Cytosine deamination rates analysis to determine species origin in ancient mixed and processed Animal tissues

Matar Rachel and Merheb Maxime*
Department of Biotechnology, American University of Ras Al Khaimah AURAK, Ras Al Khaimah, P.O Box 10021, UNITED ARAB EMIRATES
*maximemerheb@yahoo.fr

Abstract
In living organisms, Cytosine deamination is one of the most common forms of DNA lesion where cytosine is converted to uracil. In such DNA molecules, DNA polymerases will incorporate deoxyadenosine as complementary to the uracil where at those positions a deoxyguanosine should normally have been incorporated, thus, C to T and G to A transitions will appear in the amplified DNA sequences. While such lesion is very abundant in ancient DNA, in the present study our goal in Cytosine deamination analysis is not only to check sequences authenticity but also to compare signatures of sheep and cattle mtDNA substrates in order to determine the species origin of an ancient leather and glue samples in a complicated mixture.

Keywords: Cytosine deamination, ancient DNA, C to T rate, mitochondrial DNA, species identification, mixed tissues.

Introduction
DNA modification is a serious concern for most ancient samples. Cytosine deamination to form uracil represents the most abundant nucleotide modification and appears as a common signature of all aDNA templates. In addition to Cytosines deamination, post-mortem DNA damage can generate many other lesions. Hydantoins (oxidative pyrimidine derivatives) and formamidopyrimidines have been detected in various ancient DNA extracts. Ancient DNA extracts show also large apurinic site (AP site) contents and Single Strand Breaks (SSB) which prevent DNA elongation and hamper PCR amplification success.

Cytosine deamination is quick and causes its conversion to uracil. In such DNA template, DNA polymerases will incorporate deoxyadenosine residues at positions where a deoxyguanosine would normally have been incorporated, thus, C to T and G to A transitions will appear in the amplified DNA sequences. While such lesion is very abundant in ancient DNA, our goal in Cytosine deamination analysis is not only to check sequences authenticity but also to compare signatures of sheep and cattle DNA in both leather and glue samples.

Based on C to T signatures in different animal tissues mixed altogether to cover the hummer heads in an Erard piano created in 1802, we were able to identify the animal origin of the hide and glue in that mixture. Thanks to non-inhibitive DNA extraction method, PCR reactions have revealed that sheep and cattle are the origin of leather/glue combination. Therefore, to identify the species origin of the hide and the glue, we focused on the cytosine deamination rate.

Material and Methods
DNA was extracted and amplified using appropriate ancient DNA techniques and respective of the most scrupulous ancient DNA authentication criteria. DNA extraction and pre-PCR manipulation were performed in highly isolated laboratory.

Samples
The first sample: First we extracted the leather layers covering a hammer head that was in a Grand Piano created by Sebastien Erard in 1802. The hummer is provided by the Museum of Cité De La Musique in Paris where the piano is conserved. The hummer head was covered with 4 leather layers that have undergone independent extractions:

Extraction (a): First, second and third leather layers are together. These layers are containing glue that we cannot isolate.

Extraction (b): The heart of the fourth leather layers. Using very fine tools, sampling of this leather layer was performed delicately under optical microscope. We have taken very small amount of materials (15 mg) but we were aware to take only from the core of this leather layer. Anyway leather and glue separation was quite impossible.

The second sample: The hummer head is constituted with animal based components, leather and glue. Where the leather isolation from glue was impossible, we had chance to find an isolated drop of glue taken from the same hummer head.

DNA Extraction: Samples were extracted separately including an ancient cave bear tooth as carrier effect control. 221 mg of leather (a), 15 mg of leather (b) and 25 mg of the isolated glue manually reduced using sterile tweezers and blades. 10 ml extraction buffer containing 0.5M EDTA (pH 8), 1% N-Laurylsarcosine, 0.5mg/ml Proteinase-K were added to reduced samples and incubated for 16 hr at 37°C under rotation. Compared to bones digestion, we need much more N-Laurylsarcosine and Proteinase-K to perform complete leathers digestion. The
remaining leather or glue powder was collected by centrifugation and only the supernatant was further used in the silica extraction protocol\(^\text{16}\) with modification of the binding and washing buffers composition.

Binding buffer consists of 5M GuSCN, 50mM Tris, 25mM NaCl, 20 mM EDTA, Triton X-100 (1.3%) and pH 8. Washing buffer consists of 70% ethanol. Elution was performed with 300 µL for the leather sample (a and b extracts), and 150 µL for the glue sample of the Elution Buffer EB (Qiagen ®).

PCR: We attempted to amplify the three samples extracts with the five specific primers listed in table 1 (Pig, Cattle, Goat, Chamois and Sheep). PCR reactions were conducted in a total volume of 25 µl using 2.5 units of Taq Gold (Applied Biosystems®) together with 2 mM MgCl\(_2\), 250 µM of each dNTP and 0.5µM of the different species specific primers listed in table 1. In all amplifications four negative controls were added: PCR control, Aerosol control, Extraction control and Carrier effect control. We note that animal serum was avoided in all PCR reactions. A 10 min activation step at 94°C was followed by 55 cycles of denaturation (94°C, 45s), annealing (55°C), extension (72°C, 45s) and a last extension step at 72°C (10 min). PCR products were loaded on 2 % agarose gel electrophoresis.

DNA sequencing: PCR products were further deep sequenced after MID ligation on the 454-FLX Roche sequencing platform (Cogenics®) or cloning (Topo® TA cloning Kit; Invitrogen®). In the latter case, colonies for insertion were screened by PCR (35-45 cycles of denaturation: 94°C, 30 sec; annealing: 55°C, 30 sec and elongation: 72°C, 45 sec) into a 22µl reaction mix consisting in 1µl of universal M13 primers (Forward: 5’-GTGTGTTTCCAGTCAACGTGTTG; Reverse: 5’-TTTCACACAGGAAACAGGTAT) and 1X PCR supermix (Invitrogen) or 1X PCR Master Mix (Promega®) and electrophorograms were checked by eyes and sequences were aligned using Seaview\(^\text{17}\). 454-FLX results were analyzed using Galaxy (http://main.g2b2x.psu.edu/)\(^\text{1,11,17}\).

Sequence Identification: All sequences were identified by SAP (Statistical Assignment Package) software,\(^\text{6,15}\) a method using Bayesian approach to statistical assignment. The method has advantages compared to the online BLAST search tool by including phylogenetic information and providing statistically meaningful measures of confidence (posterior probabilities) to the taxonomic assignment.

Cytosine deamination analysis: C to T + G to A substitution rates were calculated for glue and leather samples in D-loop DNA sequences amplified with the specific cattle and specific sheep primers and sequenced (454-FLX Roche and cloning). For each position in DNA sequence carrying a C or G we divided the number of sequencers revealing a substitution (C to T; G to A) by the total number of sequences, yielding \(\lambda I\) for that particular position. To take into account the length and the base composition of the sequence we divided the sum of all \(\lambda I\) per sample/marker by the number of all potentially affected positions (C and G positions in the sequence), yielding \(\lambda’\), the average deamination rate over all positions\(^\text{3,11}\).

Results

PCR, sequencing and species identification: With each specific couple of primer listed in table 1, three independent PCR have been performed for both samples where in each PCR we added four controls described previously. Agarose gel electrophoresis has shown that we have amplicons in specific sheep and specific cattle PCRs for leather as well as for glue (Table 2). No amplicon has been shown for all controls and not for other specific PCR (pig, chamois and goat). For each PCR, two positive amplicons have been sequenced and the numbers of sequences are listed in table 2 for each PCR for leather and glue samples. Using SAP software (Statistical Assignment Package), sequences identities for all amplicons were matching the target species (98% Bos and 100% Ovis in leather and glue).

Cytosine deamination analysis: As explained before, \(\lambda’\) is the average of deamination rate over all C and G positions (\(\lambda’ = C + T \rightarrow G + A\) substitution rate). \(\lambda’\) was calculated on sequences obtained by three independent PCR (Table 2). Thus, substitution rate in specific sheep and specific cattle amplification were calculated for glue and leather samples. Where rates are very similar between extract -a- and -b- in leather sample, we calculated the average rates for both.

Cytosine deamination rates are as follows:

\[
\begin{align*}
\lambda’_{\text{sheep-leather}} &= 0.031; \\
\lambda’_{\text{sheep-glue}} &= 0.035; \\
\lambda’_{\text{cattle-leather}} &= 0.015 \\
\lambda’_{\text{cattle-glue}} &= 0.004.
\end{align*}
\]

Discussion and Conclusion

PCR, sequencing and species identification: Amplifications followed by sequencing and sequences identifications showed the presence of sheep and cattle mtDNA in both samples (leather and glue). While the extracted glue drop has any contact with the leather, these results allow concluding that the glue has been made from sheep and cattle tissues. A natural leather piece is normally made from only one species. As we described before, the extracted leather is not pure but co-extracted with glue. Therefore, the correlation used to conclude the origin of the glue sample cannot be applied in the leather case where we detected sheep and cattle mtDNA. Thus, conventional PCR, sequencing and sequences identification results are not enough to determine the origin of the leather. Sheep and cattle mtDNA presence in both samples lead to two eventual possibilities (i) No amplifiable mtDNA from the leather and the detected mtDNA is only from the glue or (ii) We have amplifiable mtDNA from leather but its identity is quenched by the glue mtDNA if leather origin is sheep or cattle.
Cytosine deamination analysis: Sheep mtDNA shows light difference in the cytosine deamination rates between leather and glue samples (Ratio: $\lambda_{\text{sheep-leather}} / \lambda_{\text{sheep-glue}} \geq 0.9$). Thus, sheep mtDNA sequences have almost the same cytosine deamination signature in glue and leather samples ($\lambda_{\text{sheep-leather}} = \lambda_{\text{sheep-glue}}$). This result suggests that leather fraction in the first sample has not offered sheep mtDNA but this last is issued only from the glue fraction.

Contrarily in cattle mtDNA, the cytosine deamination rate is obviously higher in leather sample than in glue sample ($\lambda_{\text{cattle-leather}} / \lambda_{\text{cattle-glue}} \geq 5$). If leather fraction in the first sample has not offered cattle mtDNA, we should obtain almost the same $\lambda$ for cattle sequences (same modification signature) in leather and glue samples. But $\lambda_{\text{cattle}}$ is 5 times higher in leather than in glue. This suggests that leather has offered cattle mtDNA showing much more cytosine deaminated sites. Cytosine deamination rates are detailed in table 3.

Acknowledgement
We thank Paleogenetic and Molecular Evolution team at the IGFL and Palgène platform for technical support. This work funding was supported by the Cité de la Musique, Paris; ANR and CNRS-France.

References

Table 1
<table>
<thead>
<tr>
<th>Primers ID</th>
<th>Mitochondrial gene</th>
<th>Target length</th>
<th>Sequence</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIG85L530</td>
<td>D-loop</td>
<td>84 bp</td>
<td>ACATACAATATATGTGACCCCCCA TTAATGCACGACGTACATAGG</td>
<td>Pig</td>
</tr>
<tr>
<td>PIG85ANCR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOV8516335F</td>
<td>D-loop</td>
<td>87 bp</td>
<td>ACCCCCAAAGCTGAAGTTCT TTTAATACTGATAAGGCTC</td>
<td>Cattle</td>
</tr>
<tr>
<td>BOV85REV1R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAP732F</td>
<td>D-loop</td>
<td>72 bp</td>
<td>ACTRTATATCTACACCTACAC CATAAAATGTAAGTACATACA</td>
<td>Goat</td>
</tr>
<tr>
<td>CAP732R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rup-12S-L</td>
<td>12sRNA</td>
<td>76 bp</td>
<td>CCCTCTCCAAGYRAATACAGGA TGTACGACTTGTCTCCTCTTG</td>
<td>Chamois</td>
</tr>
<tr>
<td>Rup-12S-H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheep-87F</td>
<td>D-loop</td>
<td>71 bp</td>
<td>CCTGTCCATTAGATCGAGCTTG GAGGGATCCTTGTAAGGCGG</td>
<td>Sheep</td>
</tr>
<tr>
<td>Sheep-157R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Results of specific PCR of leather extract (a and b) and glue extract. SAP results are presented in posteriori probability ($p$-$p$)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Leather extract (a)</th>
<th>Leather extract (b)</th>
<th>Glue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers</strong></td>
<td>BOV8516335F/BOV85REV1R: specific cattle</td>
<td>BOV8516335F/BOV85REV1R: specific sheep</td>
<td>BOV8516335F/BOV85REV1R: specific sheep</td>
</tr>
<tr>
<td>Seqencing</td>
<td>Cloning</td>
<td>Cloning</td>
<td>454</td>
</tr>
<tr>
<td># of sequences</td>
<td>94</td>
<td>89</td>
<td>398</td>
</tr>
<tr>
<td>SAP results ($p$-$p$)</td>
<td>Cattle (98%)</td>
<td>Sheep (100%)</td>
<td>Cattle (98%)</td>
</tr>
</tbody>
</table>
Table 3
Cytosine deamination rate here we focus C/T+G/A modifications in leather and glue. λ’ cattle means C/T+G/A modifications for sequences amplified with specific cattle primers. λ’ sheep means C/T+G/A modifications for sequences amplified with specific sheep primers. To compare cytosine deamination rates we reported the ratios R and R’.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Leather extract (a)</th>
<th>Leather extract (b)</th>
<th>Glue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primers</td>
<td>BOV8516335F/BOV85REV1R specific cattle</td>
<td>BOV8516335F/BOV85REV1R specific cattle</td>
<td>BOV8516335F/BOV85REV1R specific cattle</td>
</tr>
<tr>
<td>seqencing via</td>
<td>Cloning</td>
<td>Cloning</td>
<td>Cloning</td>
</tr>
<tr>
<td># sequences</td>
<td>94,0000</td>
<td>89,0000</td>
<td>398,0000</td>
</tr>
<tr>
<td>λ’ A to G</td>
<td>0,0046</td>
<td>0,0050</td>
<td>0,0005</td>
</tr>
<tr>
<td>λ’ T to C</td>
<td>0,0020</td>
<td>0,0056</td>
<td>0,0023</td>
</tr>
<tr>
<td>λ’ G to A</td>
<td>0,0106</td>
<td>0,0197</td>
<td>0,0105</td>
</tr>
<tr>
<td>λ’ C to T</td>
<td>0,0118</td>
<td>0,0135</td>
<td>0,0046</td>
</tr>
<tr>
<td>λ’ C/T+G/A</td>
<td>0,0225</td>
<td>0,0331</td>
<td>0,0151</td>
</tr>
</tbody>
</table>

λ’ C/T+G/A
R=λ’ sheep_leather / λ’ sheep_glue ~ 0,9
R’=λ’ cattle_leather / λ’ cattle_glue ~ 5


(Received 02nd December 2015, accepted 15th January 2016)