# Comprehensive Analysis of NGS Data from Modern and Ancient DNA with Parabon Fx™

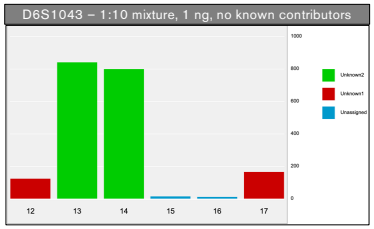
Ellen Greytak, PhD*; Janet Cady, PhD; and Steven Armentrout, PhD
Parabon NanoLabs, Inc.
*Correspondence to: ellen@parabon.com

## Introduction

Next-generation sequencing (NGS) (also known as massively parallel sequencing, MPS) has greatly expanded the range of analyses that can be performed on forensic samples, while still producing length-based STR profiles compatible with legacy capillary electrophoresis profiles. However, software systems provided by the NGS kit manufacturers are limited to a single NGS platform and lack many tools that practitioners regularly need, impeding both casework and research. Parabon's Fx™ Forensic Analysis Platform is designed to address this challenge by providing a single, secure, user-friendly interface to dozens of tools to analyze and compare NGS data from multiple sequencing platforms, while also enabling easy access to raw data and intermediate analysis files. This poster demonstrates some of the new analyses that can be performed in Fx and shows screenshots of how the results are displayed. Fx includes tutorials and detailed write-ups of how each tool works and how the results can be interpreted.

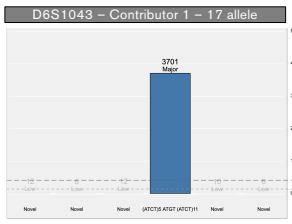
## STR Mixture Deconvolution

- STR mixtures can be deconvoluted in Fx using length-based or sequence-based alleles (uses EuroForMix continuous model[1])
- User specifies number of contributors and known contributors (if any) under each hypothesis
- Produces deconvoluted genotypes with log likelihood ratios (LLRs)
- Calculates weight-of-evidence for statistical assessment
- Performs optional non-contributor analysis

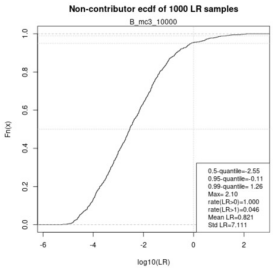



| Allele | Sequence | ReadCount | Contributor |
|---|---|---|---|
| 12 | (ATCT)12 | 123 | Unknown1 |
| 13 | (ATCT)13 | 840 | Unknown2 |
| 14 | (ATCT)14 | 798 | Unknown2 |
| 15 | (ATCT)15 | 14 | Unassigned |
| 16 | (ATCT)5ATGT(ATCT)10 | 10 | Unassigned |
| 17 | (ATCT)5ATGT(ATCT)11 | 164 | Unknown1 |

| | Likelihood | Log Likelihood |
|---|---|---|
| **Under Hd** | 3.18e-302 | -694.225 |
| **Under Hp** | 6.05e-297 | -682.067 |
| **Ratio** | 191000 | 5.28 |



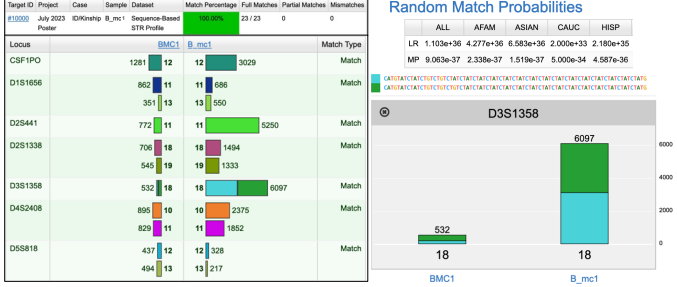

Contributor 1 actual genotype:
(ATCT)12 / (ATCT)5 ATGT (ATCT)11



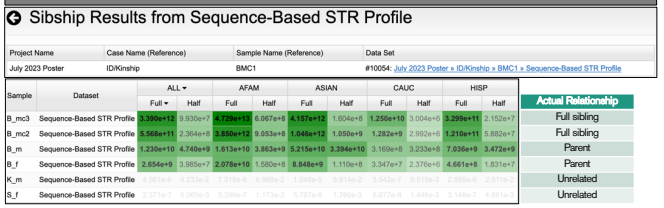

## Cross-Kit STR Sequence-Based Identity & Kinship

- Each NGS kit targets a slightly different region around each STR → sequence alleles differ in flanking SNPs
- Fx cross-kit analysis trims sequence alleles to the overlapping region and uses population allele frequencies trimmed to that same region
- Identity matching (RMP) or parentage/sibship (LR vs. unrelated)

Sequence-Based STR Identity Matching:
IonTorrent PrecisionID GlobalFiler (BMC1) vs. ForenSeq Panel B (B_mc1)

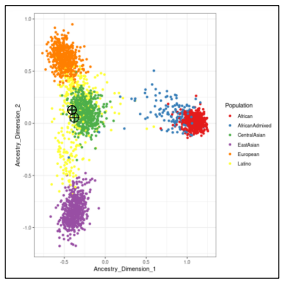



Sequence-Based STR Kinship Testing:
IonTorrent PrecisionID GlobalFiler vs. ForenSeq Panel B




## SNP Ancestry

- Fx uses principal components analysis (PCA), Admixture, & FrogKB population likelihoods to infer ancestry from SNPs (e.g. ForenSeqB, PrecisionID Ancestry)
- Samples that would be called as "Admixed American" by other tools can be distinguished as Central/South Asian vs. Admixed Asian/European vs. Latino




Global Ancestry Analysis

The table below lists the proportional membership of this genome relative to seven global ancestry regions.

| Ancestry Proportions | |
|---|---|
| Central Asia | 100.00% |
| Africa | 0.00% |
| America | 0.00% |
| East Asia | 0.00% |
| Europe | 0.00% |
| Oceania | 0.00% |

Global Ancestry Analysis

The table below lists the proportional membership of this genome relative to seven global ancestry regions.

| Ancestry Proportions | |
|---|---|
| Europe | 63.50% |
| East Asia | 36.50% |
| Africa | 0.00% |
| America | 0.00% |
| Central Asia | 0.00% |
| Oceania | 0.00% |

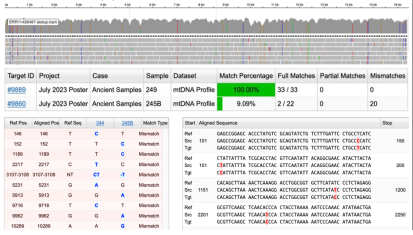
Population Likelihood Analysis

The table below shows this genome's 10 most likely populations from FrogKB.

| Population | Region | Genotype Probability | Population | Region | Genotype Probability |
|---|---|---|---|---|---|
| Kachari | Asia | 2.4545e-15 | Chuvash | Europe | 1.1131e-19 |
| Gujarati(GIH) | Asia | 8.6251e-16 | Kazak | Asia | 9.0712e-20 |
| Sri_Lankan_Tamil(STU) | Asia | 4.0454e-16 | Qinghai_Tibetans | Asia | 3.6840e-20 |

## Mitochondrial DNA Analysis

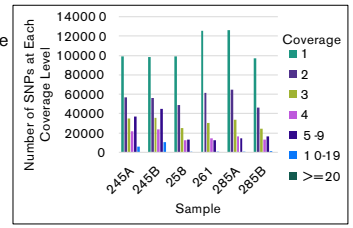
- Automatic detection of sequenced region; custom flat or per-site error rates
- View the alignment in the chromosome browser, perform string- or haplotype-based lineage matching with or without including heteroplasmy and homopolymer regions, view the differences in context, and predict haplogroup using the AQME approach[2]



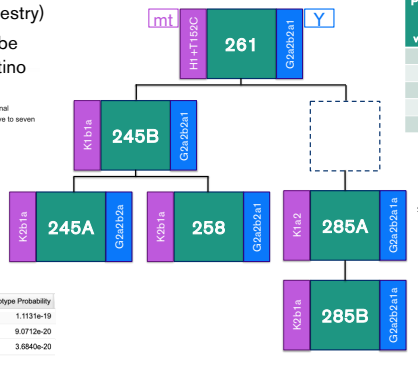
| Haplogroup | Score | Shared | Missing | Private |
|---|---|---|---|---|
| K2b1a | 304 | A11467G, A12308G, G12372A, A1811G, T8698C, A3480G, G9055A, C14167T, A10550G, T11299C, T14798C, T16224C, T16311C, T146C, T9716C, C2217T, G5231A, A14037G, C11869A, G13135A, A263G, C7028T, G11719A, A4769G, A15326G, A2706G, A1438G, A8860G, A73G, A750G, C14766T | None | N3107C, T16519C |
| K2b1a3 | 296 | A11467G, A12308G, G12372A, A1811G, T8698C, A3480G, G9055A, C14167T, A10550G, T11299C, T14798C, T16224C, T16311C, T146C, T9716C, C2217T, G5231A, A14037G, C11869A, G13135A, A263G, C7028T, G11719A, A4769G, A15326G, A2706G, A1438G, A8860G, A73G, A750G, C14766T | A723G | N3107C, T16519C |

## Distant Kinship from Low-Coverage Sequencing

- Distant kinship can be inferred in Fx even from low-coverage SNP capture or WGS data: calculates LLR of data under different relationship hypotheses vs. unrelated hypothesis
- FASTQ files downloaded for 6 ancient French individuals (4850-4500 BC) with known pedigree[3]
- Sequencing coverage is too low for direct genotype calling: <1.5% of SNPs have ≥10X coverage
- Genotype likelihoods calculated for 94,752 SNPs and used for extended kinship out to 4th degree[4] in Fx
- Determined mtDNA & Y (using Yleaf approach[5]) haplogroups in Fx




| Actual | 245A | 245B | 258 | 261 | 285A |
|---|---|---|---|---|---|
| 245B | Parent-Child (PC) | | | | |
| 258 | Full Siblings | PC | | | |
| 261 | Grand-parent (GP) | PC | GP | | |
| 285A | First cousin (1C) | Avuncular | 1C | GP | |
| 285B | 1C once-removed (1C1R) | Great-avuncular | 1C1R | Great-GP | PC |




| Predicted Log Likelihood Ratio (LLR) vs. Unrelated | 245A | 245B | 258 | 261 | 285A |
|---|---|---|---|---|---|
| 245B | 1073.14 | | | | |
| 258 | 577.85 | 722.34 | | | |
| 261 | 261.29 | 809.03 | 130.08 | | |
| 285A | 82.39 | 255.76 | 56.11 | 211.86 | |
| 285B | 26.61 | 79.13 | 20.77 | 66.87 | 633.14 |

[1] Bleka, et al. (2016). EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics*, **21**, 35–44.

[2] Sturk-Andreaggi, et al. (2017). AQME: A forensic mitochondrial DNA analysis tool for next-generation sequencing data. *FSI: Genetics*, **31**, 189–197.

[3] Rivollat, et al. (2023). Extensive pedigrees reveal the social organization of a Neolithic community. *Nature*, **620**, 600–606.

[4] Gorden, Greytak, et al. (2022). Extended kinship analysis of historical remains using SNP capture. *FSI: Genetics*, **57**, 102636.

[5] Ralf, et al. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Mol Biol Evol*, **35**(5), 1291–1294.