Oral Signature Screening Database for Palaeoproteomic Analyses of Dental Calculus

Authors: Madeleine Bleasdale[1], Nicole Boivin[1], Kristine Korzow Richter[1]

Affiliations

1. Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany.

Rationale

As the proteomic analysis of dietary peptides retrieved from archaeological materials becomes more common[1–3], new methods of authentication and screening need to be considered. Here we present an Oral Signature Screening Database (hereafter OSSD) developed for ancient dental calculus. By using the OSSD as a screening tool we hope users will be able to quickly establish if samples have an oral signature allowing them to identify potentially problematic samples as well as those that may be good candidates for endogenous dietary proteins.

The extraction and identification of ancient dietary proteins from dental calculus is a growing field of research in archaeology[2–5]. For such studies it is fundamental that the dietary proteins reported are endogenous (i.e they became entrapped in the calculus during formation) opposed to modern contamination. One method to investigate whether such proteins are truly "ancient" is through the estimation of deamidation rates of glutamate and aparagene[6–8]. While these methods have great potential when applied to well-preserved archaeological materials[8], there are some challenges when they are applied to poorly preserved ancient dental calculus samples. Firstly, it requires that deamidation is not induced during extraction[9,10]. Secondly, a large number of endogenous peptides are needed in order to have adequate deamidation sites for statistical models. Thirdly, it cannot authenticate individual peptides, only the entire identified sample or a sufficiently large subset of an identified sample[3]. Finally, due to variations in deamidation rates between different peptides and different sites within a peptide[11], it is best suited to samples that have a high coverage of a small number of proteins.

Although proteins have demonstrable longevity and can survive in some contexts for millions of years[12] there are many challenges when working with proteins extracted from poorly preserved archaeological materials. The extraction methods used for low abundant samples often need to be optimized in ways that may induce deamidation in the laboratory. As previously discussed, it can also be difficult to use deamidation rates on low-abundant samples because there is a smaller percentage of endogenous peptides[2] in a highly complex sample[13]. Often there are few dietary peptides retrieved which, collectively, do not offer a sufficient number of deamidation sites for statistical analysis.

Genetic studies have shown calculus to be an excellent record of the human oral microbiome[14,15]. It is therefore anticipated that a well preserved dental calculus sample would include an oral signature[13]. However, this “oral signature” is not routinely assessed or reported in a standardised format in proteomic studies of ancient dental calculus. In part, this

is because the oral microbiome is diverse and strongly related to environment and diet, and there are reported distinctions between the oral microbiomes of modern plaque and ancient calculus[16]. While there is already a comprehensive database for the Human Oral Microbiome (eHOMD)[17] we selected a restricted list of oral bacteria proteins found commonly in ancient dental calculus in order to produce a screening tool that requires minimal computational time.

Here we present a Oral Signature Screening Database (OSSD) which includes common contaminants, proteomes from a subset of the most common oral microbes, and human inflammatory response proteins commonly found in archaeological dental calculus samples. At this time the OSSD is not comprehensive and we expect and encourage discussion. As more ancient dental calculus results are published the database will be tested, refined and new versions will be made available. Nevertheless, even in its current format the OSSD provides a quick method of assessing the "oral signature" of a sample which can enable further exploration of the authenticity of results.

Methods

The protein list for the database was created by finding commonalities amongst published datasets for dental calculus[2,5] as well as unpublished results generated by the Palaeoproteomics Lab Group in Jena (MPI-SHH). The full list of proteins in the database are presented in Table 1. Proteins were divided into four categories: lab contaminants, common contaminants, oral microbiome and immune response. Common lab contaminants include trypsin, the enzyme used during the extraction process, and serum albumin which is often a contaminant in modern proteomics facilities. The common contaminants list includes collagens and keratins which are frequently introduced through sample handling and proteins associated with the burial environment.

The primary aim of the OSSD is to have a quick method to assess the oral signature in archaeological dental calculus samples. While we acknowledge oral biomes can contain numerous bacterial species and be highly variable, it is not the purpose of the OSSD to fully capture this diversity. Therefore we only selected a subset of common oral bacteria identified in ancient dental calculus samples, such as members of the "red complex" (*Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*) which are associated with periodontal disease[15,18]. In order to ensure a short run time (<30 mins) for the OSSD when used with common MS/MS data analysis tools we selected 11 bacteria proteomes in total. We recognise the list of oral bacteria is not extensive and that the OSSD is not a substitute for the comprehensive oral database eHOMD (expanded Human Oral Microbiome Database) which contains over 700 microbial species[17]. The OSSD is a screening tool and we would therefore recommend the use of other databases, such as eHOMD, for in-depth assessment of bacterial proteomes. In addition to oral bacteria discussed above, we also included some commonly identified proteins associated with the human immune response (Lysozyme C, Neutrophil elastase, Cathepsin G).

The OSSD was tested on a number of published dental calculus samples and associated blanks: 11 samples from PXD009603[2], 15 samples for PXD012893[3] and 14 samples PXD008217[5]. In addition, we tested it against published results of archaeological and

modern bones and sediments: 16 from PXD014657[19] and 12 from MassIVE MSV000083687 (doi:10.25345/C5G04C)[20], as well two internal bone extractions (unpublished). The database was tested on Byonic Protein Metrics Inc.[21] with the following settings: non-specific digestion, a precursor mass tolerance of 5ppm, a fragment mass tolerance of 0.05Da, carbamidomethyl of cysteine as a fixed modification, variable modifications (2 common, 1 rare) as deamidation of asparagine and glutamate (2 common), oxidation of lysine and methionine (2 common), phosphorylation of serine and threonine (1 common), glutamate or glutamic acid to pryo-glutamate (1 rare), and acetyl at the n-terminus (1 rare). Proteins were manually assigned to each of the four categories and totals calculated. Proteins were considered authentic if they had at least four peptides assigned and had a log probability of greater than one or greater than the highest scoring decoy whichever was higher. For all of the samples the average number of contaminant proteins was 6. Therefore, in order to pass OSSD threshold and confirm an "oral signature", samples required at least 12 total proteins or twice the number of average contaminants, with 50% of proteins assigned to oral microbiome or immune response.

Results and Discussion

To test the OSSD we ran it against published and unpublished results from bone (n = 24), calculus (n = 32) and soil (n = 1), as well as associated extraction blanks (n = 11). Samples were extracted with different methods (gelatinization, GASP, FASP, SP3) in different laboratories and are from modern and archaeological contexts from across the world. As anticipated, all the bones and blank samples failed to meet the OSSD threshold (Table 2). Thirty-one out of thirty-two calculus samples passed our OSSD threshold, including all calculus samples that were reported to have milk peptides.

While we acknowledge this a preliminary version of the OSSD we would encourage its use as an initial screening step before more in-depth data analysis. Although more testing is needed to identify cut-off values for authenticity for different methods and regions of the world, we believe the database can be used to quickly identify potentially problematic samples as well as those with the greatest potential to provide authentic, endogenous dietary-related proteins. In addition, more robust testing needs to be conducted on samples with no oral signature but known sources of modern contamination from handling, such as those from teaching collections. Finally, this method does not fix the problem of the need to authenticate individual peptides. Samples which have an endogenous oral signature could still be contaminated with modern dietary peptides which would not be detected using this method. Special consideration should be made for ancient samples which could have been treated with animal-based glues as part of conservation practices[22]. Additionally, using the whole proteome of oral microbes likely allows for overlap between proteins found in both oral microbes and soil microbes.

Further work is needed to expand and improve the OSSD as more published datasets become available but the first version of the OSSD provides a firm framework to try and validate samples. We encourage discussion to improve this database for future research.

Compatibility and Version History

The OSSD is a FASTA file which has been tested in both MASCOT and Byonic and is suitable for other pipelines including, but not limited to, MaxQuant and Peaks. The first version (v.0) is just one FASTA file with all of the proteins including several dairy proteins. The version we recommend using is the first official version (v.1.0) which contains FASTA files for the different components (lab contaminants, common contaminants, oral microbiome proteins, and human immune response proteins) as well as a complete FASTA file which contains all of the proteins together. Dietary proteins are not included in this version.

References

1. Yang, Y. *et al.* Proteomics evidence for kefir dairy in Early Bronze Age China. *J. Archaeol. Sci.* **45**, 178–186 (2014).
2. Hendy, J. *et al.* Proteomic evidence of dietary sources in ancient dental calculus. *Proc. Biol. Sci.* **285**, (2018).
3. Charlton, S. *et al.* New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeol. Anthropol. Sci.* **11**, 6183–6196 (2019).
4. Hendy, J. *et al.* A guide to ancient protein studies. *Nat Ecol Evol* (2018) doi:10.1038/s41559-018-0510-x.
5. Jeong, C. *et al.* Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11248–E11255 (2018).
6. van Doorn, N. L., Wilson, J., Hollund, H., Soressi, M. & Collins, M. J. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **26**, 2319–2327 (2012).
7. Simpson, J. P. *et al.* The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *J. Archaeol. Sci.* **69**, 29–38 (2016).
8. Ramsøe, A. *et al.* DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *J. Archaeol. Sci.* **115**, 105080 (2020).
9. Hao, P., Ren, Y., Datta, A., Tam, J. P. & Sze, S. K. Evaluation of the effect of trypsin digestion buffers on artificial deamidation. *J. Proteome Res.* **14**, 1308–1314 (2015).
10. Procopio, N. & Buckley, M. Minimizing Laboratory-Induced Decay in Bone Proteomics. *J. Proteome Res.* **16**, 447–458 (2017).
11. Robinson, N. E. & Robinson, A. B. Prediction of protein deamidation rates from primary and three-dimensional structure. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4367–4372 (2001).
12. Demarchi, B. *et al.* Protein sequences bound to mineral surfaces persist into deep time. *Elife* **5**, (2016).
13. Jersie-Christensen, R. R. *et al.* Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nat. Commun.* **9**, 4744 (2018).
14. de La Fuente, C., Flores, S. & Moraga, M. DNA FROM HUMAN ANCIENT BACTERIA: A NOVEL SOURCE OF GENETIC EVIDENCE FROM ARCHAEOLOGICAL DENTAL CALCULUS. *Archaeometry* **55**, 767–778 (2013).
15. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
16. Velsko, I. M. *et al.* Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, 102 (2019).
17. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010**, baq013 (2010).
18. Tanner, A. C. R. & Izard, J. Tannerella forsythia, a periodontal pathogen entering the genomic era. *Periodontol. 2000* **42**, 88–113 (2006).

19. Tsutaya, T. *et al.* Palaeoproteomic identification of breast milk protein residues from the archaeological skeletal remains of a neonatal dog. *Sci. Rep.* **9**, 12841 (2019).
20. Richter, K. K. *et al.* What's the Catch?: Archaeological application of rapid collagen-based species identification for Pacific Salmon. *J. Archaeol. Sci.* **116**, 105116, (2020).
21. Bern, M., Kil, Y. J. & Becker, C. Byonic: Advanced Peptide and Protein Identification Software. in *Current Protocols in Bioinformatics* (eds. Baxevanis, A. D., Petsko, G. A., Stein, L. D. & Stormo, G. D.) vol. 79 1393 (John Wiley & Sons, Inc., 2012).
22. Nicholson, G. J., Tomiuk, J., Czarnetzki, A., Bachmann, L. & Pusch, C. M. Detection of bone glue treatment as a major source of contamination in ancient DNA analyses. *Am. J. Phys. Anthropol.* **118**, 117–120 (2002).