THE UNIVERSITY of ADELAIDE

BIOINFORMATICS AND SYSTEMS MODELLING

# Ancient DNA

Bastien Llamas

Monday 24 October 2022

BIOINF 3000 / BIOTECH 7005, Week 12

www.adelaide.edu.au/acad

@DNATimeTravel @ACAD_Research_

bastien.llamas@adelaide.edu.au

ACAD
Australian Centre for Ancient DNA

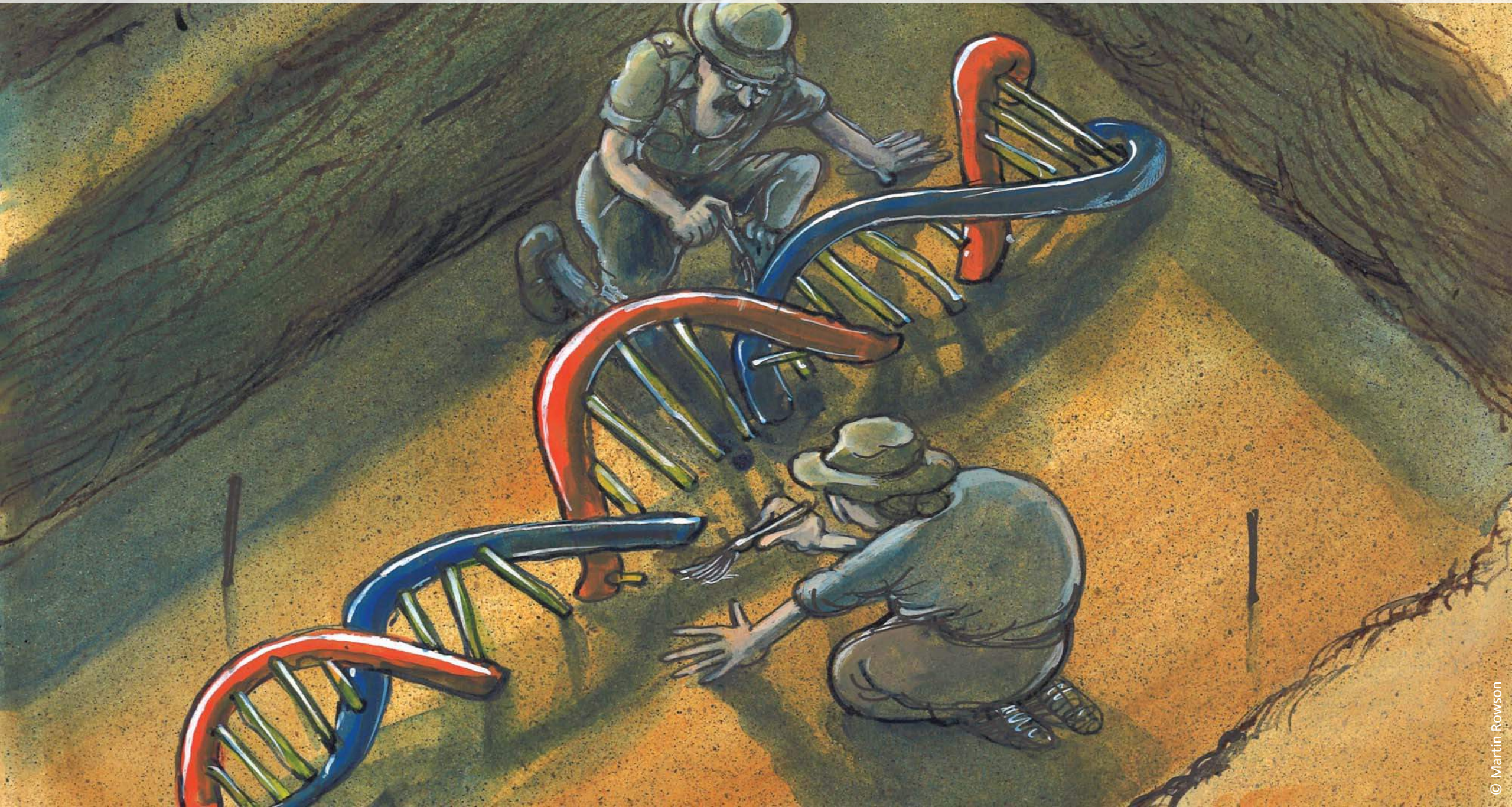
# Lecture overview

- Introduction to ancient DNA
- Sources of ancient DNA and post-mortem DNA decay
- Properties of ancient DNA
- Ancient DNA analysis

# Lecture overview

- Introduction to ancient DNA
- Sources of ancient DNA and post-mortem DNA decay
- Properties of ancient DNA
- Ancient DNA analysis

# Ancient DNA



# Dead men do tell tales

Credit: sbras.info

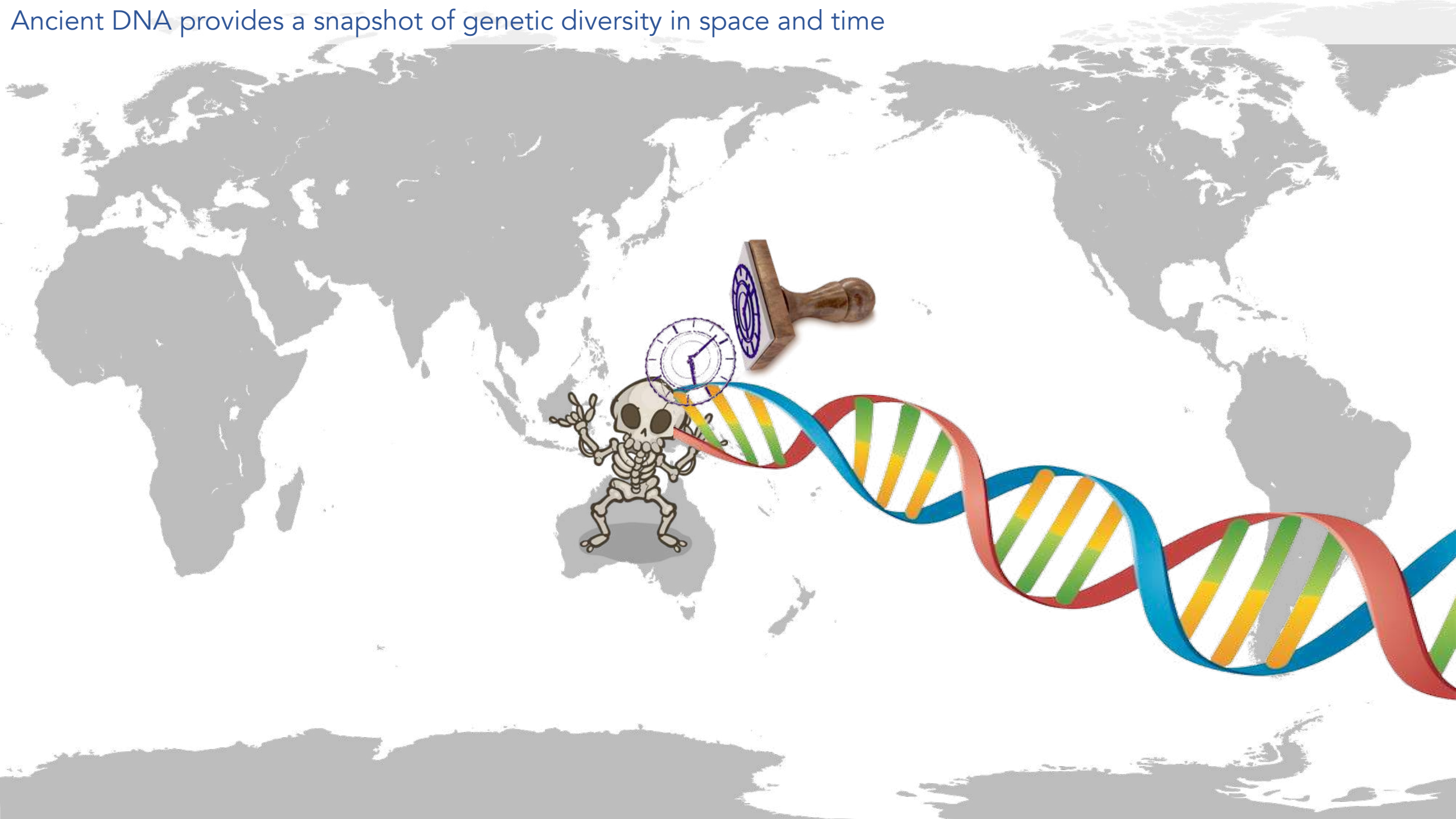
# Extinct humans

Credit: Entressangle & Daynes

# Extinct fauna

National Museum of Antropology, Mexico City

# Ancient DNA provides a snapshot of genetic diversity in space and time



# Lecture overview

- Introduction to ancient DNA
- Sources of ancient DNA and post-mortem DNA decay
- Properties of ancient DNA
- Ancient DNA analysis

# Sources of ancient DNA

BONES 

 DENTAL CALCULUS

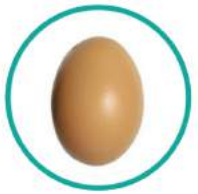
MAIZE COBS 

 MOLLUSC SHELLS

 TEETH

MUSEUM SKINS 

 GRAPE VINE & SEEDS

EGGSHELLS 

BULK-BONE 

 COPROLITES

HERBARIUM 

 PARCHMENT

 MUMMIES

RODENT MIDDEN 

 POLLEN

SPELEOTHEM 

HAIRS 

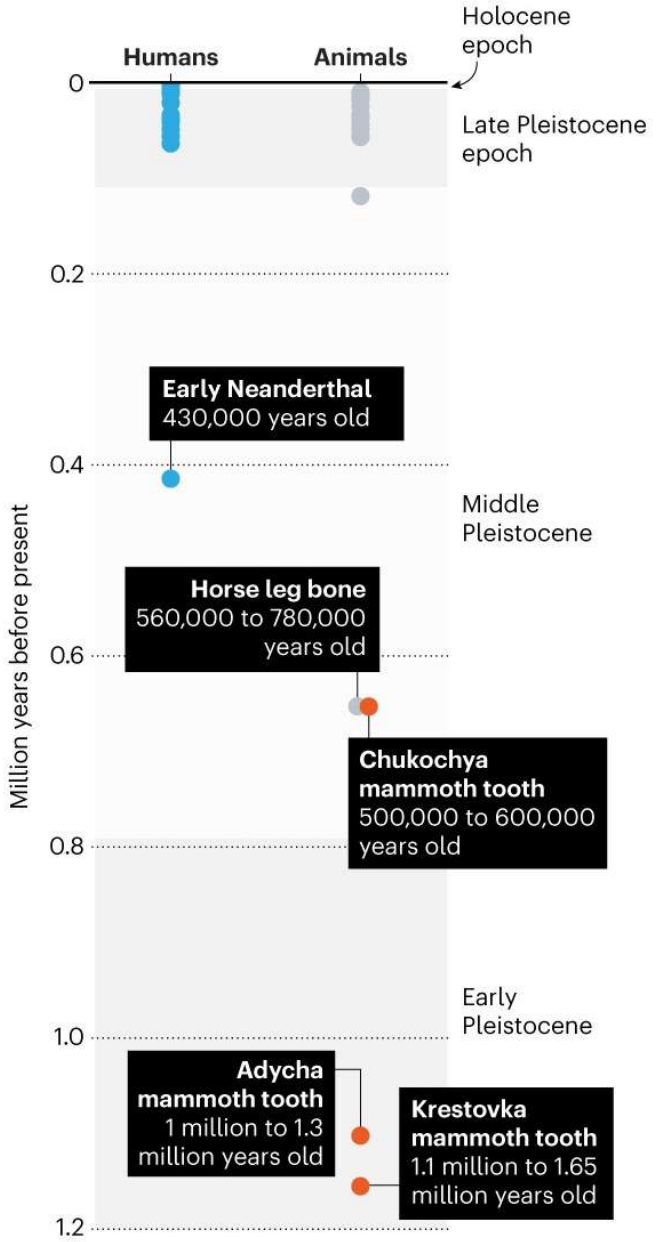
 CLOTHING

WOOD 

 SEDIMENT CORES

Adapted from www.molecularecologist.com/2017/04/unusual-sources-of-ancient-dna

# Empirical time limit of DNA preservation is 1.1 to 1.65 million years




The upper age bound for the mammoth teeth is based on a genetic dating method; the lower bound is based on the age of the sediments in which the teeth were found.

# Theoretical time limit of DNA preservation is 1 million years

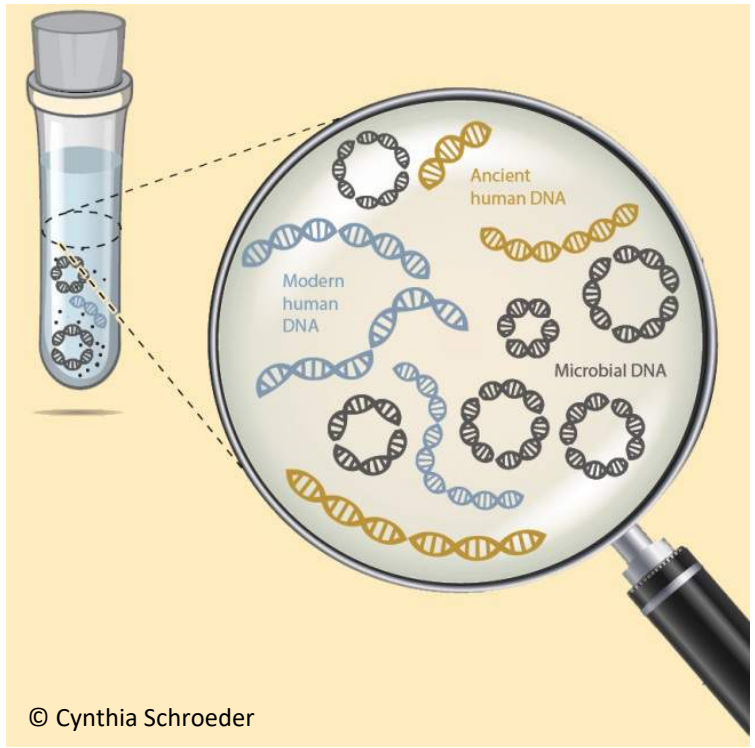

Wikimedia Commons

# Lecture overview

- Introduction to ancient DNA
- Sources of ancient DNA and post-mortem DNA decay
- Properties of ancient DNA
- Ancient DNA analysis

# The three experimental challenges of ancient DNA

## Contamination

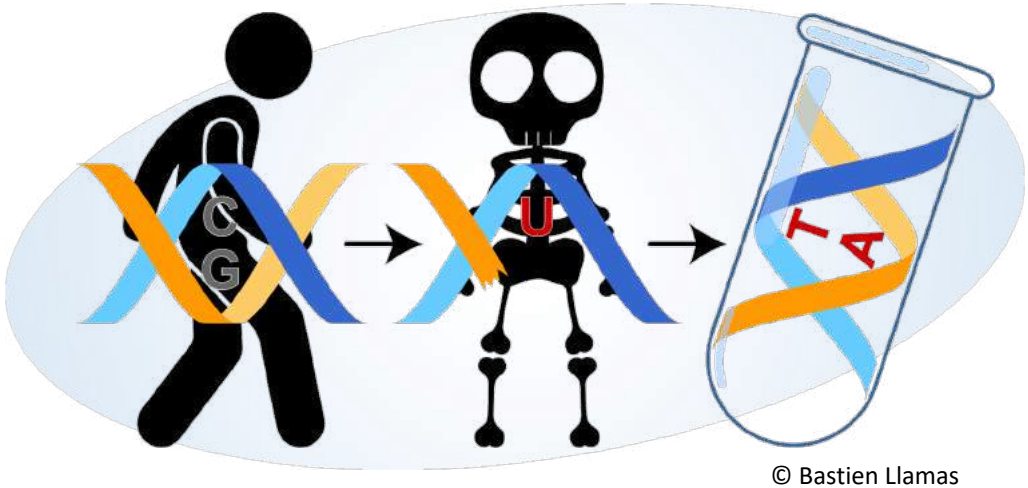



© Cynthia Schroeder

## Fragmentation

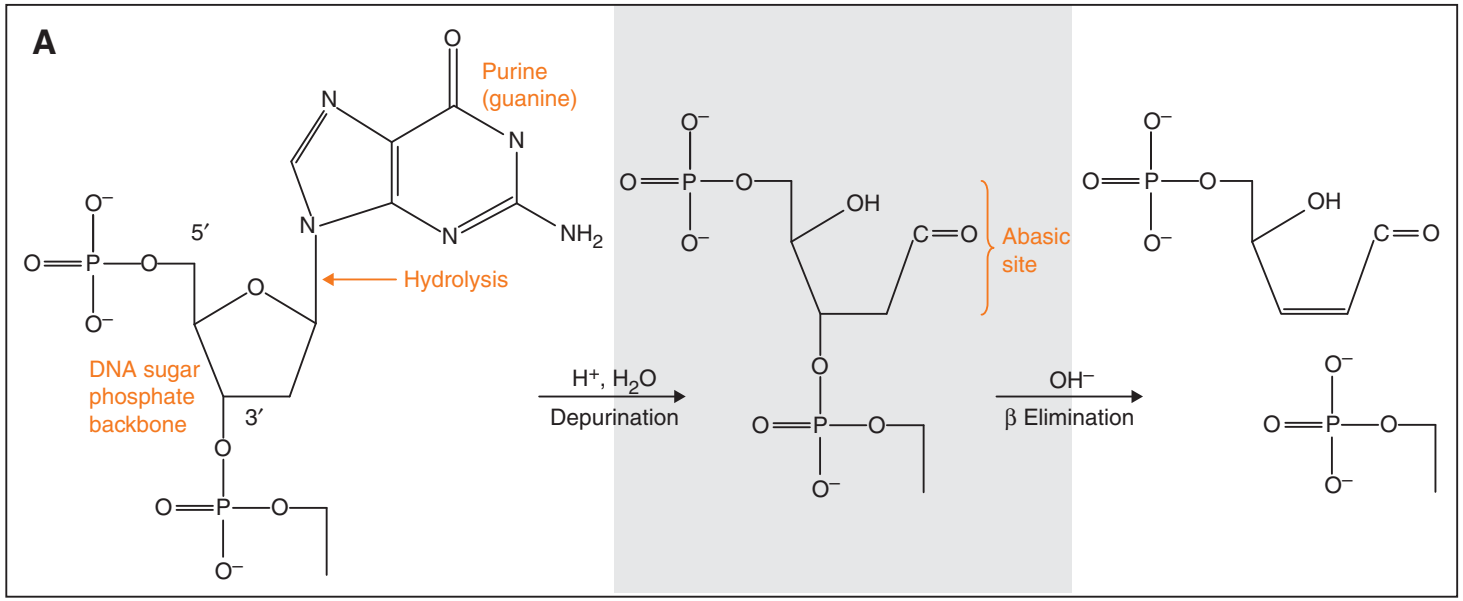
© DrexelNOW

## Miscoding lesions




© Bastien Llamas

# Ancient DNA damage: fragmentation



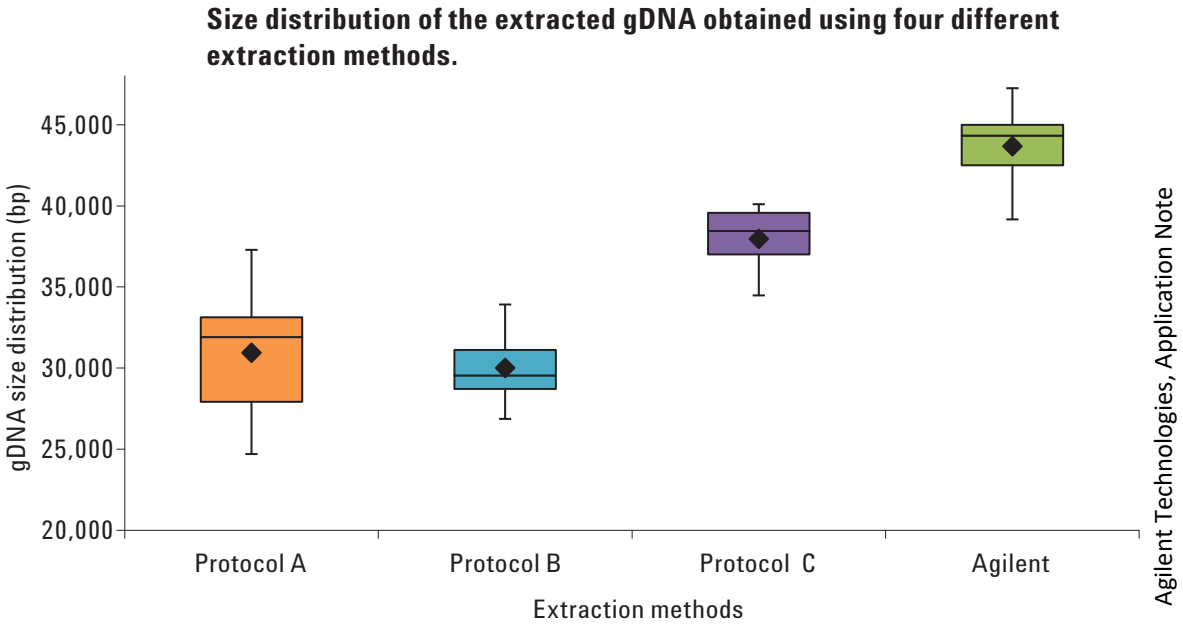

Dabney et al 2013 Cold Spring Harbor perspectives in biology

Steps to DNA fragmentation:

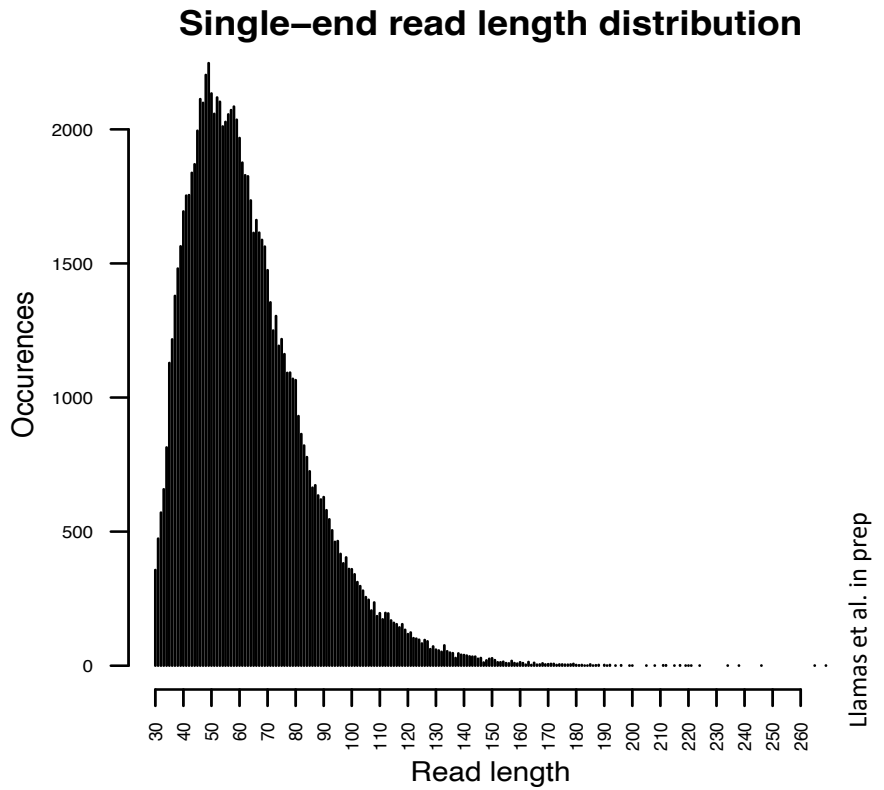
- Hydrolytic depurination (N-glycosyl bond between sugar and adenine or guanine residue is cleaved) results in an abasic site
- DNA strand fragmented through β elimination, leaving 3’-aldehydic and 5’-phosphate ends

# Ancient DNA fragmentation are short

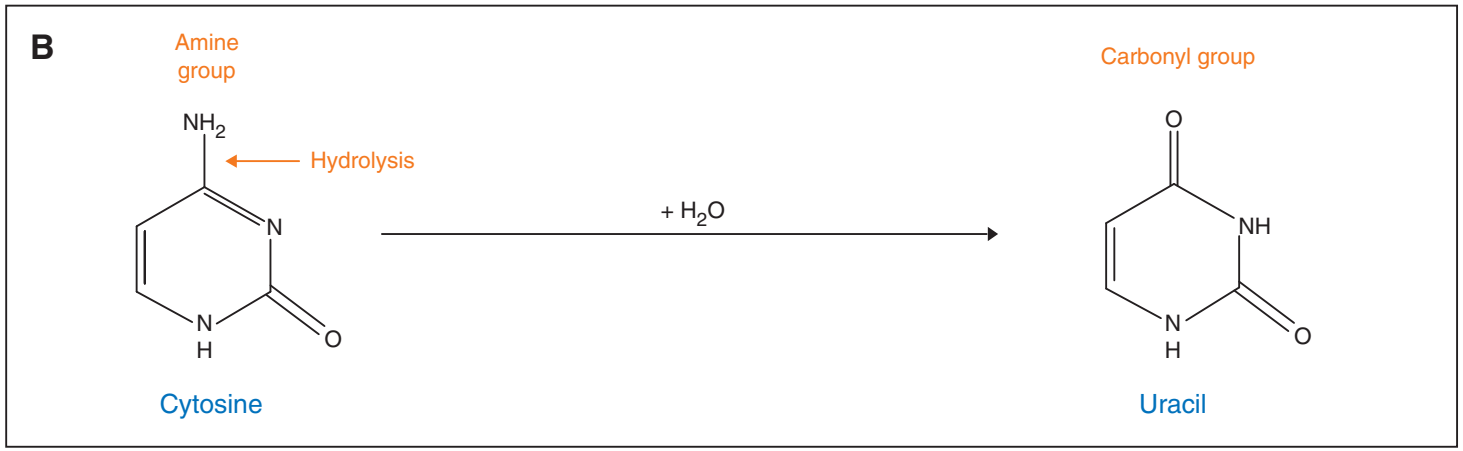
DNA from fresh samples:
Several thousands of base pairs




Ancient DNA:
Few tens of base pairs



# Ancient DNA damage: deamination

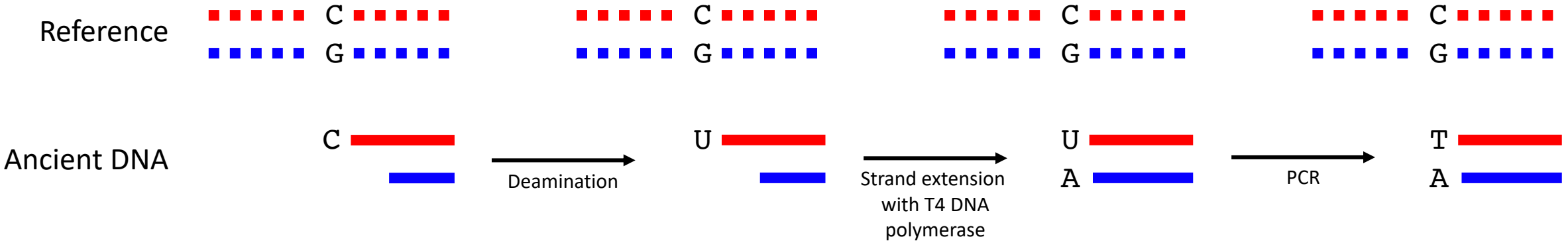


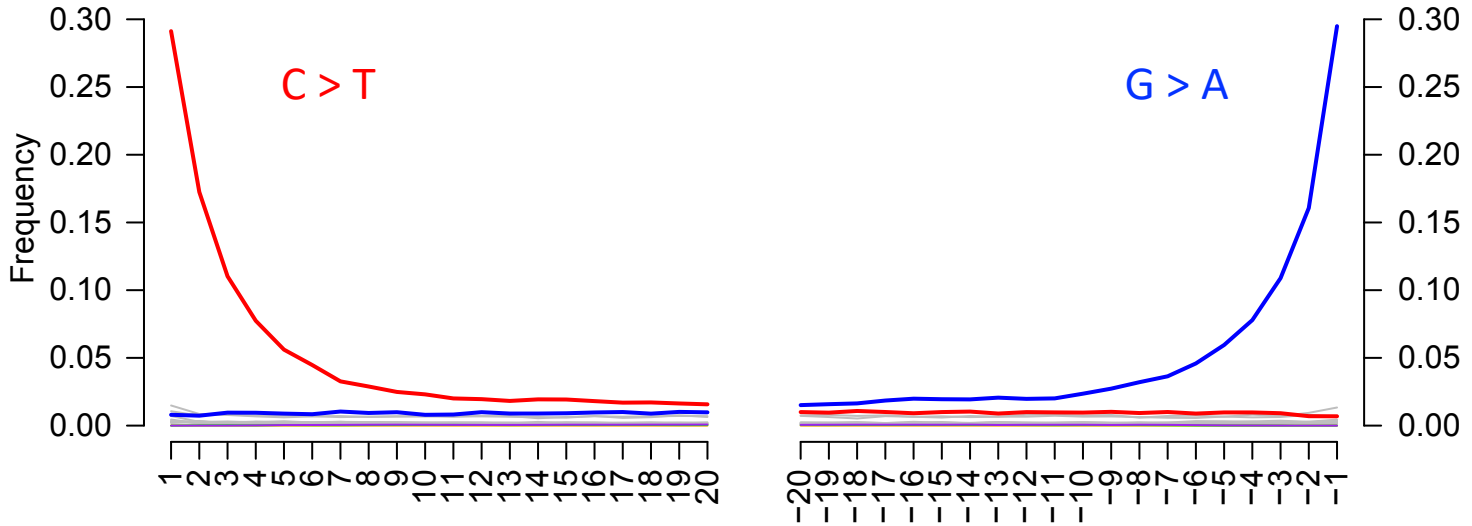

Dabney et al 2013 Cold Spring Harbor perspectives in biology

- Deamination of cytosine to uracil is the major mechanism leading to miscoding lesions in ancient DNA
- If not dealt with experimentally or bioinformatically, miscoding lesions can lead to false positive results

# Ancient DNA deamination accumulates at molecules ends

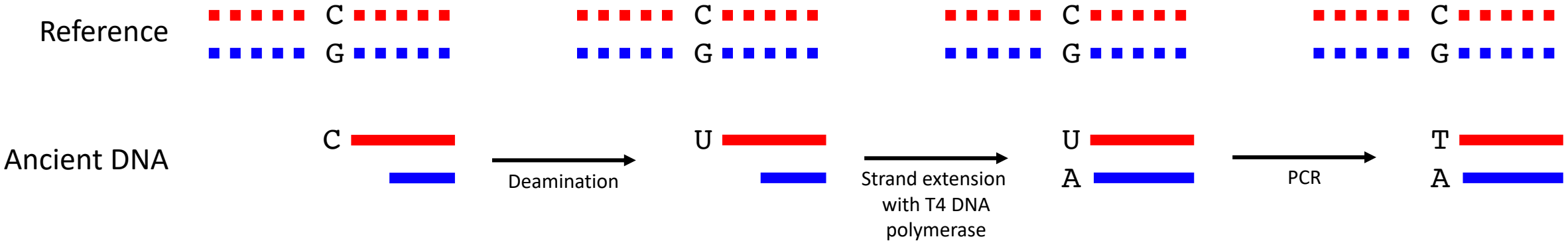
DNA polymerases will incorporate an A across from the U, and in turn a T across from the A, causing apparent G to A and C to T substitutions

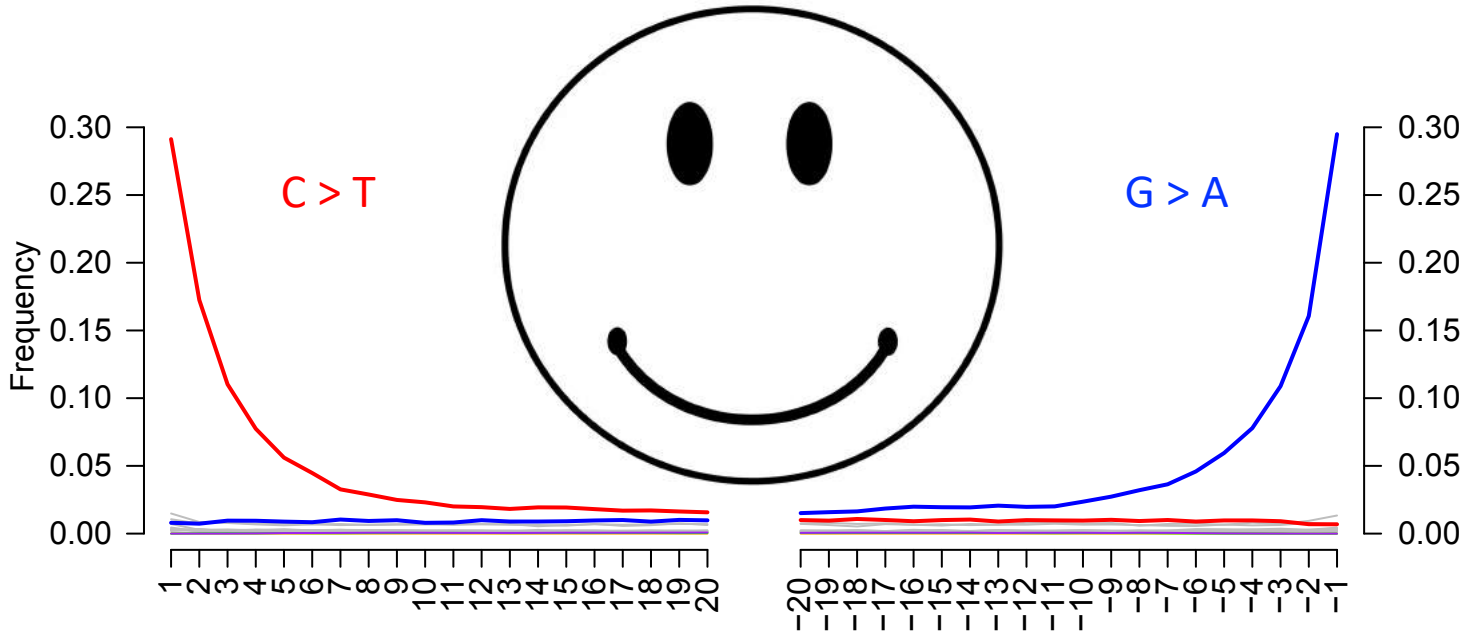


# Ancient DNA deamination accumulates at molecules ends

DNA polymerases will incorporate an A across from the U, and in turn a T across from the A, causing apparent G to A and C to T substitutions

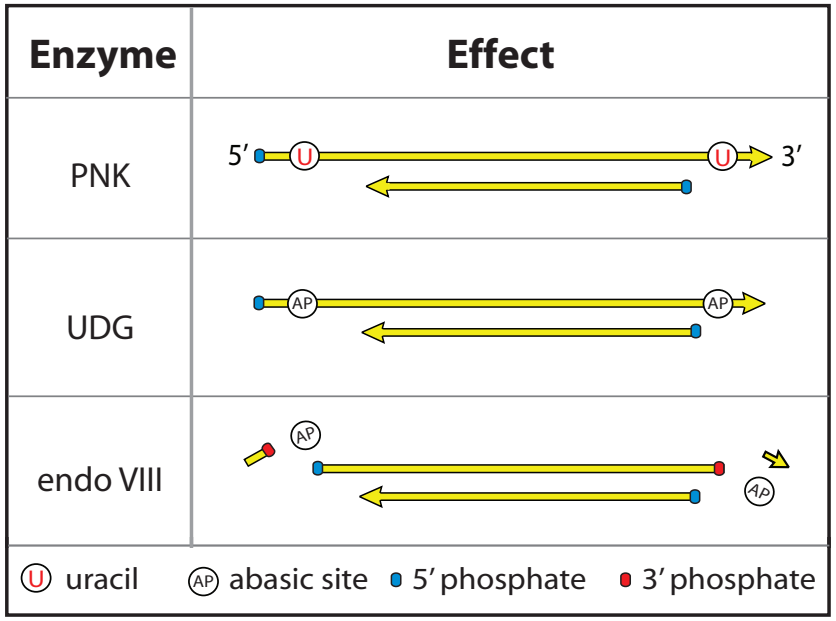


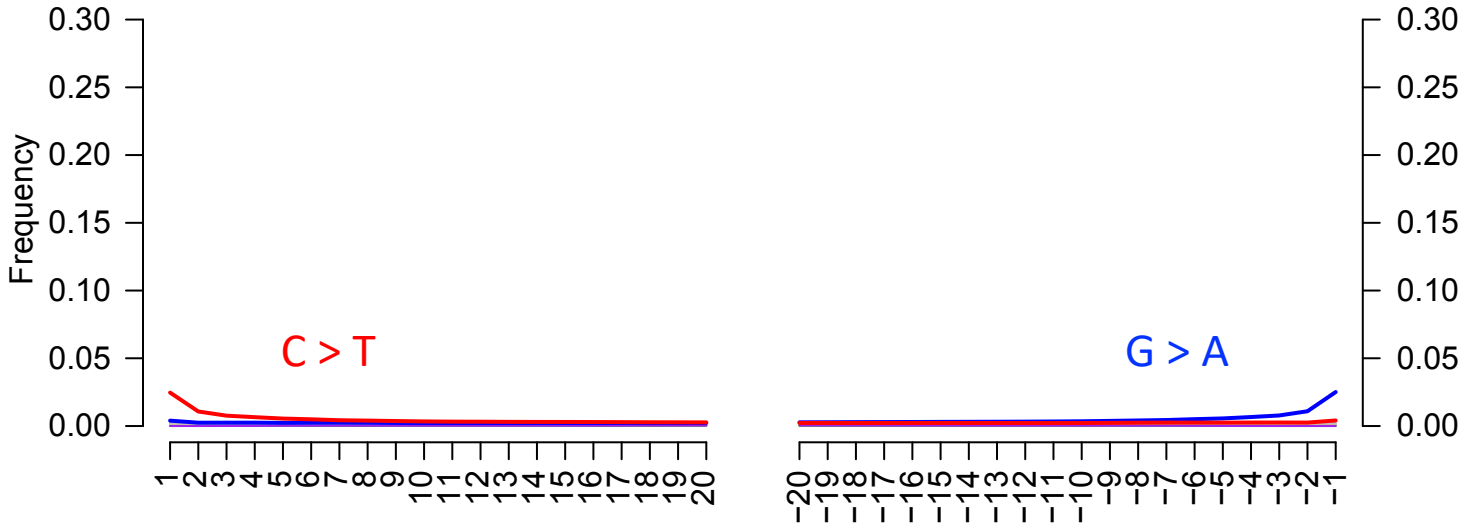
# Uracils can be removed experimentally

DNA can be “repaired” before preparing sequencing libraries by:

- Removing the uracil base with UDG (Uracil-DNA Glycosylase), leaving an abasic site
- Cleaving the DNA backbone at the 3’ and 5’ sides of the abasic site with endonuclease VIII



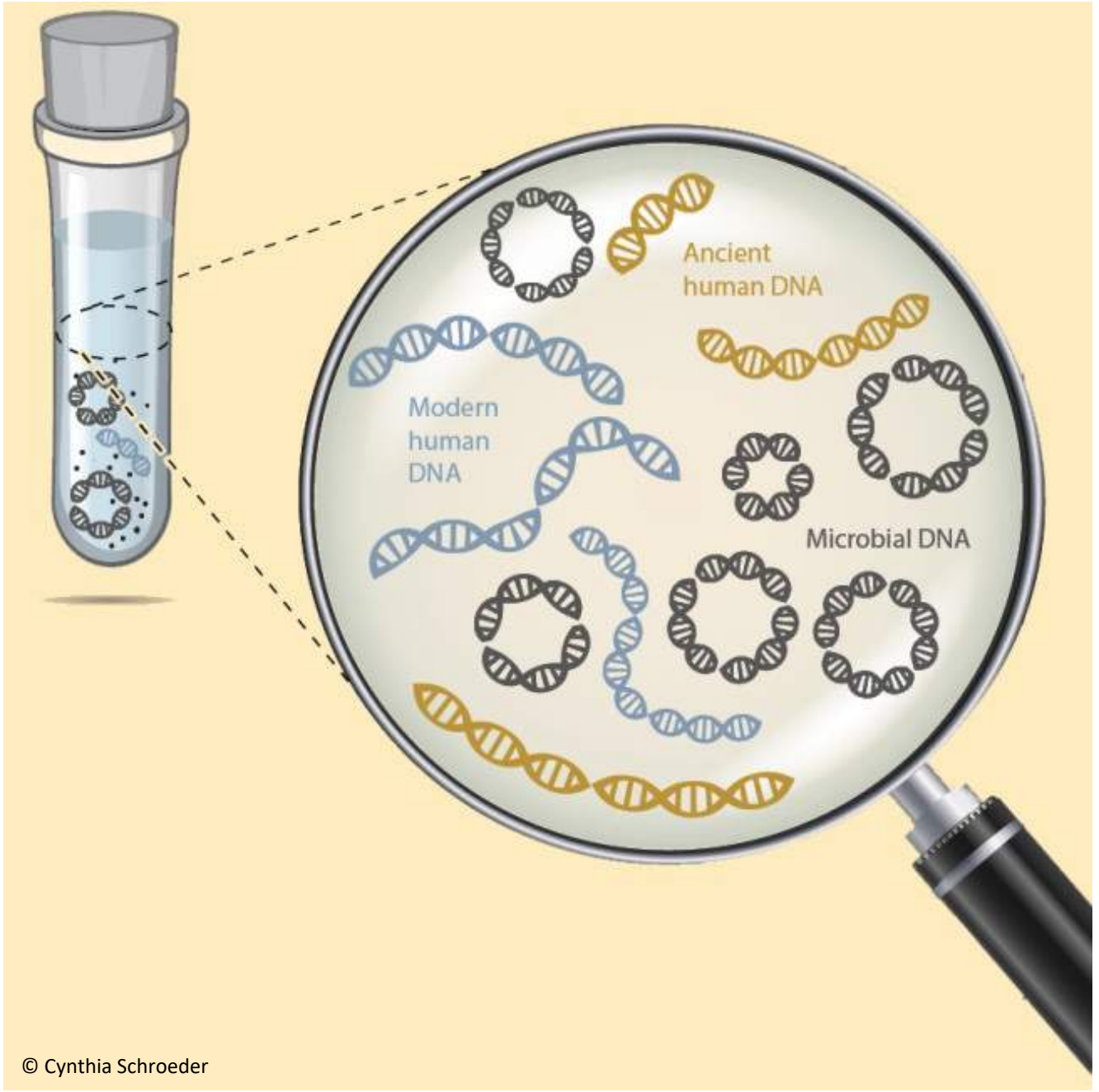

Adapted from Briggs et al. 2010 Nucleic Acids Research




Consequences of DNA repair:

- No accumulation of C-to-T and G-to-A substitutions at the DNA fragment ends
- Shorter DNA fragments

# Ancient DNA is contaminated



# The Australian Centre for Ancient DNA: a low-DNA environment


Google Maps

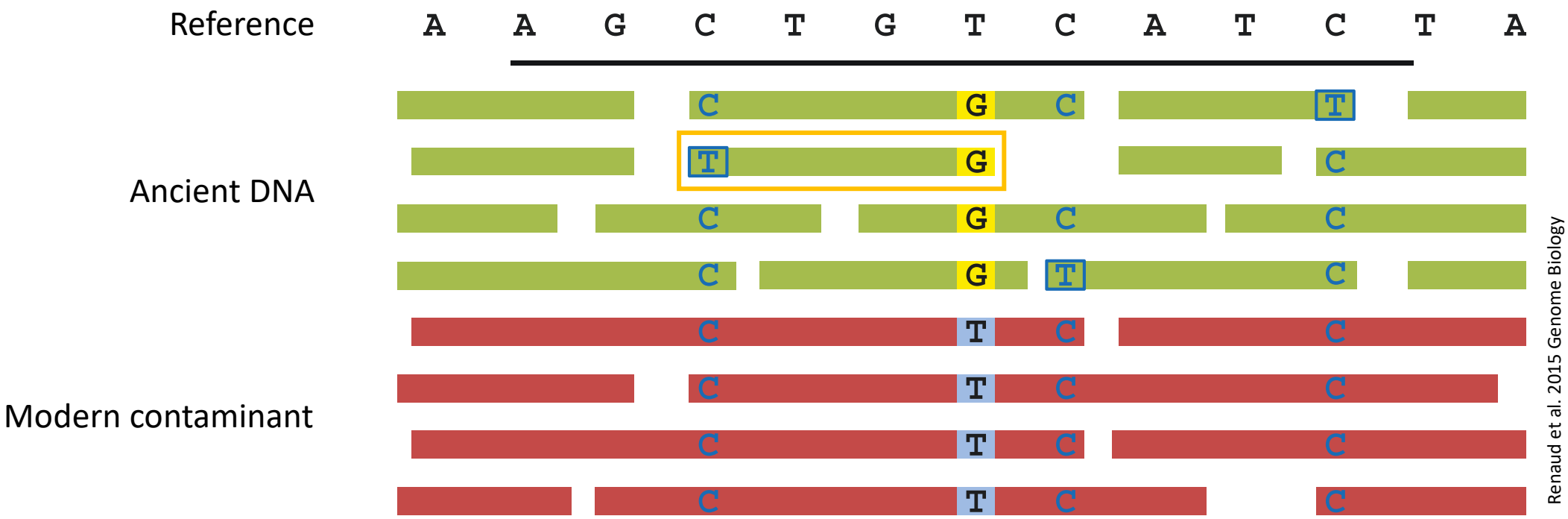
Credit: Jeremy Austin

# Strict laboratory protocols

Credit: Julien Soubrier

Credit: Julien Soubrier

# Contamination estimation from haploid DNA




In data mapped to haploid genomes (mitochondrial DNA in both sexes or X chromosome in males):

1. Identify heteroplasmic positions (here G/T)
2. Identify ancient DNA fragments that carry deaminated cytosines near the ends of the DNA fragments
3. The ancient allele is linked to cytosine deamination in some ancient DNA fragments. Estimate contamination from fragments that do not contain the ancient allele

# So, what is so special about ancient DNA?

- FACT: Ancient DNA is heavily fragmented

  EVIDENCE: Distribution of ancient DNA fragments length is skewed towards short fragments

- FACT: ancient DNA is contaminated with exogenous DNA

  EVIDENCE: Only a small fraction of the raw sequencing data aligns to the reference genome

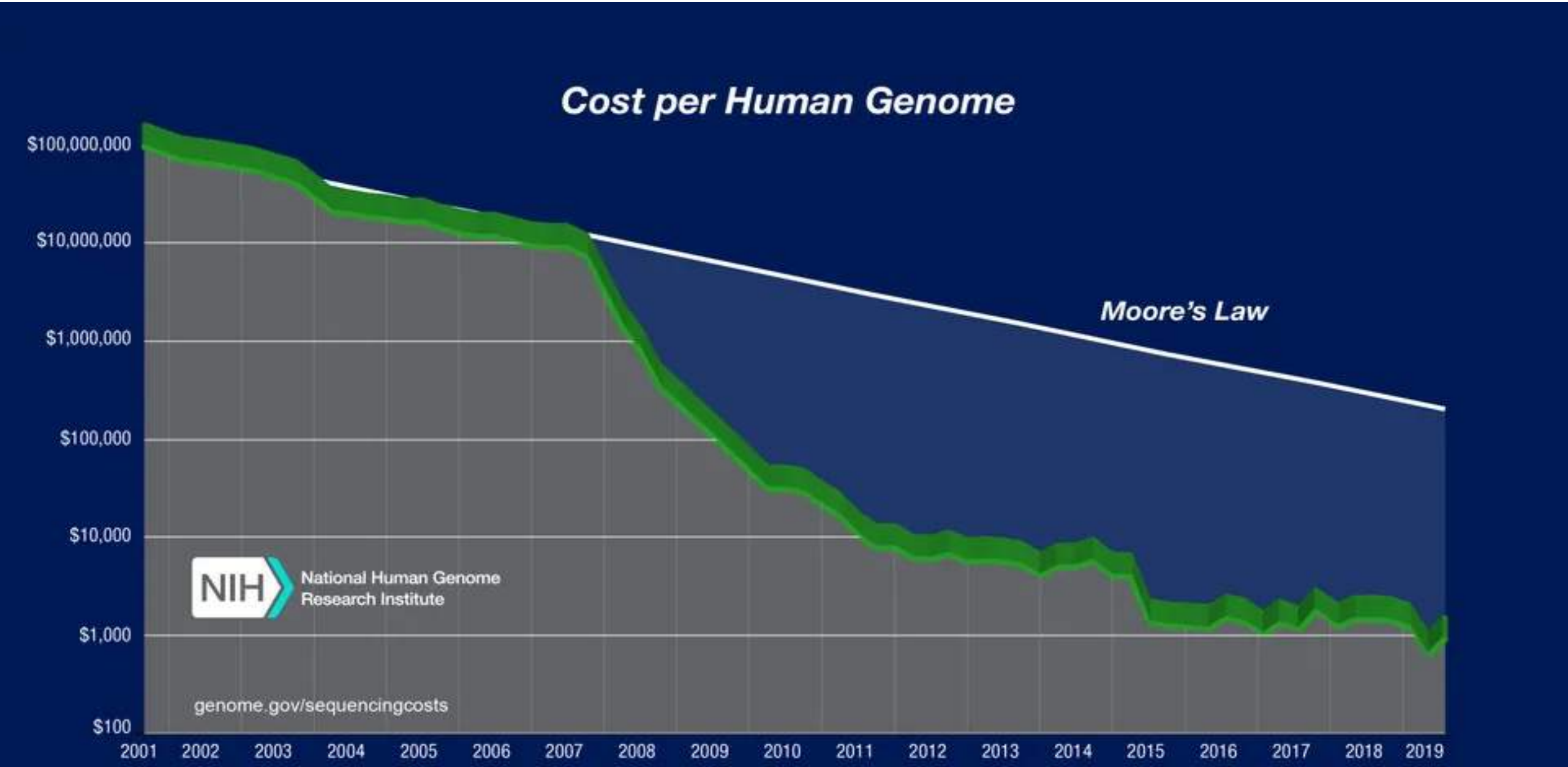
- FACT: Cytosines are deaminated into uracils

  EVIDENCE: Accumulation of C-to-T and complementary A-to-G substitutions at the ends of sequencing reads if DNA is not repaired with UDG and Endo VIII

# Lecture overview

- Introduction to ancient DNA
- Sources of ancient DNA and post-mortem DNA decay
- Properties of ancient DNA
- Ancient DNA analysis

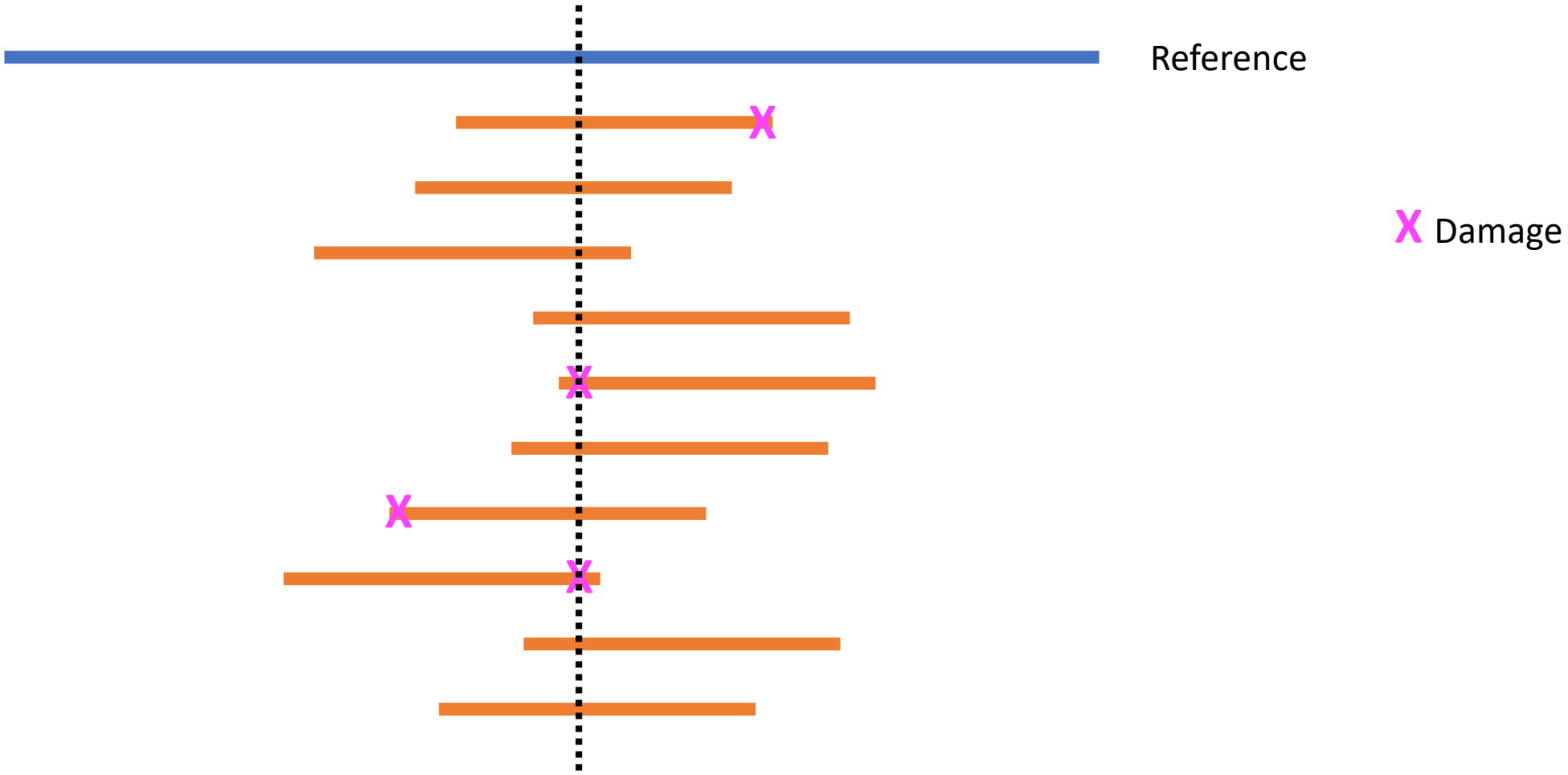
# High-throughput sequencing (HTS)



# Ancient DNA and HTS: the perfect match

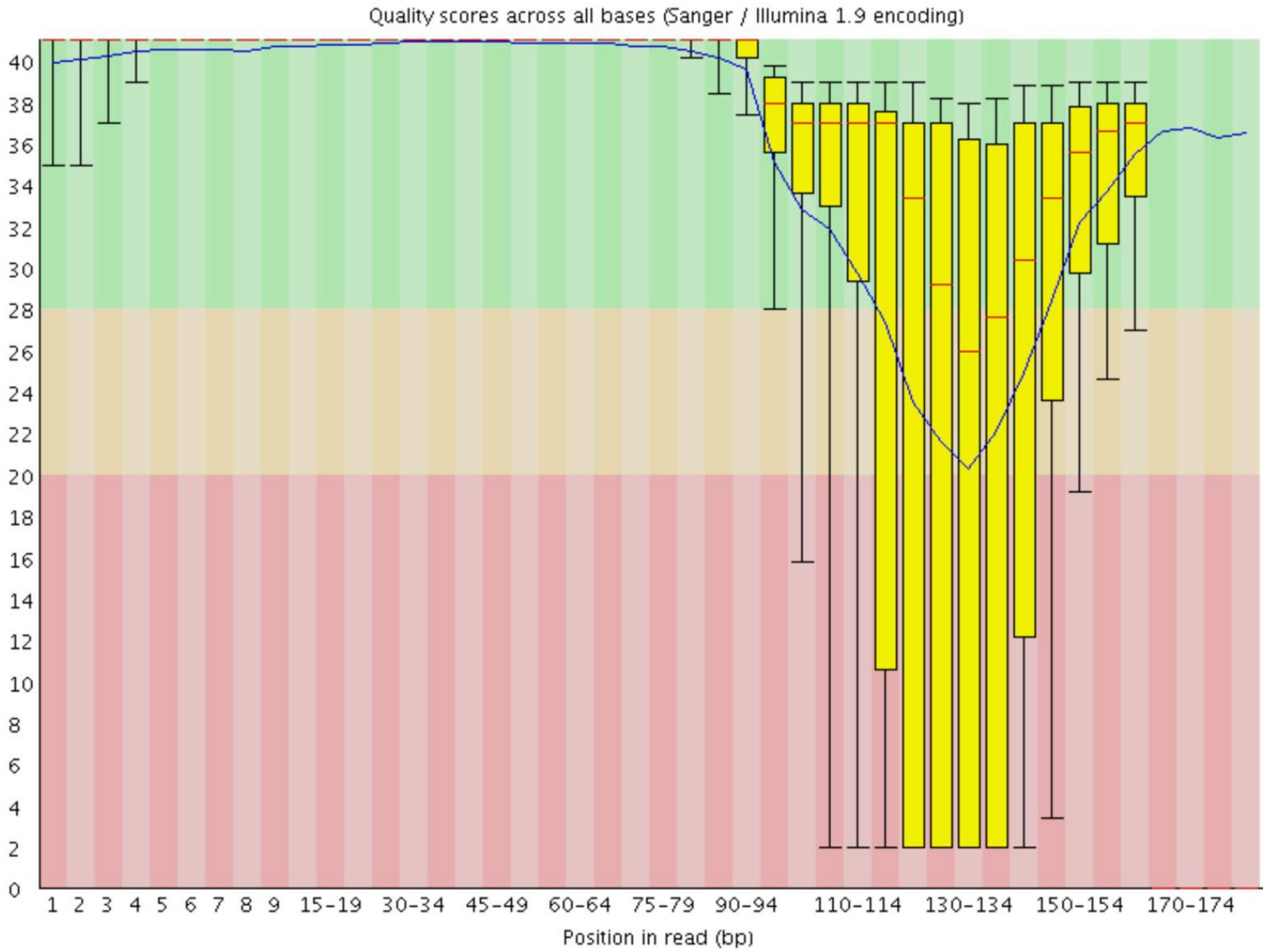
- HTS generate short DNA sequences
- Sequencing of virtually all molecules in an ancient DNA extract
- Independent replication for each genomic site

# Depth of coverage in HTS

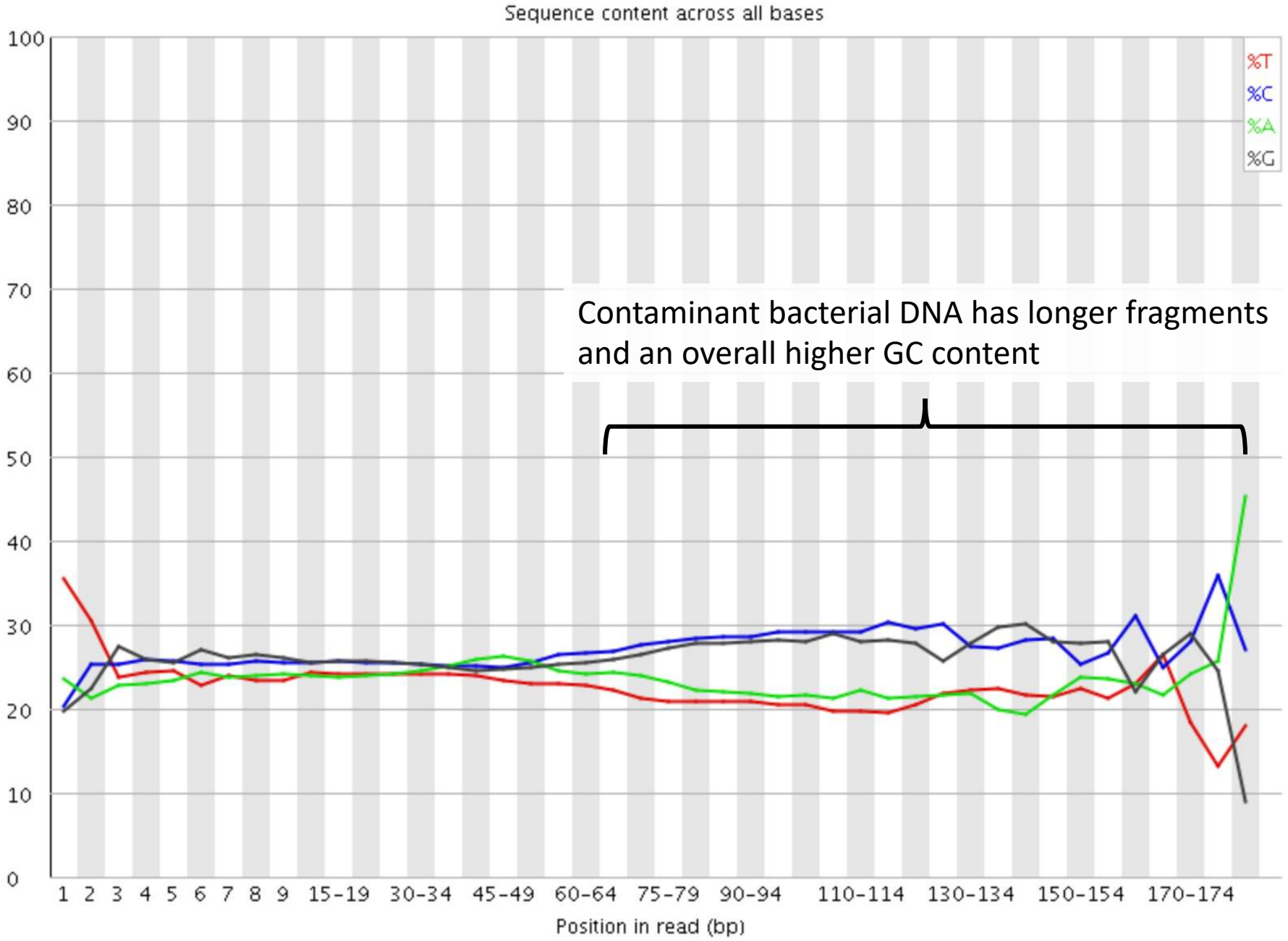



Deep sequencing of ancient DNA allows independent replication of variant sites to identify miscoding lesions and remove false positive results

# Typical base quality distribution of ancient DNA sequencing data



# Typical higher GC content towards the end of sequencing reads

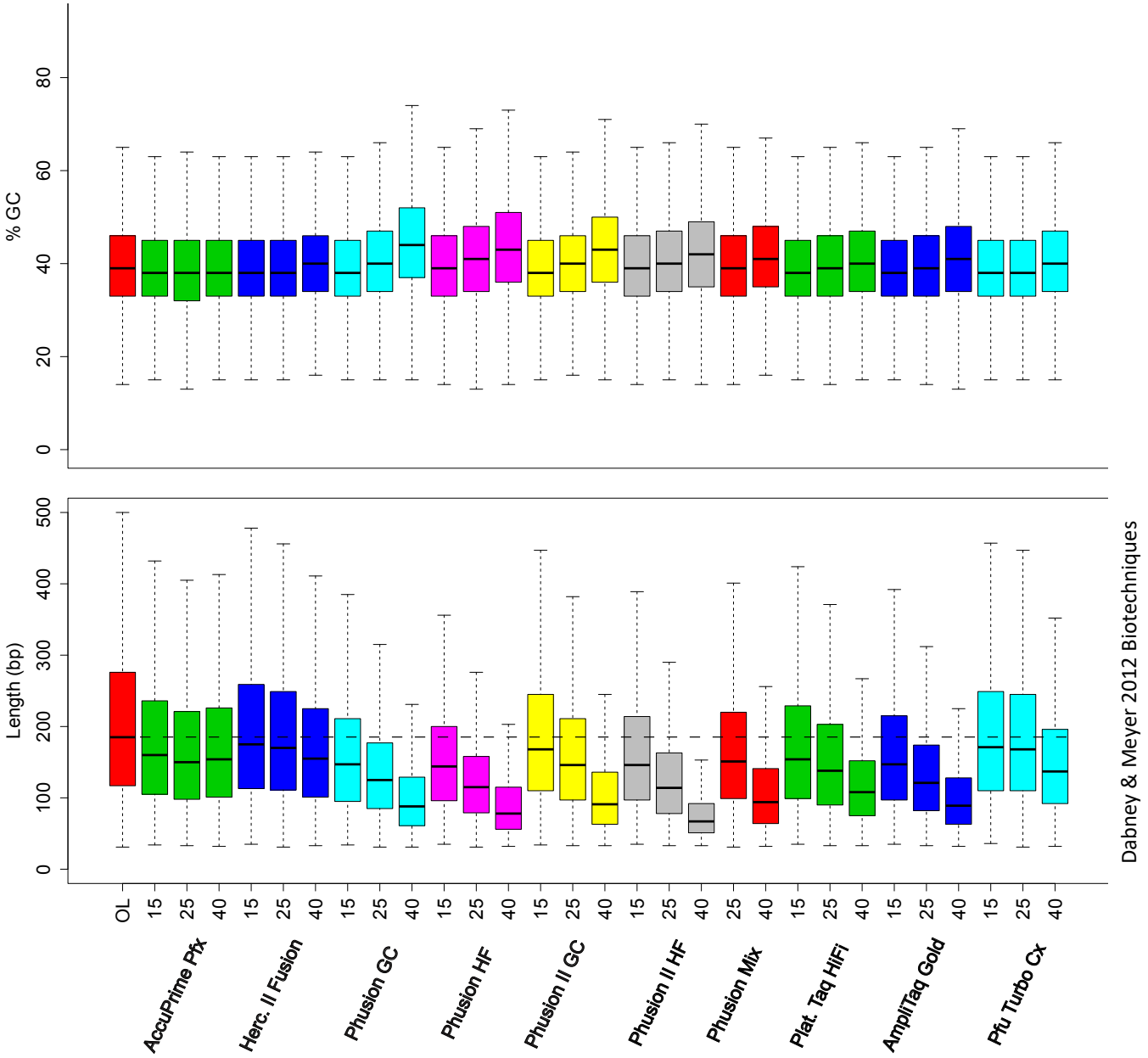


# Data biases due to experimental procedures

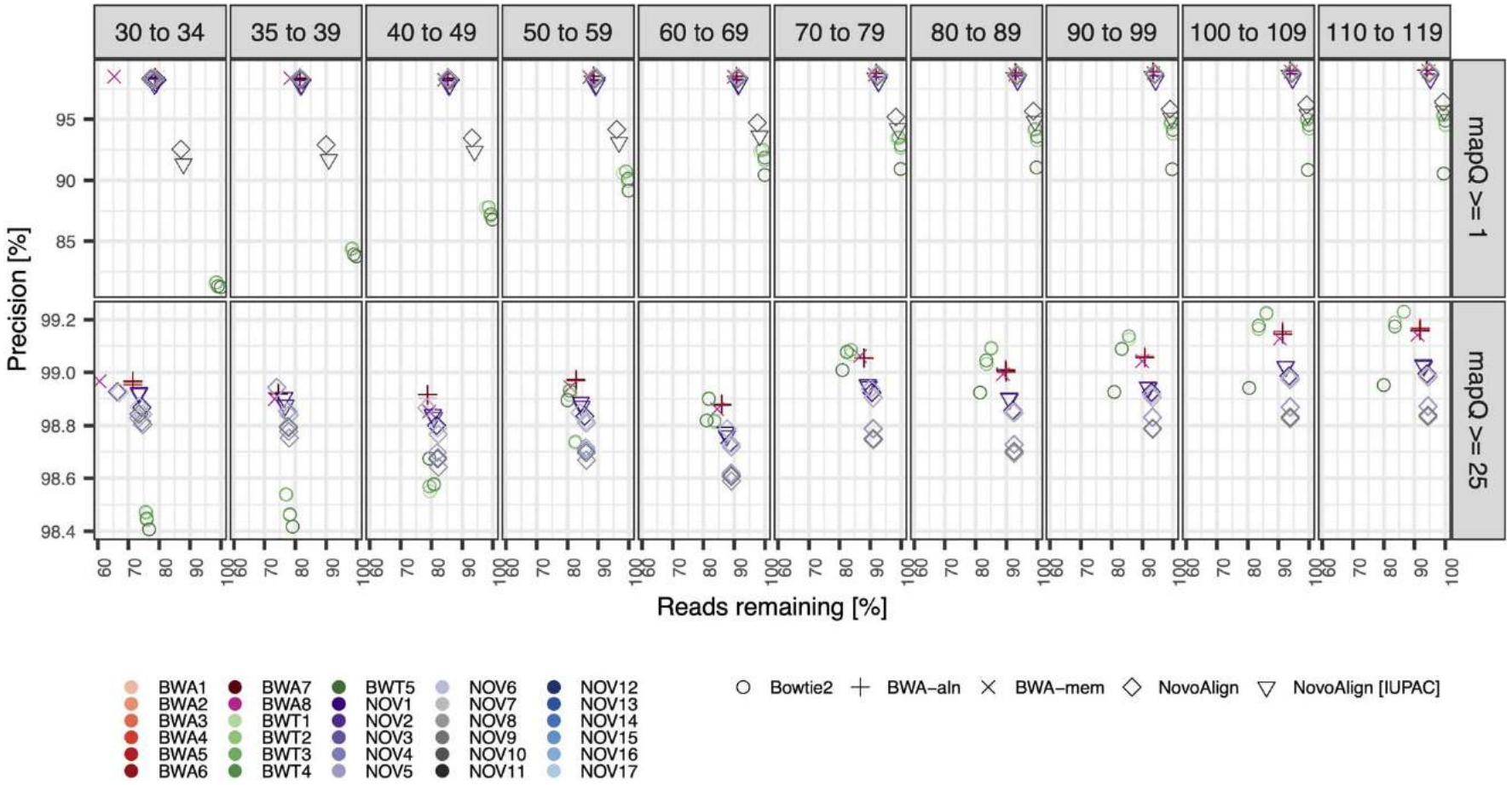
The performance of DNA polymerases used to perform PCR depends on GC content and fragment length

Some polymerases will amplify preferentially high GC-content DNA (i.e., microbial DNA)

Given the distribution of ancient DNA fragment lengths, length bias should be minimised



# Impact of read mapping strategies when processing ancient DNA sequencing data



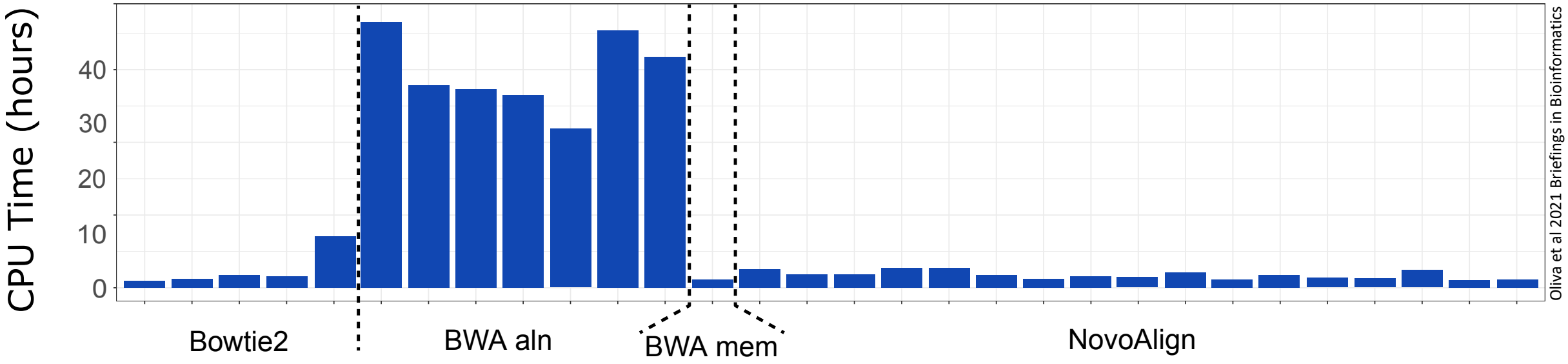

Oliva et al 2021 Briefings in Bioinformatics

Due to the short length of ancient DNA reads, mapping software performance varies for:

- Precision, where the longer the read, the more accurate the mapping
- The proportion of mapped reads

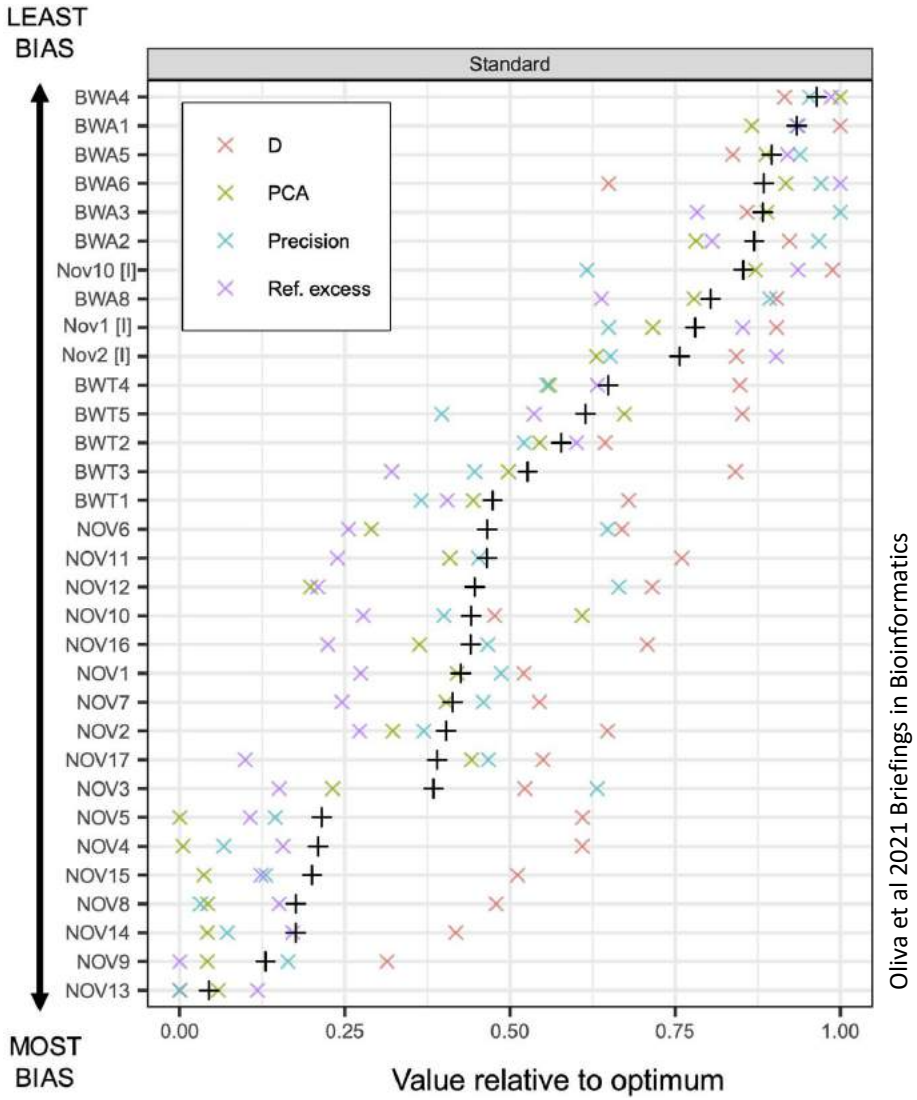
Filtering out reads with low mapping quality generally improves the mapping process

# Impact of read mapping strategies when processing ancient DNA sequencing data




Reads mapping strategy must take into account CPU time, especially if resources are limited

# Impact of read mapping strategies when processing ancient DNA sequencing data




Oliva et al 2021 Briefings in Bioinformatics

- D: D-statistics estimate the level of relationship between test populations.
- PCA: PCA allows to visualise high complexity datasets in a 2D representation. The closer two individual datasets are in the PCA space, the more similar they are.
- Precision: mapping to the correct genomic location.
- Ref. excess: reference allele bias.

Reference bias leads to wrong population genetics inferences

# Summary

- Post-mortem processes impact DNA integrity and survival
- We can observe characteristic patterns in ancient DNA sequencing data
- It is possible to repair ancient DNA damage experimentally
- Bioinformatic methods can help estimate the amount of contamination
- Mapping of ancient DNA sequencing data requires balancing between key performance indicators

**Perspective** Nature | www.nature.com

# Ethics of DNA research on human remains: five globally applicable guidelines

https://doi.org/10.1038/s41586-021-04008-x

Received: 20 March 2021

Accepted: 8 September 2021

Published online: 20 October 2021

Check for updates

Songül Alpaslan-Roodenberg[1,2], David Anthony[3,4], Hiba Babiker[5], Eszter Bánffy[6], Thomas Booth[7], Patricia Capone[8], Arati Deshpande-Mukherjee[9], Stefanie Eisenmann[5,10], Lars Fehren-Schmitz[11,12]✉, Michael Frachetti[13], Ricardo Fujita[14], Catherine J. Frieman[15], Qiaomei Fu[16], Victoria Gibbon[17], Wolfgang Haak[5], Mateja Hajdinjak[7], Kerstin P. Hofmann[6], Brian Holguin[18], Takeshi Inomata[19], Hideaki Kanzawa-Kiriyama[20], William Keegan[21], Janet Kelso[10], Johannes Krause[10], Ganesan Kumaresan[22], Chapurukha Kusimba[23], Sibel Kusimba[23], Carles Lalueza-Fox[24], Bastien Llamas[25,26], Scott MacEachern[27], Swapan Mallick[1,28,29], Hirofumi Matsumura[30], Ana Y. Morales-Arce[31], Giedre Motuzaite Matuzeviciute[32], Veena Mushrif-Tripathy[9], Nathan Nakatsuka[1], Rodrigo Nores[33], Christine Ogola[34], Mercedes Okumura[35], Nick Patterson[4,29], Ron Pinhasi[2,36], Samayamantri P. R. Prasad[37], Mary E. Prendergast[38]✉, Jose Luis Punzo[39], David Reich[1,4,28,29]✉, Rikai Sawafuji[40], Elizabeth Sawchuk[41,42], Stephan Schiffels[5,10], Jakob Sedig[1,4]✉, Svetlana Shnaider[43], Kendra Sirak[1,4]✉, Pontus Skoglund[7], Viviane Slon[44,45], Meradeth Snow[46], Marie Soressi[47], Matthew Spriggs[15,48], Philipp W. Stockhammer[10,49]✉, Anna Szécsényi-Nagy[50], Kumarasamy Thangaraj[37,51], Vera Tiesler[52], Ray Tobler[23,25], Chuan-Chao Wang[53,54], Christina Warinner[10,55], Surangi Yasawardene[56] & Muhammad Zahir[5,57]

## Box 1

### Five globally applicable guidelines for DNA research on human remains

(1) Researchers must ensure that all regulations were followed in the places where they work and from which the human remains derived.
(2) Researchers must prepare a detailed plan prior to beginning any study.
(3) Researchers must minimize damage to human remains.
(4) Researchers must ensure that data are made available following publication to allow critical re-examination of scientific findings.
(5) Researchers must engage with stakeholders from the beginning of a study and ensure respect and sensitivity to other stakeholder perspectives.

**Thank you for your attention!**

Don’t forget to become a member of

AUSTRALIAN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY SOCIETY

https://www.abacbs.org/

GSA GENETICS SOCIETY OF AUSTRALASIA INC.

https://genetics.org.au/


www.adelaide.edu.au/acad

@DNATimeTravel @ACAD_Research_

bastien.llamas@adelaide.edu.au


THE UNIVERSITY of ADELAIDE

ACAD Australian Centre for Ancient DNA