

## Preparations for shotgun sequencing analyses in forensic genetics

Børsting C. \*, Jepsen A.H., Andersen J.D., Kampmann M.L.

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

\*claus.boersting@sund.ku.dk

### Abstract

Shotgun sequencing of a trace or reference sample has at least three important purposes: 1) Increase the evidential weight in cases involving close relatives, 2) Human identification of DNA from trace samples that are too degraded for standard short tandem repeat (STR) analyses, and finally 3) Provide single nucleotide polymorphism (SNP) profiles for forensic investigative genetic genealogy (FIGG) searches. The genotypes obtained from shotgun sequencing will not include the standard STRs and thus, searches in the traditional forensic DNA databases will not be possible. Instead, the SNP profile will be used for comparison to a reference sample. Shotgun sequencing potentially provides genotypes of millions of SNP loci covering the entire human genome. In addition to human identification and FIGG searches, the SNPs may provide information for population assignment, forensic DNA phenotyping, and identification of haploid lineages.

Shotgun sequencing of forensically relevant reference samples was tested using four extraction methods (Chelex®, PrepFiler™ Automated Forensic DNA Extraction Kit (Thermo Fisher Scientific), and EZ1&2 DNA Investigator Kit (Qiagen) for FTA™ cards and whole blood) and three library building protocols (two for double-stranded DNA and one for single-stranded DNA). Three individuals donated whole blood and buccal swabs. The libraries were generated and sequenced on a NovaSeq 6000 (Illumina).

Extracts from Chelex® and PrepFiler™ extraction methods were likely single stranded and therefore best suited for the single stranded library building protocol, whereas the extracts from EZ1&2 methods were double stranded and the double-stranded library building protocols showed the most promising results.

Sequence analyses were conducted with an in-house developed pipeline involving BWA-MEM and GATK. Two parameters (read depth (DP) and genotype quality (GQ)) were used to evaluate the genotypes. Stringent criteria for DP and GQ will be needed for human identification, whereas more relaxed criteria might be sufficient for FIGG searches.