Rootless hair - an underappreciated, non-invasive sample type for genomic analyses in animal husbandry contexts

Varsha Rao[1], Kelly M. Harkins[1], Christopher J. Troll[1], Charles Vaske[1], Colin Naughton[1], Jessica M. Morgan[1], and Richard E. Green[2*]

ABSTRACT

Next-generation sequencing (NGS) technology has revolutionized the way we explore the living world. However, in many scenarios, including in animal husbandry contexts, it may be difficult to obtain biological samples for routine genetic analyses. The use of rootless hair has only recently emerged in forensic genetics as a viable sample type for human genome-level analyses, including genotyping, because a single strand contains only picograms of highly fragmented DNA (< 50 bp). These characteristics have historically hindered conventional PCR-based testing as well as standard NGS library preparation methods.

Here we demonstrate the utility of a single-stranded NGS library preparation approach developed by Claret Bioscience that enables conversion of ultra-short DNA for whole genome sequencing. Using ancient DNA extraction methods coupled with the SRSLY PicoPlus NGS library preparation kit, we generated sequence-ready libraries in under 3 hours from as little as 100 pg of DNA from rootless hair/feathers. The samples were collected directly (e.g. cut or sheared) or indirectly (e.g. shed) from a variety of living animals including horse, sheep, chicken, alpaca, gibbon, dog, cat, as well as extinct animals like the wooly rhino and mammoth.

By mapping the sequencing reads to the respective reference genomes we show that SRSLY generated high quality DNA data, even from 50,000+ year old samples. The fragment length distribution of all samples show the same mean length and ~10bp periodicity as observed in modern human hair samples. Even with low depth of sequencing (<4 Million reads), data from rootless hair was used to reconstruct complete mitochondrial genomes at high coverage, which can be used in evolutionary analysis, in determination of mtDNA derived phenotypical features, and more.

Molecular complexity estimates show that each tested hair strand contained tens or hundreds of millions of unique DNA fragments, enabling a wide range of whole genome analyses, like SNP profiling. This novel sample type and genomic approach opens new possibilities for non-invasive wildlife sampling, livestock breed evaluation and conservation, ecological analyses, and other agricultural applications.

REFERENCES

1 Troll, C. J. et al. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. BMC Genomics 20, 1023, doi:10.1186/s12864-019-6355-0 (2019).

2 Cohen, J. D. et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci U S A 114, 10202-10207, doi:10.1073/pnas.1704961114 (2017).

3 Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. Bioinformatics 33, 2381-2383, doi:10.1093/bioinformatics/btx183 (2017).

4 Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

5 Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10, 512-526, doi:10.1093/oxfordjournals.molbev.a040023 (1993).

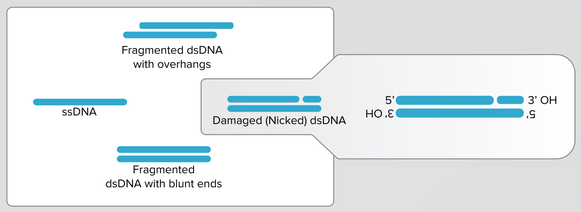
6 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35, 1547-1549, doi:10.1093/molbev/msy096 (2018).

DNA EXTRACTION FROM ROOTLESS HAIR

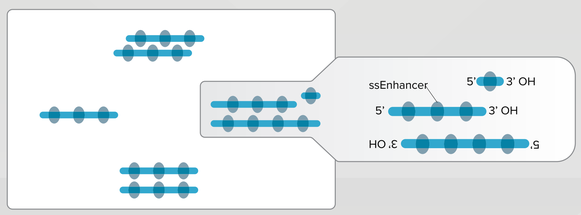
Figure 1. DNA can be extracted from challenging samples including rootless hair, fur, fossilized bones, plasma. These contain highly fragmented DNA that were considered to be unsuitable for NGS library preparation.

LIBRARY PREPARATION WORKFLOW

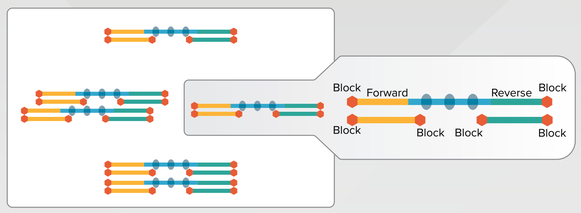
1. DNA INPUT POOL



2. DENATURE AND CAPTURE AS SINGLE-STRANDED DNA



3. PHOSPHORYLATE TEMPLATE DNA AND LIGATE ADAPTERS



4. INDEX PCR

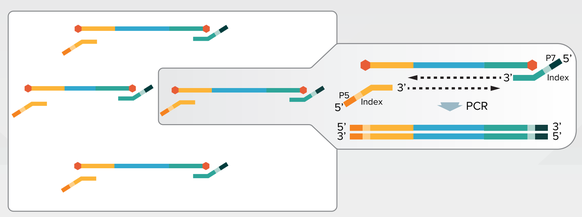


Figure 2. ClaretBio's SRSLY method simple yet robust workflow allows yet high quality NGS library prepartion[1].

DNA FRAGMENTATION VARIES BETWEEN SAMPLE TYPES

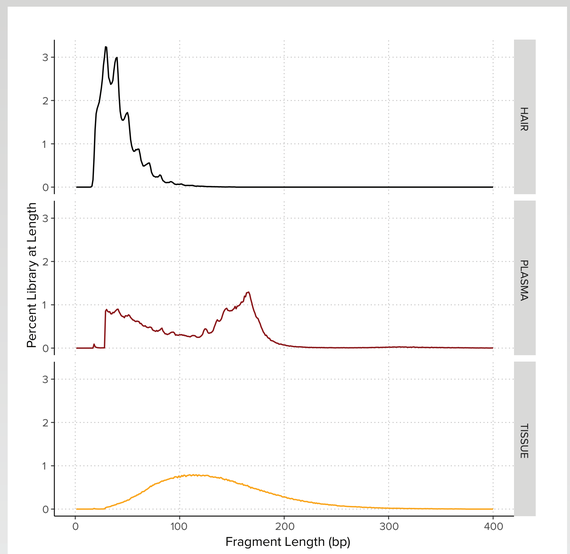


Figure 3. Comparion of DNA fragment size from various sources. SRSLY libraries were generated according to SOP with DNA obtained form Hair (250pg), Plasma (1ng) and Spleen FFPE curl (10ng). Libraries were sequenced on an Illumina Miseq to a depth of 0.5 million reads and mapped to the human genome using BWA[2]. The insert size distribution shows that DNA from hair is highly fragmented (< 100 bp), in comparison to other sources of fragmentary and/or damaged DNA such as plasma and tissue derived DNA.

GENOMIC DISTRIBUTION OF ALIGNED READS

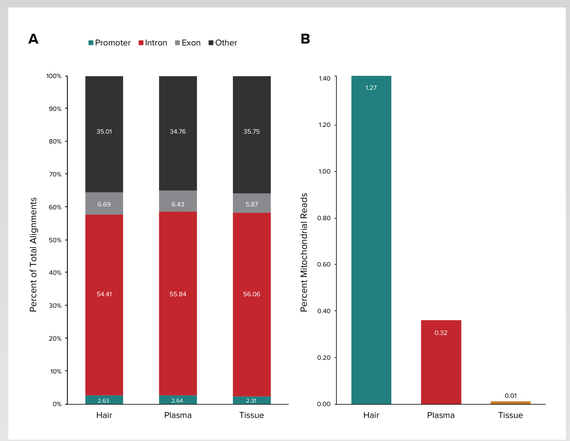


Figure 4. Rootless hair largely recapitulates cellular genomic composition, expect for mitochondrial reads. Aligned bam files from human hair, plasma and tissue DNA were using in this analysis. A) Similar genomic composition of aligned reads from each sample type. Using a gene model that describes the genomic regions pertaining to protomer, intron and exon region using AnnotatoR tool[3]. B) Rootless hair has a higher percent of mitochondrial reads. Reads mapping to mitochondrial chromosomes were determined and normalized to read depth using samtools idxstats[4].

NGS ANALYSIS FROM HAIR, FUR AND FEATHER DNA

Rootless Hair Collected From	Input DNA Amount	PCR Cycles	Total DNA Yield Post Index PCR	Total Read Pairs Sequenced	Average Mitochondrial Coverage
Alpaca	220 pg	15	183.2 ng	3,779,907	13x
Cat	350 pg	14	1276 ng	3,636,029	324x
Chicken	600 pg	12	234 ng	3,062,552	20x
Dog	200 pg	15	510 ng	1,680,182	14x
Gibbon	200 pg	15	382 ng	1,831,741	51x
Horse	250 pg	15	360 ng	3,610,330	25x
Sheep	2000 pg	12	1492 ng	2,711,282	18x

Table 1. SRSLY library were generated from 200-2000 pg rootless hair. Inputs lower than 500 pg were amplified for >12 cycles. High yields were obtained from each input type showing that sequence-ready libraries can be generated from rootless hair. Reads were mapped to respective reference genomes and mitochondrial genomic coverage was determined[3].

HIGHLY FRAGMENTED DNA FROM ROOTLESS HAIR

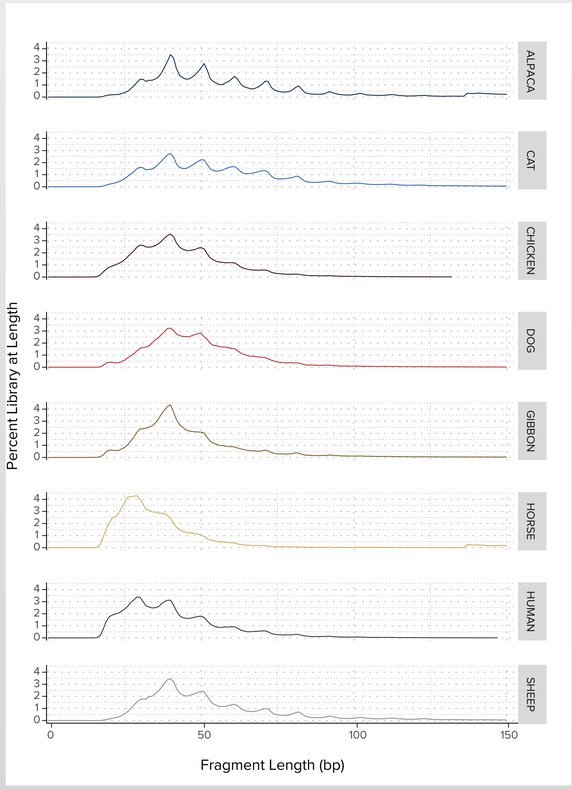


Figure 5. Insert size distribution of libraries made from rootless hair, fur and feather from various livestock and field samples. Libraries were prepared from <250 pg of input DNA. The reads were sequenced to depth of >1 million reads, mapped to respective genomes. SRSLY captures highly fragmented DNA from each of the sample type tested. The fragment size is < 100bp and shows a ~10 bp periodicity.

MITOCHONDRIAL GENOME COVERAGE

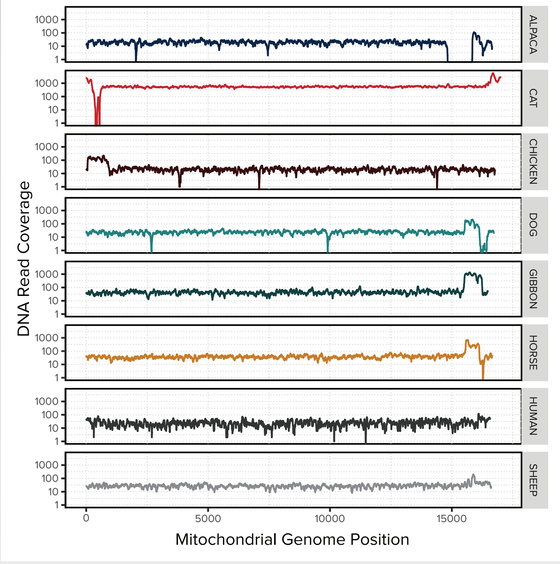


Figure 6. Mitochondrial genomic features can be captured using NGS data from rootless hair to study molecular phylogeny, mtDNA derived pathogenic variants. Libraries from each organism were mapped to their respective genome and read coverage at each position of the mitochondrial genome was calculated[4]. The cat mitochondrial DNA showed very high depth of coverage.

EVOLUTIONARY ANALYSES WITH ROOTLESS HAIR

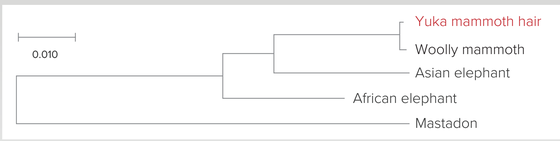


Figure 7. Paleogenomic analyses using rootless hair. Mammoth hair consensus mitochondrial genome sequence was reconstructed using mia (mapping iterative assembler)[5] against the complete *Mammuthus primigenius* mitochondrion reference genome (gi|124056416|ref|NC_007596.2) The mitogenome coverage was 39x. The consensus sequence of the Yuka Mammoth hair was then aligned to the mitogenomes of *Mammuthus primigenius, Elephas maximus, Loxodonta africana,* and *Mammut americanum* (16193 positions) using MEGA[6]. A phylogenetic tree was inferred using a maximum likelihood method. The branch lengths are measured as substitutions per site.

SERIOUS ADVANTAGES

Features	SRSLY	dsDNA	Application
Single strand	√√	✗	cfDNA, oligo, methyl-seq, cDNA
Native ends	√√	✗	cfDNA, oligo
Nicked DNA	√√	✗	cfDNA, aDNA, FFPE
Short fragments	√√	✗	aDNA, hair DNA, cfDNA
Damage DNA capture	√√	✓	FFPE, aDNA, hair DNA
Short protocol time	√√	✓	cfDNA, cDNA
Directional	√√	✓	cDNA, methyl-seq

Table 2. SRSLY can be used in a variety of NGS applications. The method captures fragments that are lost to traditional double stranded prep which expands the utility of the method. In Methyl-Seq and RNA-Seq applications, the methyl-conversion and cDNA synthesis steps result in single-stranded products which are ideal templates for SRSLY. Ancient DNA, FFPE DNA and cfDNA contain damaged, nicked or short DNA which are captured efficiently by this approach.

CLARET BIOSCIENCE | P.O. Box 3052 | Santa Cruz, CA 95063 | +1.707.412.8484 • **FOR MORE INFORMATION** | visit www.ClaretBio.com or write to info@ClaretBio.com

[1]Claret Bioscience LLC, Santa Cruz, CA, 95060, USA, [2]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA, [3]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA, *Corresponding Author