Age Classification of Genomic DNA from Metagenomes

Anton Thomma[1,2], Stefanie Widder[2], Thomas Rattei[1,2]
[1] Department für Computational Systems Biology, Faculty of Life Science, University of Vienna, A-1090 Wien, Austria
[2] Department of Genome Oriented Bioinformatics, Technische Universität München, D-85354 Freising, Germany

Mummies or osseous fragments from protected and covered spaces, where the exposure to the environment and its influences was limited, are particularly valuable for archaeological research. For instance, Neanderthal bones were found in caves unreachable for animals[1], the mammoth remains were retrieved from permafrost-preserved soil[2] and the Tyrolean iceman Ötzi was conserved in a glacier of the Similaun mountains. Aside from morphological and phenotypical classification, the determination of DNA sequences and the subsequent genome analyses have been first applied to mitochondrial DNA and then been extended to genomic DNA. The determination of genomic sequences is accelerated with next-generation sequencing technologies cutting costs and time for sample preparation. Typically also microbial DNA is sequenced in addition to the target (main) organism. The sequenced DNA is mainly analysed using established bioinformatical methods for genome assembly and mapping to close reference genomes.

However, ancient DNA was exposed to degradation, radiation and microbial or fungal activity at the place of discovery. For instance, deamination is one of the most common forms of hydrolytic DNA damage in living organisms[3]. Particularly for cytosine, this leads to the conversion to uracil, which is sequenced as thymine[4]. On the one hand this degradation has to be considered in all bioinformatic analyses of ancient DNA (eq. tuning of parameters or establishing specific workflows). On the other hand these degradation patterns can be used to distinguish ancient DNA from modern contaminations, which is particularly important for microbial sequences.

The Tyrolean iceman Ötzi was found in 1991 and is currently studied and conserved at the EURAC Institute for Mummies and the Iceman in Bolzano, Italy. Within the consortium investigating the ancient DNA fragments extracted from a pelvis bone sample, our aim is the comprehensive characterization of this ancient metagenome which was sequenced on a SOLiD sequencer. As in other palaeontological samples, this metagenome is composed by ancient human sequences from Ötzi and ancient microbial sequences. Due to the discovery record, we also expect microorganisms that recently contaminated the sample. Besides all other project specific bioinformatic analyses, a method for the age classification of the microbial sequences has therefore to be developed, evaluated and applied.

For this age classification, we evaluated and optimized the tool mapDamage[5], which has been very recently developed for Illumina sequencing data. Training data consisting of ancient and recent DNA was necessary to analyse the specific mapping patterns. The human sequences Ötzi metagenome provided the ancient DNA reads whereas modern DNA reads were obtained of the HG00101 library[6]. Mapping of the Ötzi sequencing DNA fragments was performed with an optimized parameter set for sensitive selection of the human Ötzi sequences.

Among other characteristics, the Ötzi reads showed a highly reduced thymine frequency at the last position in the reference genome before the mapped read, followed by an over representation of thymine at the first position of the read. For the modern DNA from the HG00101 mapping, the thymine frequency is not significantly lower at the last position in the reference genome before the mapping. Adenine shows contrary behaviour in Ötzi and modern human mapping. In Ötzi reads the adenine frequency increases before the read and in modern HG00101 reads the frequency decreases.

These differences allowed clear differentiation between the ancient (Ötzi) and the modern (HG00101) human DNA.

We will now apply these methods to the microbial species in the Ötzi metagenome and will show preliminary results on the poster.

1 Noonan, J.P. et al. Sequencing and Analysis of Neanderthal Genomic DNA. Science 314, 1113 -1118 (2006).
2 Hofreiter, M. DNA sequencing: Mammoth genomics. Nature 456, 330-331 (2008).
3 Lindahl, T. Instability and decay of the primary structure of DNA. Nature 362, 709-715 (1993).
4 Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research 29, 4793 -4799 (2001).
5 Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics (2011)
6 SRA Experiment: ERX008207: AB SOLiD System 3.0 paired end sequencing, http://www.ebi.ac.uk/ena/data/view/ERX008207