The Little Book of Smiley Plots

A collection of ancient DNA patterns and their causes

The SPAAM Community

2025-01-05

Table of contents

Preface

A key part of any ancient DNA project is to show that the DNA is exactly that - that the DNA is ancient, rather than from modern contamination.

A key authentication method is to show the presence of elevated C to T deamination patterns (and the complementary G to A) at the end of DNA molecules - known as damage patterns - originally reported by (Briggs et al. 2007).

These patterns can be plotted in what have been colloquially known as 'Smiley Plots. However, there can be a wide range of smiley plots, some which show valid ancient DNA, and others that do not - either due to not actually having true ancient DNA but also from laboratory and/or bioinformatic artifacts.

This book aims to act as a reference guide to interpreting ancient DNA damage plots, providing a wide range of example 'smiley plots', with descriptions of what the describe and what can cause them. As an added bit of fun, each type of 'smiley plot' comes with a artistic interpretation of the line shape contributed by members of the ancient DNA community.

Introduction

What are damage patterns?

Damage patterns on ancient DNA molecules occur due to increased miscoding lesions at the end of molecules. When DNA molecules start to decompose (i.e., repair mechanisms are lost once an organism dies), the very long DNA molecules start to fragment due to 'nicks' occurring on the sugar-phosphate of one of the strands, weakening the structure and causing the molecule to cleave into two. However, this cleavage is not necessarily 'clean', i.e., occurs on both strands at the same position. Rather, when the two uneven 'nicks' cause the DNA molecule to cleave into two, this results in a 'jagged' break - with the molecules having 'overhangs' of one strand being longer than the other of each of the new two now-'independent' molecules.

The resulting single-stranded overhangs leave the nitrogenous-bases 'exposed' on the overhang to the surrounding environment. It was found

It is important to note that the library construction method will influence damage, e.g., is the library constructed from double-stranded DNA or single-stranded DNA, is the polymerase in the initial library amplification proof-reading or not, and so on.
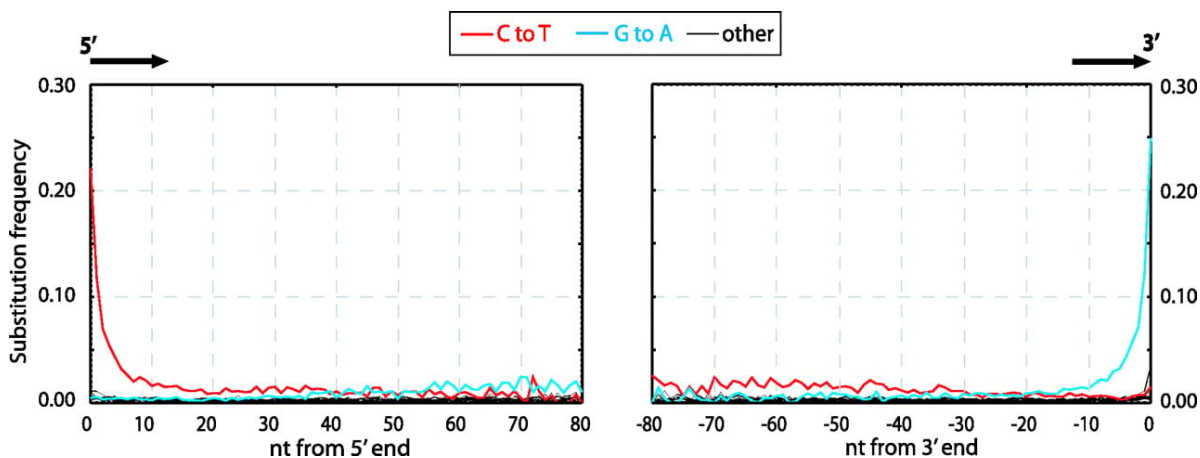
Figure **??**

Figure 1: First reported misincorporation lesion 'smiley plot' from Neanderthal DNA (Briggs et al. 2007). Reproduced here under free access.

4

How to generate damage patterns

There is a range of software that can generate damage pattern plots from ancient DNA NGS libraries. The vast majority of tools require to be of sequencing reads aligned to a reference genome. Here we make suggestions of some tools that you can use to generate such plots.

Genomics

These tools general take BAM files as input (i.e., after mapping of FASTQ files to a reference genome using a short-read aligner)

- mapDamage

    - Source: https://github.com/ginolhac/mapDamagee
    - Documentation: https://ginolhac.github.io/mapDamage
    - Citation: (Jónsson et al. 2013)

- DamageProfiler

    - Source: https://github.com/Integrative-Transcriptomics/DamageProfiler
    - Documentation: https://damageprofiler.readthedocs.io/en/latest/
    - Citation: (Neukamm, Peltzer, and Nieselt 2021)

Metagenomics

- MaltExtract

    - Source: https://github.com/rhuebler/MaltExtract
    - Documentation: https://github.com/rhuebler/MaltExtract
    - Citation: (Hübler et al. 2019)

Part I

Valid Smiley Plots