

Detecting Bacterial Species from Ancient Human Skeletal Samples

Ariel Owens, Daisy McGrath and Tsai-Tien Tseng

This paleopathological study aims to identify *Mycobacterium tuberculosis* complex (MTBC), *Mycobacterium avium* complex (MAC) and other *Mycobacterium* species *in silico* from skeletal samples that belonged to 28 Polish individuals in the Neolithic period under PRJNA422903 from the Sequence Read Archive (SRA). After next-generation sequencing (NGS), bioinformatics methods are heavily relied upon for identification of pathogens from complex samples. We implemented a bioinformatics pipeline, with custom-built databases, utilizing the following software tools: Trim Galore! and Kraken2. After adapter trimming, Kraken2 was used for taxonomic classifications. We have found that *Mycobacterium* is present in all 28 individuals. The average percentage of MAC present in the genus *Mycobacterium*, in all 28 individuals, is 6%. Reads from MTBC makes up an average of 7% of the *Mycobacterium* genus. We have identified previously unreported strains of MTBC and MAC such as *Mycobacterium tuberculosis* XDR1219, which is an extensively drug-resistance strain. Our analysis also revealed 14.8% of reads from MTBC belong to *Mycobacterium avium hominissuis*, which was commonly found in humans and pigs. Additionally, strains of *Mycobacterium simiae* complex were also discovered. *M. simiae* has been commonly found among immunocompromised individuals. In conclusion, our bioinformatics pipeline has been more effective than other published approaches. This approach broadens the potential scope of paleoepidemiology both to older, sub-optimally preserved samples and to pathogens with difficult intragenetic taxonomy. It is therefore suitable for other studies in paleopathology using NGS technologies.