Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification

Kirsten A. Ziesemer[1], Allison E. Mann[2], Krithivasan Sankaranarayanan[2], Hannes Schroeder[1,3], Andrew T. Ozga[2], Bernd W. Brandt[4], Egija Zaura[4], Andrea Waters-Rist[1], Menno Hoogland[1], Domingo C. Salazar-Garcia[5], Mark Aldenderfer[6], Camilla Speller[7], Jessica Hendy[7], Darlene A. Weston[1,8], Sandy J. MacDonald[7], Gavin H. Thomas[7], Matthew J. Collins[7], Cecil M. Lewis[2], Corinne Hofman[1], and Christina Warinner[2]

1. Leiden University, the Netherlands; 2. University of Oklahoma, Department of Anthropology, USA; 3. University of Copenhagen, Center for Geogenetics, Denmark; 4. University of Amsterdam, the Netherlands; 5. University of Cape Town, South Africa; 6. University of California, USA; 7. University of York, UK; 8. University of British Columbia, Canada

Introduction:

Many of our host microbiota are essential for critical biological functions including immune development, nutrition, and protection from pathogens. Industrial societies are characterized by a loss of microbial diversity due to anthropogenic changes to our environments. As such, understanding the ancestral state of the human microbiome is necessary. Dental calculus is an excellently preserved source for ancient oral microbial communities. Characterization of the ancient oral microbiome has primarily been accomplished through targeted amplification of the V3 region of the 16S rRNA gene. In practice, however, amplification of this region in ancient samples often results in highly skewed taxonomic profiles. ***The purpose of this study is to systematically test whether targeted amplification of variable regions of the 16S rRNA gene accurately reconstructs ancient human microbiome communities.*** Ancient dental calculus samples (n=107) were selected for analysis from seven geographically diverse sites: Middenbeemster, the Netherlands (n=76); Rupert's Valley, St. Helena (n=15); Anse à la Gourde, Guadeloupe (Caribbean) (n=5); Lavoutte, St. Lucia (Caribbean) (n=5); Tickhill, Yorkshire, UK (n=4); Samdzong, Nepal (n=1); and Camino del Molino, Spain (n=1) (Figure 1).

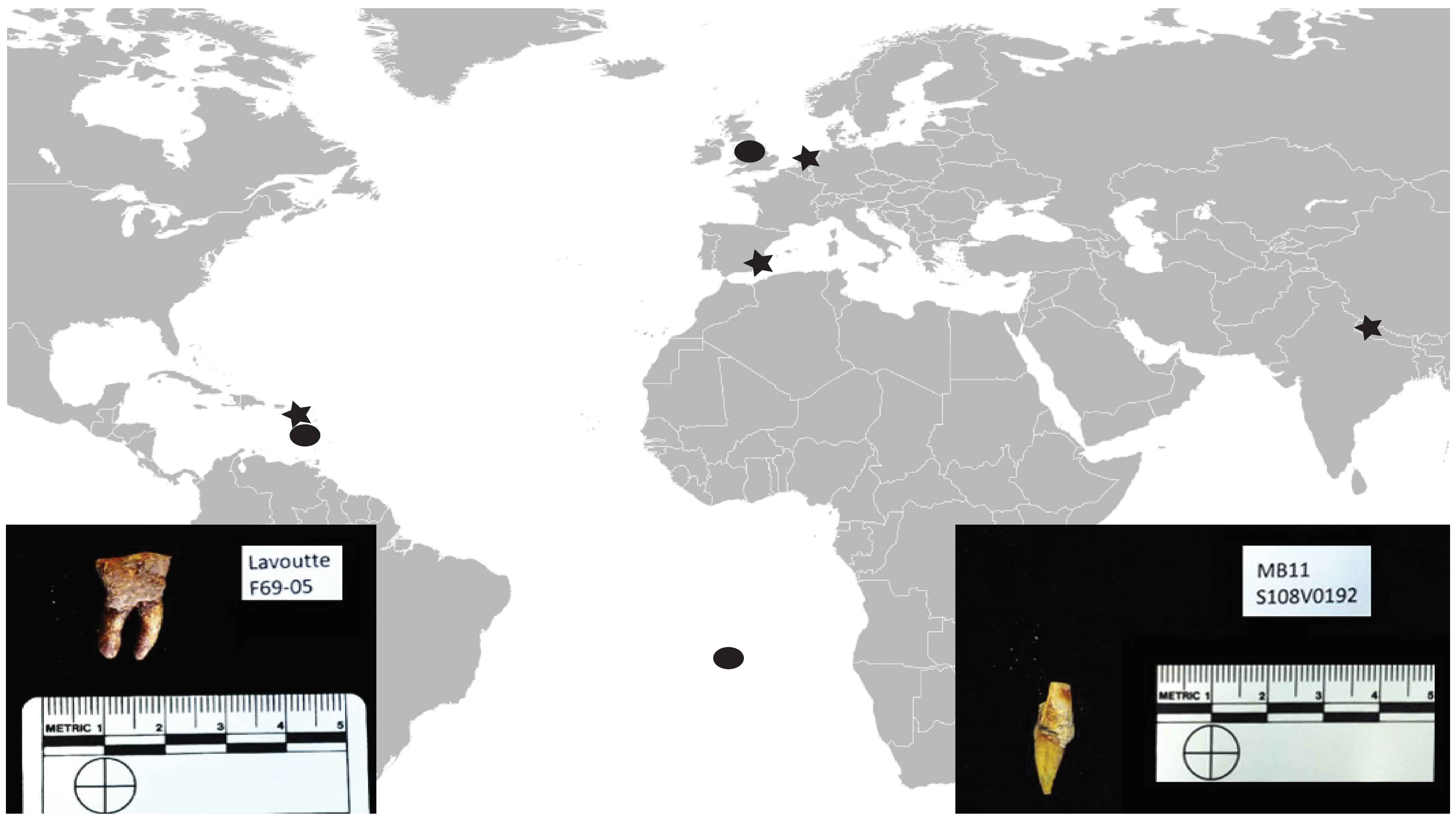

FIGURE 1: Map of sites used for targeted (●) and both targeted and non-targeted (★) sequencing.

The problem:

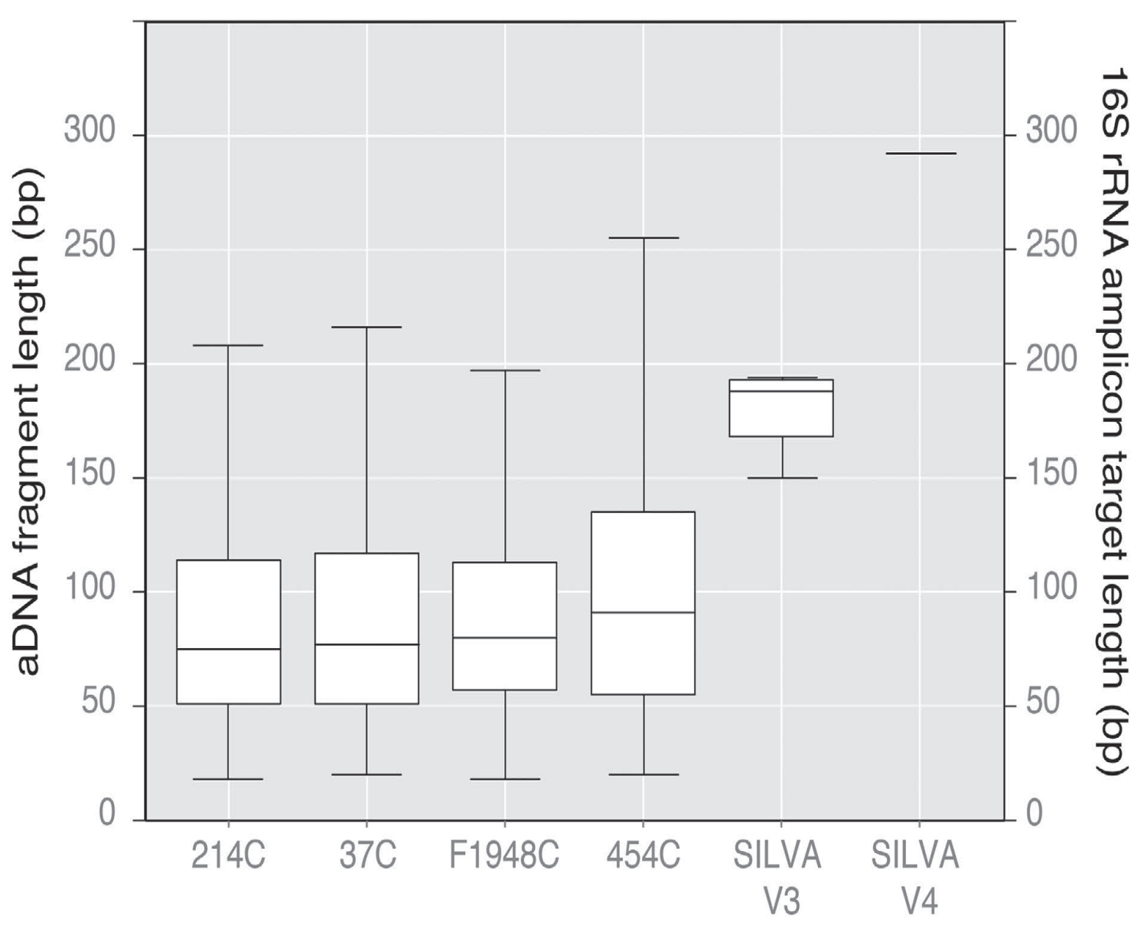

FIGURE 2: Length distribution of aDNA with calculated V3 and V4 16S rRNA amplicon lengths

- Samples predominated by oral-associated genera (Figure 3a), but! Most have low oral microbiome contribution (Figure 3b).
- Unexpectedly high *Euryarchaeota* (in red) (Figure 3c).
- Possible causes: contamination, postmortem bacterial growth, or DNA fragmentation patterns ➡ amplification bias.
- Ancient DNA (aDNA) is highly fragmented (median <100bp) (Figure 2).
- V3 & V4 regions longer than median aDNA length (Figure 2).

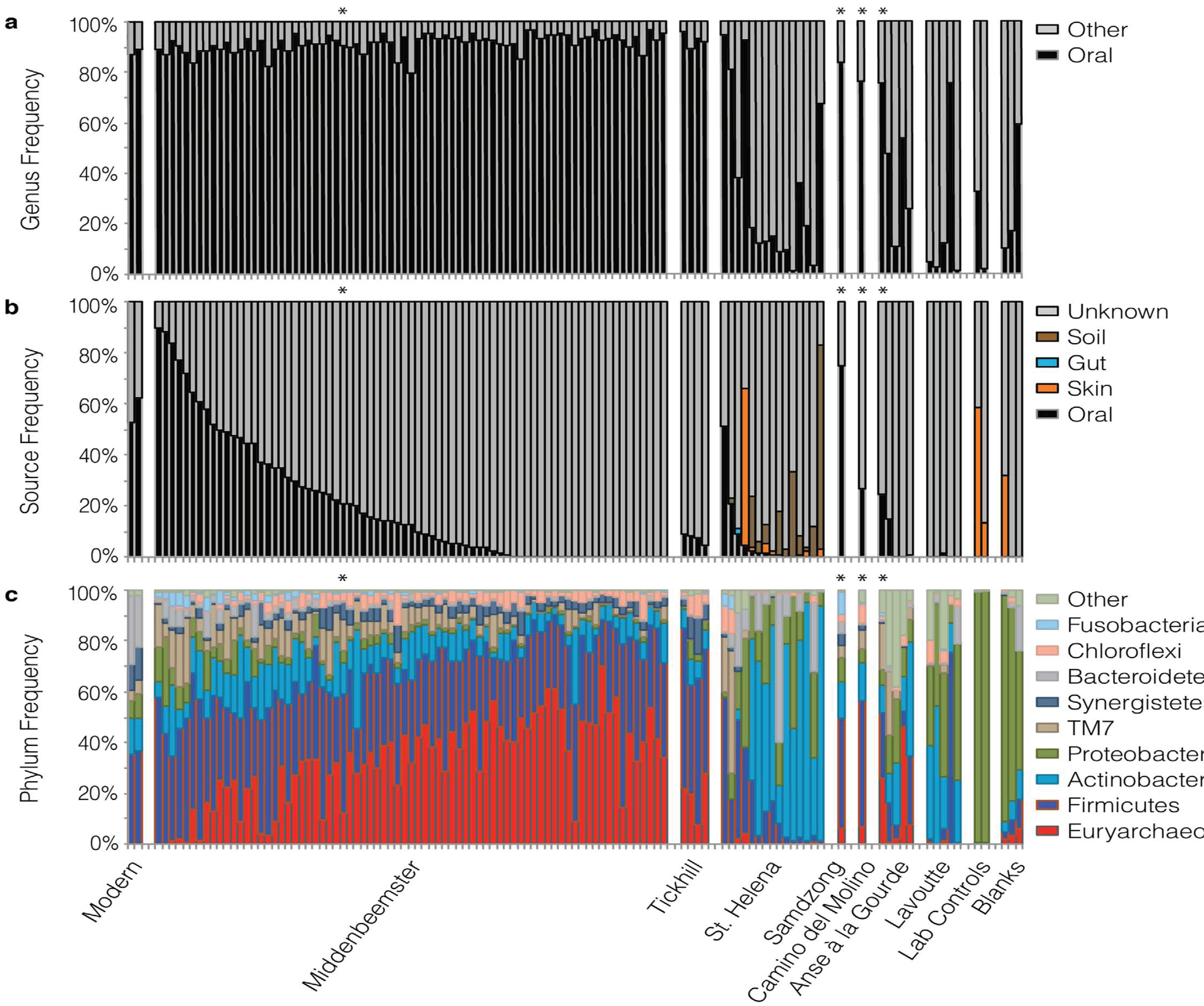

FIGURE 3: (a) Frequency of oral-associated genera in dental calculus and control samples; (b) Bayesian source-tracking results[1]; (c) Phylum frequency from V3 amplicon data. Starred samples (*) selected for shotgun sequencing

Methods:

Shotgun data was generated for a subset of the dental calculus samples to compare to paired amplicon data. Reads clustered into operational taxonomic units (OTUs) at 97% sequence similarity using QIIME[2] pipeline and Greengenes 13.8[3] reference database. Primer amplicons for variable regions of the 16S rRNA gene generated using PrimerProspector[4] and the SILVA 111[5] database, further assessed for important aDNA metrics: length variation; taxonomic resolution; and taxonomic coverage. Results were compared to amplicon and shotgun sequencing results to predict which approach would produce the lowest community reconstruction bias in aDNA studies.

Results and Discussion:

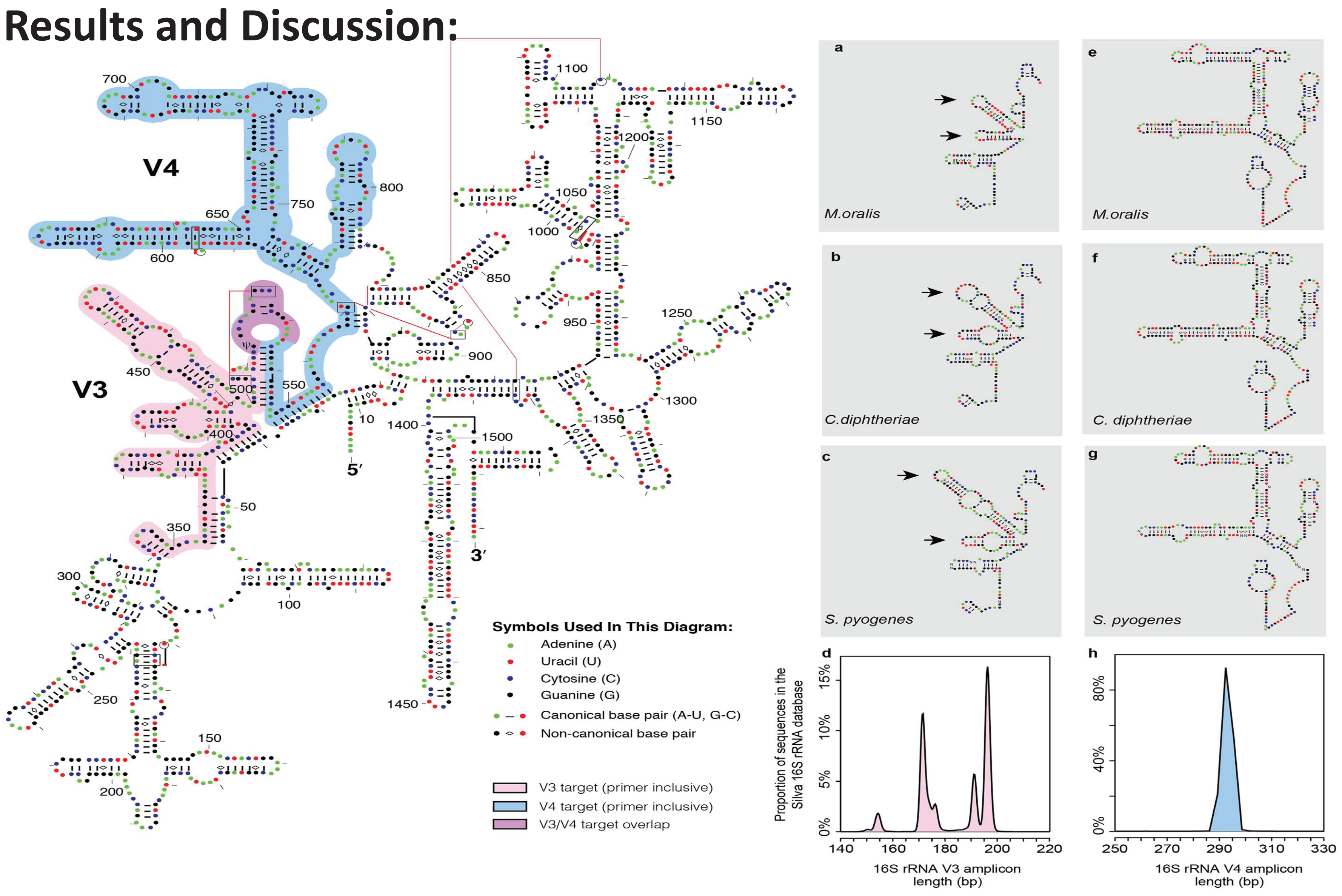

FIGURE 4: Simplified secondary structure of the E. coli 16S rRNA gene. (a-c) High sequence variation in V3 region of three archaeal and bacterial taxa. (e-g) Low sequence variation of V4 region of three archaeal and bacterial taxa. (d) Length distribution of V3 region in archaeal and bacterial taxa. (h) Length distribution of V4 region in archaeal and bacterial taxa.

While the V4 region is relatively size invariant (Figure 4h), it is impractical for aDNA research due to its length (Figure 2). The V3 region is shorter, but comparative analysis shows that its structure varies extensively across taxa (Figure 4a-c arrows), with predicted V3 amplicon lengths ranging from 150 to 194bp (Figure 4d). *In silico* analyses of other 16S rRNA variable regions exhibit similar limitations for studies of ancient oral microbial ecosystems (Table 1).

	Length Statistics[a]			Taxonomic Resolution[b]	Amplicon Taxonomic Coverage By Phylum[c]													
Hypervariable region	Min. amplicon length	Max. amplicon length	Max.-min. length	OTUs	Firmicutes	Bacteroidetes	Proteobacteria	Actinobacteria	Spirochaetes	Fusobacteria	TM7	Tenericutes	Synergistetes	SR1	Chlorobi	Chloroflexi	Euryarchaeota	Chlamydiae
V1	140	276	136	773	0.42	0.26	0.39	0.30	0.29	0.04	0.42	0.32	0.25	0.13	0.31	0.31	-	0.03
V1	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V1-V2	314	379	65	41,781	0.40	0.38	0.38	0.32	0.30	0.31	0.41	0.30	0.44	0.39	0.53	0.22	-	0.03
V1-V2	315	380	65	44,144	0.41	0.38	0.38	0.39	0.30	0.31	0.42	0.32	0.44	0.42	0.55	0.32	-	0.04
V3	216	233	17	18,051	0.14	0.38	0.72	0.55	0.10	0.01	0.03	0.02	0.30	-	0.32	0.01	-	0.00
V3	190	216	26	29,731	0.65	0.83	0.85	0.80	0.32	0.04	0.83	0.17	0.74	0.20	0.96	0.23	-	0.00
V3	174	200	26	46,118	0.99	0.99	0.99	0.96	0.31	0.98	0.96	0.98	0.99	0.98	0.98	0.70	-	0.71
V3	148	192	44	49,397	0.99	0.99	0.99	0.99	0.98	0.99	0.97	0.99	0.99	1.00	0.98	0.98	0.91	0.71
V4	290	295	5	56,463	0.98	0.98	0.98	0.90	0.88	0.98	0.95	0.99	0.99	1.00	0.98	0.96	0.97	0.97
V5	144	148	4	27,009	0.97	0.97	0.98	0.98	0.78	0.95	0.08	0.87	0.94	0.08	0.98	0.37	0.00	1.00
V5	141	146	5	32,063	0.98	0.98	0.98	0.98	0.99	0.98	0.96	0.98	0.99	0.08	0.99	0.87	0.96	1.00
V6	152	167	15	50,892	0.99	0.99	0.99	0.99	0.99	0.99	0.80	0.98	0.99	0.01	0.99	0.96	0.18	1.00
V6	160	172	12	47 332	0.98	0.97	0.98	0.99	0.90	0.96	0.96	0.59	0.99	0.01	0.99	0.98	-	0.99

TABLE 1: (a) Length statistics (b) Number of unique OTUs generated per primer pair (c) Estimate of proportion of sequences in SILVA SSU 111 database that the primer pair will amplify per phylum. Darker cells indicate lower taxonomic coverage.

Shotgun data were next compared to paired amplicon data. Taxa with short V3 amplicon lengths are overrepresented (Figure 5). Conversely, taxa with long V3 amplicon lengths are extremely underrepresented (Figure 5). This in part explains the unusual taxonomic frequencies detected as short V3 amplicon taxa, including members of *Euryarchaeota,* may be observed approximately 20x higher in the amplicon data than the shotgun, while those taxa with long V3 regions such as *Neisseria* may be more than 80x under reported.

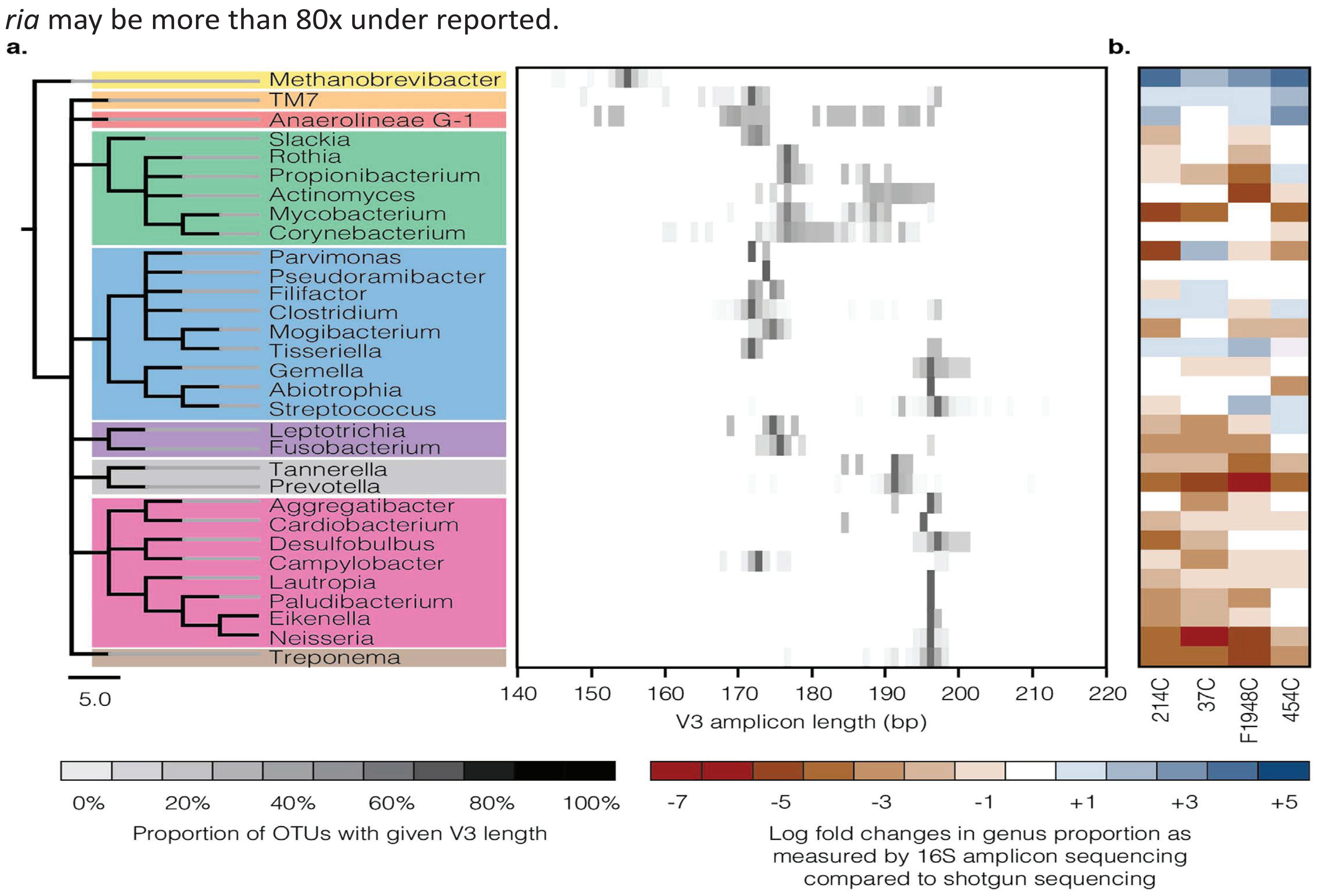

FIGURE 5: Phylogenetically organized 16S rRNA V3 amplicon lengths (a) compared to fold change differences in amplicon versus shotgun datasets (b)

Conclusions:

These results demonstrate the limitations of targeted V3 amplicon sequencing techniques in the characterization of ancient human-associated microbial communities. Analysis of other commonly targeted variable regions of the 16S rRNA gene show that no single variable region can overcome these challenges because of length restrictions, insufficient taxonomic coverage or resolution, or a combination thereof. In conclusion, given these issues, shotgun metagenomics presents an alternative for ancient microbiome characterization. Unlike targeted amplicon-based methods, shotgun sequencing is not impacted by small fragment lengths or target length variation, making it a superior reconstruction method.

References cited:
1. Knights et al. (2011) *Bayesian community-wide culture-independent microbial source tracking*. Nature Methods 8(9):761-763
2. Caporaso et al. (2010) *QIIME allows analysis of high-throughput community sequencing data*. Nature Methods 7(5):335-336
3. DeSantis et al. (2006) *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB*. App Environ Microbiol 73:5069-72
4. Walters et al. (2011) *PrimerProspector: de novo design and taxonomic analysis of PCR primers*. Bioinformatics 27(8):1159-1161
5. Quast et al. (2012) *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools*. Nucleic Acids Research 41:D590-D596