Commentary

AMELY deletion is not detected in systematically sampled reference populations: A Reply to Štamfelj

Glendon J. Parker a,*, Tammy Buonasera a, b, Julia M. Yip a, Jelmer W. Eerkens b, Michelle Salemi c, Blythe Durbin-Johnson d, Randall Haas b, Jane E. Buikstra e, Haagen Klaus f, David M. Rocke d, Brett Phinney c

a Department of Environmental Toxicology, University of California, Davis, Davis, CA 95616, USA
b Department of Anthropology, University of California, Davis, Davis, CA 95616, USA
c Proteomic Core Facility, University of California, Davis, Davis, CA 95616, USA
d Department of Public Health Sciences, University of California, Davis, CA 95616, USA
e School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287, USA
f Department of Sociology and Anthropology, George Mason University, Fairfax, VA 22030, USA

ARTICLE INFO

Keywords
Sex estimation
Proteomic sex estimation
Juvenile sex estimation
Amelogenin
Amelogenin deletion
Enamel proteome

ABSTRACT

Biomolecular sex estimation promises to fill a major gap in the bioarchaeological record by providing estimates of biological sex for skeletal remains with degraded or ambiguous osteological sex-specific markers. Genomic and proteomic sex estimation, like all analytical methods, have limitations and require frameworks to address the problems of low signal samples and the inevitable conflicting results when other methods are used. Proteomic sex estimation is based on the detection of sex-chromosome specific amelogenin protein fragments in enamel using mass spectrometry. Enamel from male individuals contains amelogenin fragments from both the X- and Y-chromosome versions of amelogenin, and enamel from female individuals contains fragments from only the X-chromosome protein. The method is sensitive, robust, quantifiable and reproducible. Researchers have developed, and continue to develop, frameworks to address theoretical problems associated with low levels of detection and conflicting sex estimates that will inevitably occur when multiple methods are used on a sufficiently large dataset. Štamfelj reminds readers that structural variants of the Y-chromosome that delete the AMELY gene have been detected in forensics and clinical casework. Since this phenomenon would also account for the absence of the AMELY protein in enamel it should therefore be mentioned as an alternative hypothesis by investigators, along with female sex and low peptide signals in mass spectrometry. In his meta-analysis Štamfelj concludes that this is an intrinsic limitation of biomolecular sex estimation, particularly when examining South Asian populations, and should be incorporated in standard analytical sex estimation frameworks. In this comment, we test this assertion by examining the occurrence of AMELY deletion in the systematically sampled, high coverage, large scale, and well-curated populations of the 1000 Genomes Project and Exome Sequencing Project. When using SNP loci in the open reading frame of AMELY, structural deletion was not detected in either project. Confident probabilities of occurrence with associated intervals cannot be determined from null values. We conclude from this that, for now, AMELY deletion should have no bearing on routine biomolecular sex estimation.

Accurate sex estimation is a starting point for developing and addressing basic questions about the lives and culture of individuals. Historically biological sex estimation has relied on sexually dimorphic osteological markers that can be degraded or be ambiguous due to lack of development. These factors reduce the sensitivity of sex estimation, resulting in a significant loss of information from the bioarchaeological

* Corresponding author.
E-mail addresses: gjparker@ucdavis.edu (G.J. Parker), jmyip@ucdavis.edu (J.M. Yip), jweerkens@ucdavis.edu (J.W. Eerkens), msalemi@ucdavis.edu (M. Salemi), bpdurbin@ucdavis.edu (B. Durbin-Johnson), wrhaas@ucdavis.edu (R. Haas), buikstra@asu.edu (J.E. Buikstra), hklau@gmu.edu (H. Klaus), dmrocke@ucdavis.edu (D.M. Rocke), bsphinney@ucdavis.edu (B. Phinney).

https://doi.org/10.1016/j.jas.2021.105354
record, particularly of infants and children. Recent advances in biomolecular analysis of nucleotide or proteomic information provide alternatives that are scientifically rigorous, sensitive, and can also be applied to subadult or degraded materials. These methods provide, for the first time, sex estimates on these incompletely analysed populations and have the potential to fill a major gap in the bioarchaeological record.

Proteomic sex estimation relies on detection of peptides that are expressed in human enamel from genes residing on sex-specific chromosomes. The most characterized sex-specific gene family, the amelogenins, are therefore expressed in the most robust human tissue (Buonasera et al., 2020; Parker et al., 2019; Stewart et al., 2016, 2017; Welker et al., 2020; Cappellini et al., 2018; Lugli et al., 2019; Zsigand et al., 2020). Because protein is chemically stable, and the amelogenin peptides bind to the biomolecular interfaces in enamel, the sex-chromosome specific signals are likewise stable and show no decrease in signal over archaeological time and potentially over even deeper timescales (Buonasera et al., 2020; Parker et al., 2019; Welker et al., 2019, 2020; Cappellini et al., 2018; Poinar and Stankiewicz, 1999). Intact signal is affected by the fact that the Y-chromosome isoform of amelogenin (AMELY_HUMAN) is expressed at around 10% of the X-chromosome isoform (AMELX_HUMAN), at both the transcript and protein level (Parker et al., 2019; Lattanzi et al., 2005; Salido et al., 1992). Due to the high level of degradation and diagenesis in archaeological material, our previous research proposed that there could be situations where male individuals with low levels of detected enamel protein would have AMELX_HUMAN detected but not AMELY_HUMAN (Parker et al., 2019). While the confident detection of AMELY_HUMAN peptides is taken as an unambiguous indicator of male sex, the absence of AMELY_HUMAN peptides could be due to either a male false negative sample, or female sex. We further posited that the possibility of a male false negative sample would decrease as the quality of the enamel proteome increased. With this assumption we then developed a logistic regression curve to estimate the probability of female sex as a function of the signal from AMELX_HUMAN, which would be equivalently subjected to degradation and diagenesis (Parker et al., 2019).

When multiple biological analyses are applied to degraded material it is not surprising that conflicts will occur. It is incumbent on those developing novel methodologies to place each method in context, delineating limitations and providing frameworks to resolve conflicting results. With this in mind our research group conducted a large scale study on fifty five individuals from two Northern Californian archaeological contexts that were thoroughly examined using current osteologic, genomic and proteomic sex estimation methods (Buonasera et al., 2020). We conclude from this study that proteomic sex estimation resulted in confident sex estimates in all samples, with genomic and osteologic methods providing confident estimates in 64% and 25% of individuals respectively. As expected, conflicts between genomic and proteomic methods did occur. These inconsistent results occurred in individuals with lower levels of detected DNA, below a threshold of 100, 000 total genomic reads. Differences between proteomic and genomic sex estimates were also concentrated in contingent, “consistent with…” genomic sex estimates. These findings allowed us to propose a framework for reconciling potentially divergent sex estimates using osteological, genomic or proteomic approaches. The final framework for prioritizing sex estimates was: proteomic sex estimation > genomic sex estimation with greater than 100,000 total reads > genomic sex estimation with definitive estimates using the Rx method > definitive osteological sex estimation > conditional osteological sex estimation > conditional estimates resulting from other biomolecular methods (Buonasera et al., 2020; Mittnik et al., 2016). While conflicts occurred when DNA data quality was poor, no such pattern occurred with proteomic data. This is a strong indicator that, in this sample, inconsistent results were due to poor-quality DNA and not poor-quality proteomic data. Importantly proteomic sex estimation was also successfully conducted on infant and foetal remains, as well as two partial cremations.

Our efforts therefore have focused on developing frameworks to address practical limitations to proteomic sex estimation: ambiguous estimates resulting from low signal samples, and developing frameworks to resolve conflicting estimates that result from using multiple sex estimation methods.

The above comment by Dr Stamfelj on our publication in this journal in 2019 makes several broad points about the application of biomolecular sex estimation methods that rely on proteomic or genomic detection of the Y-chromosome isoform of amelogenin in archaeological material. Many of Stamfelj’s points agree with our previous discussions on proteomic and genomic sex estimation both in our introduction of the method in this journal in 2019 (Parker et al., 2019) and in a recent validation of the method published in Scientific Reports (Buonasera et al., 2020). The initial correspondence by Stamfelj however was focused primarily on the possibility of major structural variation to the Y-chromosome resulting in loss of the Y-chromosome form of amelogenin (AMELY). Stamfelj conducted a meta-analysis of incidences of Y-chromosome structural variation and concurrent loss of the AMELY gene. From this non-systematic sampling he concludes that this event is common enough, particularly in South Asian populations, to be routinely mentioned as a possibility in South Asia with no detected AMELY nucleotides or AMELY_HUMAN peptides. He asserts that this should be taken seriously as an alternative hypothesis when considering skeletal remains that only contain AMELX_HUMAN peptides or AMELY nucleotides in genomic sex estimation. This issue was not addressed in our initial or subsequent reports on the method, primarily because structural variation of this magnitude is rare (Sudmant et al., 2015). We therefore thank Stamfelj for raising the issue, allowing us to further develop our decision to not include it in our developed analytical frameworks and algorithms. At the same time we can now address a misconception in the field, namely that AMELY deletion is a common phenomenon or at least common enough to be routinely included in other explanations for the absence of AMELY_HUMAN peptides, such as low signal or female sex.

The presence of structural Y-chromosomal variants that delete the AMELY gene have been a feature of the forensic literature for more than twenty years (Santos et al., 1998; Takayama et al., 2009; Roffey et al., 2000; Michael and Brauner, 2004; Thangaraj et al., 2002; Steinlechner et al., 2002; Brinkmann, 2002; Butler, 2012). These findings reflect the large scale of forensic casework DNA-typing analyses. Research on this genetic phenomenon is extensive and well examined (Lattanzi et al., 2005; Santos et al., 1998; Takayama et al., 2009; Roffey et al., 2000; Michael and Brauner, 2004; Thangaraj et al., 2002; Steinlechner et al., 2002; Brinkmann, 2002; Jobling et al., 2007). Caution should be applied however, when cumulatively extrapolating outward from small-scale studies to draw conclusions about a larger population, since sampling biases may be introduced (Simundic, 2013; Buckleton et al., 2018). Fortunately, carefully randomized samples that investigate genetic variation in major human populations, with appropriate levels of scale, curation and quality control, are available to researchers in the form of the 1000 Genomes Project and the Exome Sequencing Project (Genomes Project Consortium et al., 2012, 2015; Tennessen et al., 2012; Fu et al., 2013). This provides an opportunity to rigorously quantify the phenomenon of AMELY deletion in well-constructed, systematically sampled reference populations and evaluate whether this possibility should be considered in a standard analytical workflow for biomolecular sex estimation.

The 1000 Genomes Project is large, consisting of 1271 female and 1233 male subjects and is split into 5 major population groups: African, European, East-Asian, South-Asian and South American. It consists of high coverage, curated genomes with consistent and high levels of quality control with appropriate systematic sampling of respective reference populations to minimize introduced biases, for example by removal of related individuals (Genomes Project Consortium et al., 2015). If the analysis of Dr Stamfelj is correct, and AMELY deletion is present in high quality reference populations, then the number of individuals containing the locus of single nucleotide polymorphisms from the coding region of AMELY should be less than the number of males in the project. The
misseense single nucleotide polymorphisms rs35815655, rs200834952 and rs2071394, that correspond to codons 70, 98 and 166 in AMELY, occur in all 1233 male subjects. This indicates that there are no examples of structural variants deleting the coding region of AMELY in any populations in the 1000 Genome Project. This includes 260 males in the South Asian reference population. It also includes the reference American population that most closely approximates the genome of the individual with conflicting estimates from Soro Mik’aya Patjxa, Peru highlighted in the analysis of Stamfelj (Parker et al., 2019). A similar result was obtained when using the larger populations examined in the Exome Sequencing Project (Fu et al., 2013). This high-quality study focused on European- and African-American populations that contained 1873 and 571 males respectively. Likewise, the same result occurred for SNPs rs35815655, and two SNPs that were unique to this study, rs376163078 and rs373182951 that correspond to codons 131 and 175 respectively. In this reference population there was no difference between AMELY SNP loci counts and the number of males in this study. Consistent with this, the findings, the comparison study of all three methods, conducted by our group on a sample of 55 archaeological skeletons described above, did not observe any examples of high quality osteological estimates being inconsistent with high quality proteomic or genomic sex estimates (Buonasera et al., 2020). The one example of a conflict between proteomic and osteological sex estimation, from the Soro Mik’aya Patjxa sample and highlighted by Stamfelj, was of an individual with poor preservation status and was not a confident osteological sex estimate (Parker et al., 2019). We can conclude from these analyses that the prevalence of AMELY deletion is sufficiently rare that it is not present in large well-constructed, curated, and systematically sampled reference populations, including the major South Asian reference population of the 1000 Genomes Project. Given the non-systematic sampling in the Stamfelj meta-analysis, the finding of an increased likelihood of AMELY deletion in South Asian populations is therefore problematic. Perhaps more importantly, given these null values, it is difficult to quantify and delineate confidence intervals. However, as demonstrated by Stamfelj, AMELY deletion is a non-zero number and the phenomenon is well documented in the forensic and clinical literature (Jobling et al., 2007). The possibility alone may be sufficient for many investigators to include it in their analytical frameworks, particularly those in the forensics field. The central question here is whether a remote and poorly defined possibility should be routinely discussed in physical anthropological and bioarchaeological analysis. In our view, at some point alternative hypotheses become unlikely and do not make a valid contribution to the analysis of sex estimation. The challenge for inclusion here is two-fold: AMELY deletion is exceedingly rare and confidence intervals are difficult to delineate. When clear population probabilities with demonstrated confidence intervals exist, then it would be appropriate to reconsider this point. Until then, it is reasonable to conclude that consideration of AMELY deletion is not necessary in routine sex estimation using proteomic or genomic methods.

References

