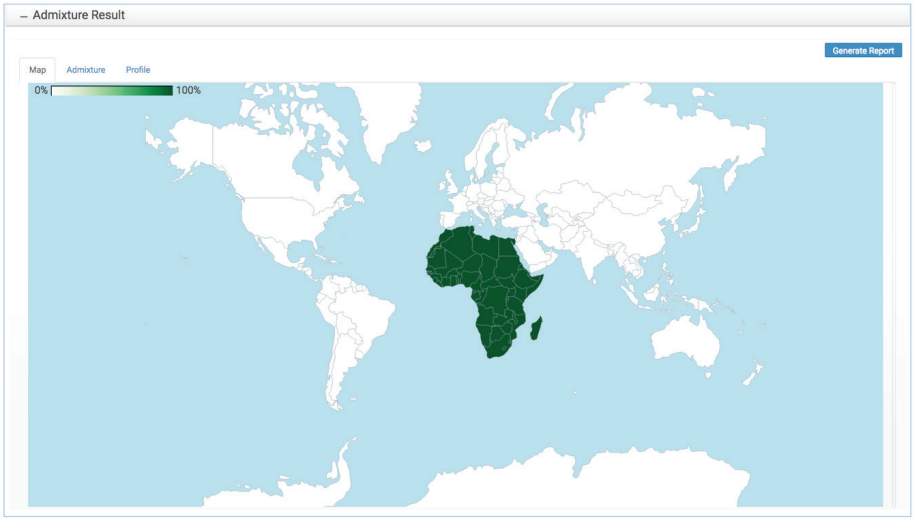

Figure 5. Graphical representation of an admixture map using the Converge SNP Analysis module. The percentage of the corresponding population in the sample is displayed. The darker the color, the greater the percentage of the corresponding population in the sample. Analysts can hover over the region to display the proportion value of the corresponding population in the sample.

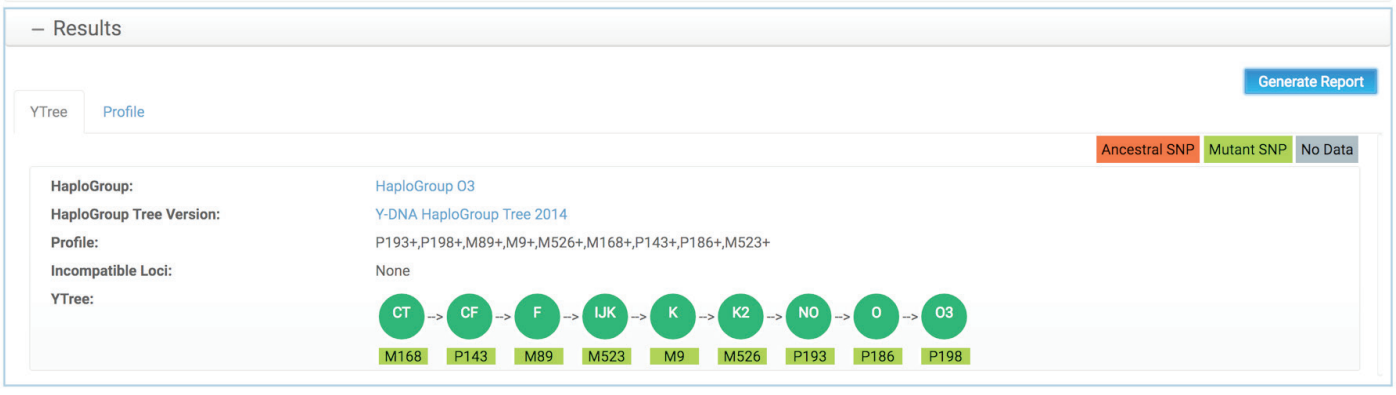

Figure 6. Y-haplogroup results from the Precision ID Identity Panel. Attributes displayed are haplogroup, derived alleles, any markers that conflict with the reported Y lineage, and the Y clades (red: ancestral SNP; green: mutant SNP; gray: no data).

STR analysis

The application of NGS is particularly helpful with degraded samples that may not provide a full STR profile using traditional CE methods. The Converge NGS Data Analysis module is used to generate profiles from the Applied Biosystems™ Precision ID GlobalFiler™ NGS STR Panel v2. The Converge NGS Data Analysis module provides information on STR allele calls, STR sequence motifs, known SNPs in flanking regions, and isometric heterozygotes (alleles of the same fragment length but containing different sequences).

Because Converge Software has an interface similar to that of Applied Biosystems™ GeneMapper™ *ID-X* Software, forensic analysts will be able to quickly evaluate NGS data using familiar process quality values (PQV) and flags such as allele number (AN), off-ladder allele (OL), peak height ratio (PHR), below stochastic threshold (BST), and control concordance (CC) (Figure 7). Preconfigured analysis settings are provided within the NGS module and may be modified as needed.

Sequence analysis of STRs also provides additional discrimination by resolving isometric heterozygotes (Figure 8) and shared STR alleles that contain SNPs in flanking regions (Figure 9). These additional sources of allelic diversity may be useful in both mixture analysis and kinship interpretation.

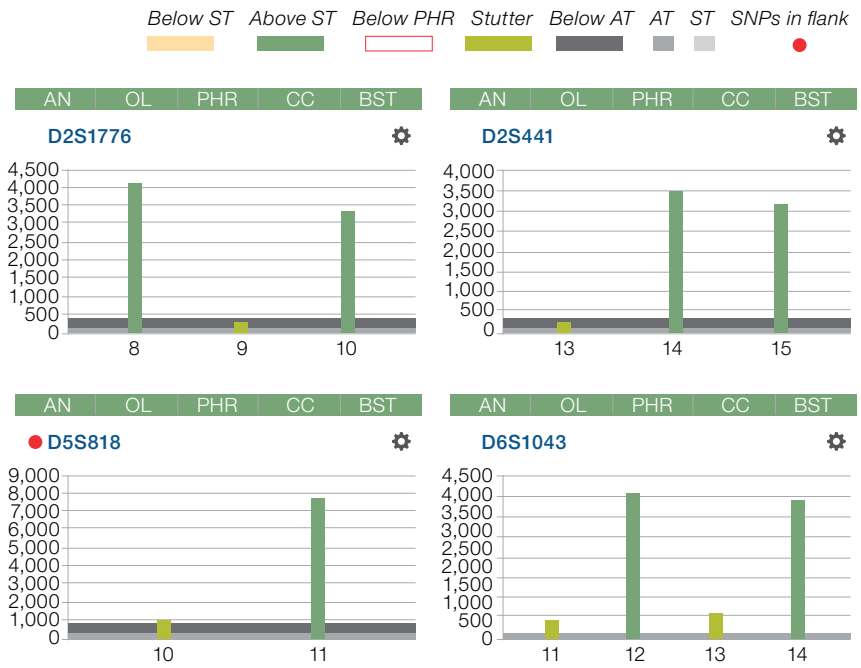

Figure 7. View of NGS secondary analysis results using Converge Software.

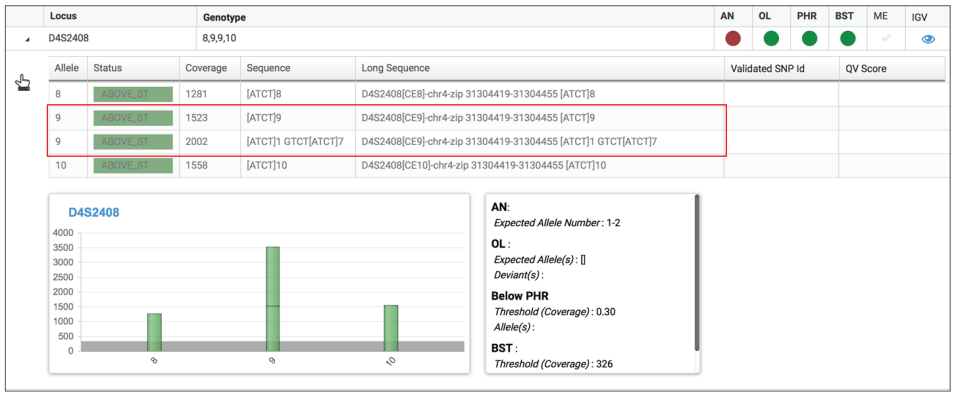

Figure 8. Detailed information from isometric heterozygote analysis with sequencing coverage and repeat motif structure.

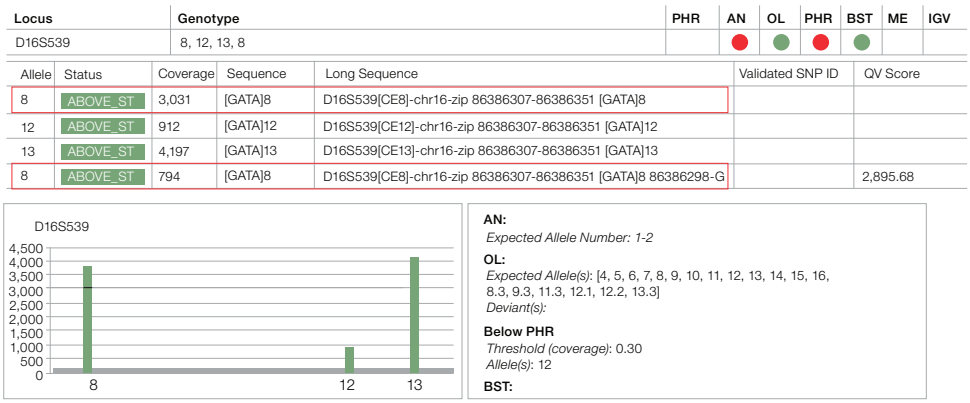

Figure 9. Detailed information on an SNP in a flanking region from a minor contributor. The SNP is identified with ISFG-recommended nomenclature. SNP information can be explored within public genomic databases via the IGV.

Reference ✔ Match ✖ Not a match ⚠ Partial match ⊘ N/A

Profile ID(s) Profile-1028 Profile-1033 Profile-1029

Locus	Profile-1028 (Precision_ID_GlobalFiler_NGS_STR_Panel_v1.0)	Profile-1033 (VeriFilerExpress_Panels_v2x)	Profile-1029 (VeriFilerExpress_Panels_v2x)
AMEL	X, Y	X, Y	X, Y
CSF1PO	10, 11, 12	10, 12	11, 12
D10S1248	13, 14	13, 14	13, 14
D12ATA63	12	-	-
D12S391	17, 18, 19	17, 18	18, 19
D13S317	9, 11, 12	11, 12	9, 10
D14S1434	13, 14	-	-
D16S539	10, 12, 13	10, 12	12, 13
D18S51	12, 17	12, 17	17, 17
D19S433	13, 15.2	13, 15.2	13, 13
D1S1656	13, 16, 17, 19.3	13, 17	16, 19.3
D1S1677	15	-	15
D21S11	28, 29, 30, 31	28, 29	30, 31
D22S1045	14, 16, 17	14, 17	16
D2S1338	18, 19, 22	19, 22	18

Figure 10. Comparison of NGS and two CE profiles (imported from GeneMapper *ID-X* Software) via the Case Management module. The NGS sample is a mixture of the two CE samples. The software displays marker-level profile comparisons as a match, non-match, or partial match.

As forensic laboratories begin to adopt sequence analysis of STR markers into casework applications, verification and concordance studies may be required. Using the NGS Data Analysis and Case Management modules of Converge Software allow for the easy comparison of NGS and CE profiles (Figure 10). This comparison ability will also be useful in the future when comparing a crime scene sample analyzed by NGS to a reference sample that has been processed using traditional CE methods.

Kinship and paternity analysis

Kinship and paternity testing can be challenging due to the sheer volume of cases to be analyzed, as well as the complexity of laboratory workflows and statistical analyses being performed. The Converge Kinship and Paternity module offers a comprehensive set of parameters to perform analysis, allowing laboratories to conduct various types of relationship testing in a fraction of the time compared to current methods.

The Kinship and Paternity module incorporates data from GeneMapper *ID-X* Software or from the Converge NGS analysis of Precision ID GlobalFiler NGS STR Panel v2. Converge Software identifies and processes Trio Paternity, Trio Maternity, Duo Fatherless, and Duo Motherless kinship tests by automatically creating cases, adding subjects and profiles, running analyses, and generating reports. Users can quickly and easily create a variety of hypotheses by drawing pedigree trees and setting comprehensive analysis parameters, including application of mutation models and population substructures (Figure 11). It also allows for genetic likelihood ratio calculations, as well as the generation of reports for any complex relationship analysis. These analyses are conducted utilizing validated algorithms.

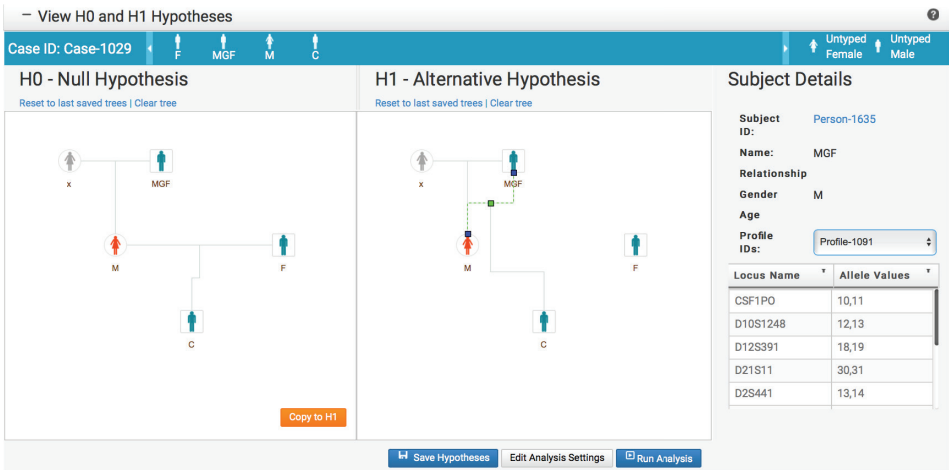

Figure 11. The Kinship and Paternity module offers an intuitive user interface to quickly perform relationship testing.

Through an intuitive interface, users can easily review detailed results, and also generate and electronically sign reports in a few simple steps. Converge Software provides one default report template; however, laboratories can use existing tools to modify this template and adjust the report to their desired format. For trio and duo testing in paternity and maternity cases, the Quick Kinship Analysis (KA) functionality of the Case Dashboard integrates all of these steps on one screen, saving time within the laboratory.

Verification

Complex analysis parameters used in kinship analysis require verification in order to determine the appropriate settings for the thresholds governing the data interpretation. Therefore, it is important that the software be tested using a variety of samples that challenge each different relationship-testing scenario. Although default settings are suggested within the software, these can be adjusted based on the outcome of each laboratory's internal evaluation of Converge Software. Optimizing any software will require an additional investment on the part of the laboratory. Extensive system verification has been performed at Thermo Fisher Scientific, and a summary of these verification studies is available as a reference guide for users interested in implementing the software.

Case management

Underlying the centralization of data creation, analysis, and storage to one easy-to-access location, the Converge Case Management module supports case, subject, and genotype profile. The module serves the needs of various users such as laboratory managers and analysts, providing intuitive data views and reports that can be configured and saved for each user. These features are easily accessible through a secured web browser under the protection of the laboratory's IT department. The information management configurability allows the user to focus on pertinent information and adapt to a laboratory's specific needs by adding new data fields. Case data management includes metadata related to the case, subjects, DNA profiles, and attachments such as pictures, Microsoft™ Word™ documents, and Microsoft™ Excel™ files, with extensive data fields available to track all required information.

Case dashboard

Ideal for quickly reviewing case status and accessing case reports, the Case Dashboard functionality provides an at-a-glance overview of the case (Figure 12). Extensive search and filtering capabilities on many data elements allow faster access to relevant information, saving time and effort. Once performed, searches can be saved and viewed at a later date. Cases can also be exported for external sharing and storage and can be imported back as needed. Archiving can be scheduled or performed manually, to remove closed cases from the active dashboard. Should the need arise, archived cases can always be retrieved in the Case Dashboard.

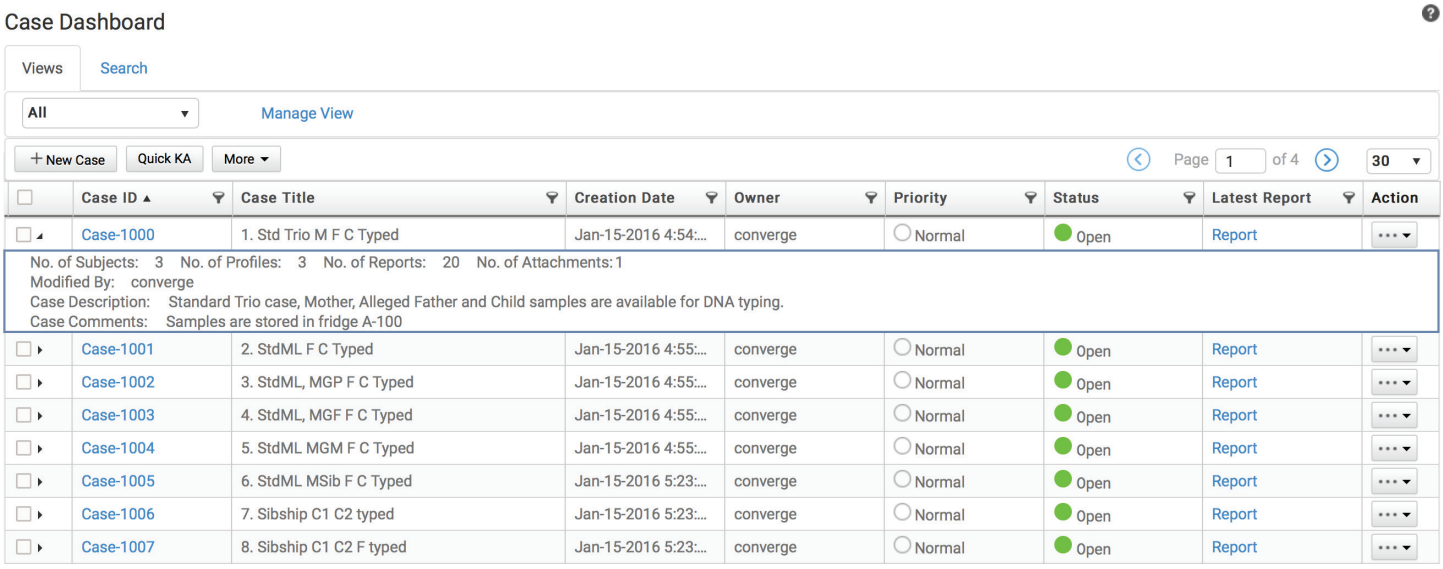

Figure 12. The Case Dashboard centralizes information on all cases in one view, allowing for a quick review.

Rely on our experienced team

The HID University offers an Advanced Converge Software Training course to assist you in optimizing your workflow with Converge Software. This two-day course provides comprehensive training on the features of Converge Software. Course content will focus on the module(s) that fit your laboratory needs. Interactive lectures and hands-on exercises will provide detailed instruction on topics such as administrative features and audit dashboard, case management, case and batch dashboard, kinship analysis, integration with Torrent Suite Software, NGS data analysis, and report generation.

Security

Easily accessible through a web browser, data can be accessed from any computer with connectivity privileges to the Converge Software server and are deployed within a laboratory's own intranet. Access to data within the software is secured through various security policies, including password expiration and account suspension. Backup and restore functionalities allow laboratories to securely maintain data and manage disaster recovery situations.

Converge Software server

The Converge Software server, based on the Linux™ operating system, hosts Converge Software using an internet protocol (IP) address configured in the system. The Converge Software server requires an administrator to manage user accounts and comes preconfigured with the following components and settings:

- Dell™ PowerEdge™ T130 II base unit
- Red Hat™ Enterprise Linux™ operating system
- Apache™ Tomcat™ application server that runs Converge Software
- PostgreSQL database server that stores the data for the server and software
- Google™ Chrome™ browser
- Automatic configuration of IP, domain name service (DNS), and Windows Internet Name Service (WINS) settings via dynamic host configuration protocol (DHCP)

Note: See the manufacturer's documentation accompanying the server or the manufacturer's website for network requirements, specifications, and product details.

Supported browsers

Converge Software is supported for use with the following browsers:

- Apple™ Safari™ v8
- Microsoft™ Internet Explorer™ v11
- Google™ Chrome™ v45

Note: Google Chrome is the only browser available on the Converge Software server.

Converge Software server components

Description
Dell PowerEdge T130 II Tower Server
Dell Flat Panel Monitor, 23 inch (or equivalent) with cable
Dell QuietKey USB Keyboard
Dell Laser USB Mouse
64 GB USB drive
Power cord*
RJ45 CAT6 Ethernet cable
USB drive (Converge Software)
USB drive (Converge Software documentation)

* Supplied with the server and monitor. Supplies 13 A/125 V, depending on the geographic location of the installation.

Converge Software server specifications

Specification	Description
Processor	PowerEdge T130 Motherboard v2 with cabled hard drives, Intel™ Xeon™ E3-1270 v5, 3.6 GHz, 8 M cache, 4C/8T, turbo (80 W)
Memory	16 GB memory (2 x 8 GB), UDIMM, 2,133 MT/s, ECC
Hard drive	2 x 2 TB, 7,200 RPM, NLSAS 12 Gbps, 3.5-inch cabled hard drive (RAID1)
Data storage	RAID1, PERC H330 integrated RAID controller for 3.5-inch hard drive

applied biosystems

Ordering information

Description	Cat. No.
Converge Software with server	A35131
Case Management and NGS Analysis modules (1 user, 3-year license)	A35987
Case Management and NGS Analysis modules (5 users, 3-year license)	A36237
Case Management and Kinship & Paternity Analysis modules (1 user, 1-year license)	A31001
Case Management and Kinship & Paternity Analysis modules (1 user, 3-year license)	A31002
Case Management and Kinship & Paternity Analysis modules (5 users, 3-year license)	A31005

Additional configurations of software licenses are available. Please contact your local representative for more information.

References

1. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394.
2. Parson W, Dur A (2007) EMPOM—a forensic mtDNA database. *Forensic Sci Int Genet* 1:88–92.

Find out more at **thermofisher.com/converge**

ThermoFisher SCIENTIFIC