Introduction to Ancient Dietary Metagenomes

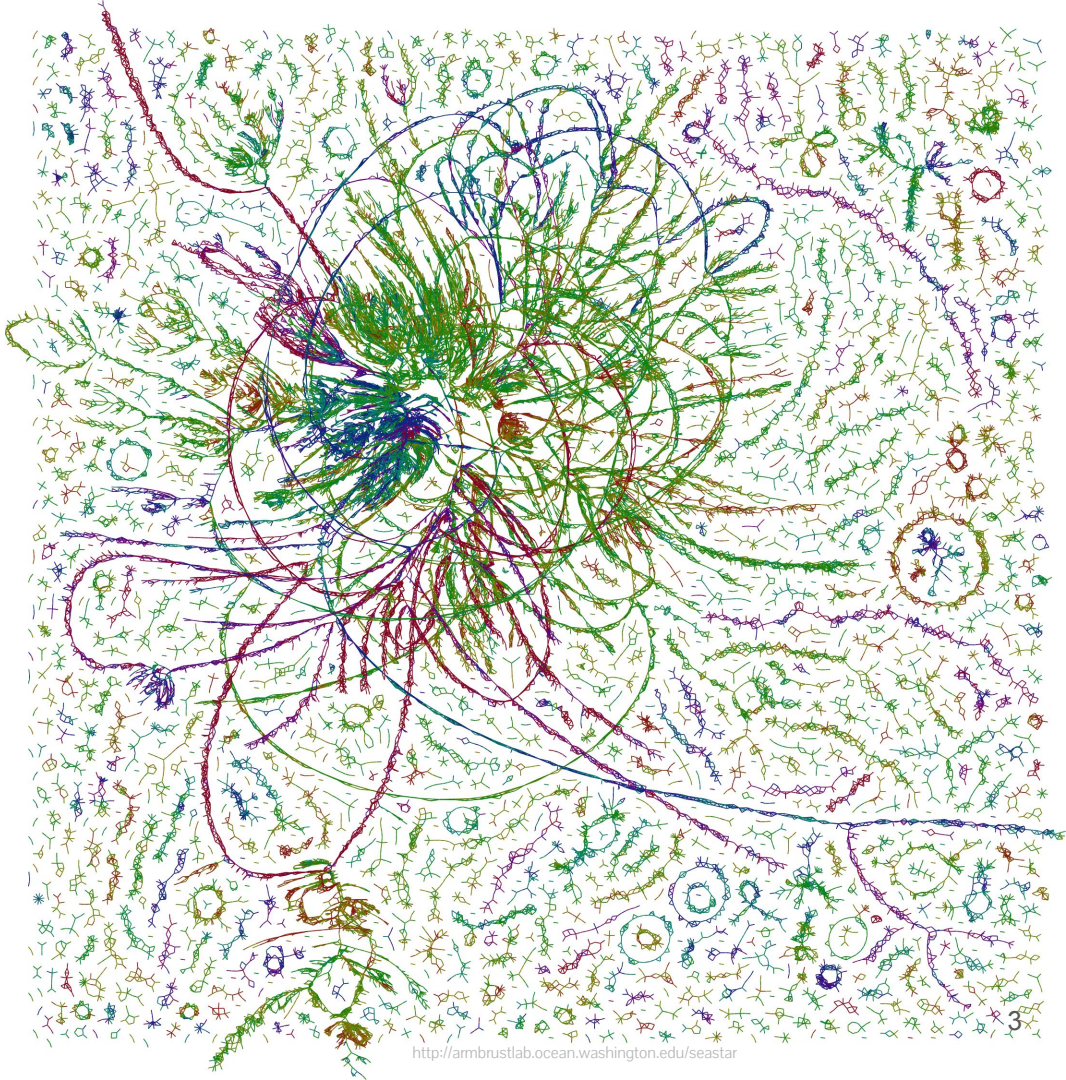
James A. Fellows Yates
(fellows[at]shh.mpg.de)

Department of Archaeogenetics
Max Planck Institute for the Science of Human History, Jena

Outline

1. Introduction
2. History of research
3. Obstacles
4. Case studies
5. Current/Future research
6. Summary

(Dietary) Metagenome:
Total genetic material of a sample (derived from consumed [eukaryotic] dietary components)



http://armbrustlab.ocean.washington.edu/seastar

James A.

Why DNA?

Hardy et al. (2015) Quat. Int.

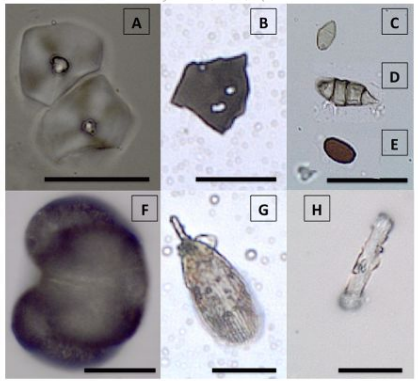


Fig. 3. Microfossils extracted from dental calculus. A – starch granules (G/9b 520–525), B – microcharcoal (G/22 705–710), C – fungi, type 1 spore (G/9b 520–525), D – fungi, type 3 spore (G/22 705–710), E – fungi, type 2 spore (G/9b 520–525), F – pine pollen grain (G/9b 520–525), G – Lepidoptera (moth and butterfly) (G/22 705–710), wing scale, H – long smooth long cell phytolith (G/9b 520–525), note fragments of calculus still adhering to it.

Microfossils

"Easy" to find

Difficult to identify to species level

Bocherens et al . (2015) Quat. Int.

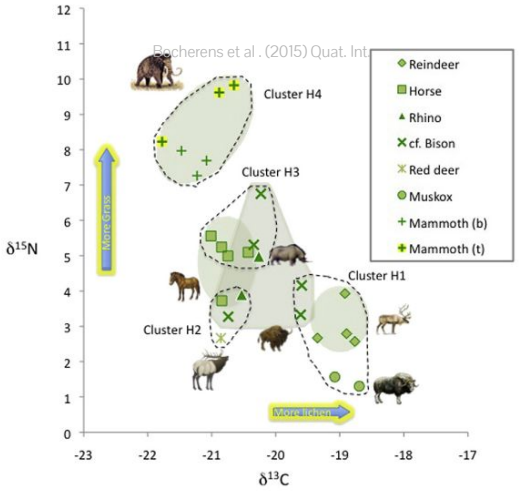
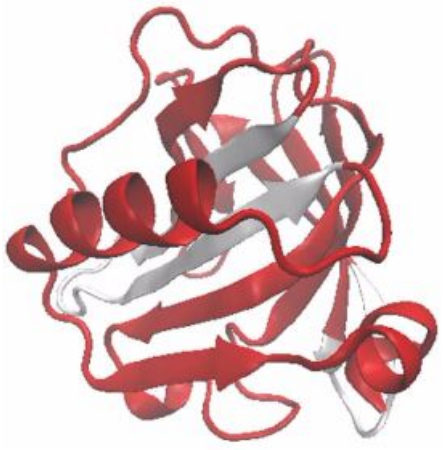


Fig. 6. $\delta^{13}C$ and $\delta^{15}N$ values of herbivorous species with grouping of the values according to species and to the cluster defined through the cluster analysis (Fig. 5).

Stable Isotopes

Well established

Reflects 'Bulk' diet



Proteomics

Give tissue specific information

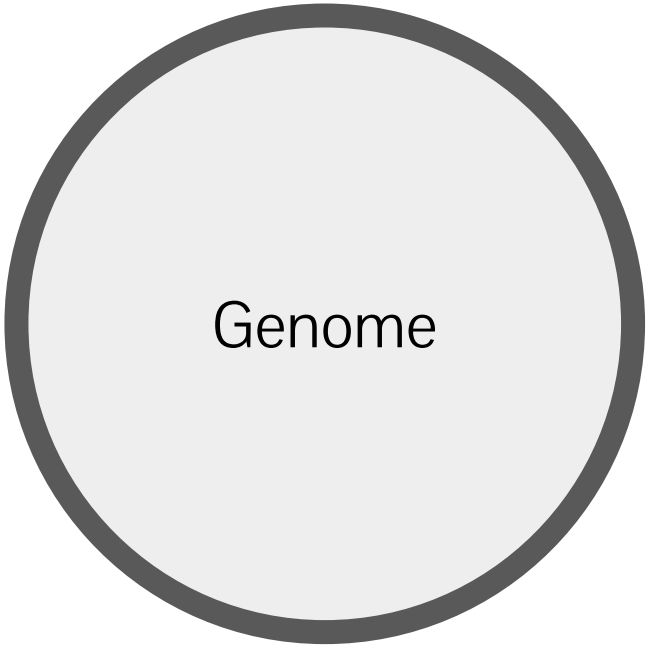
More limited databases

DNA

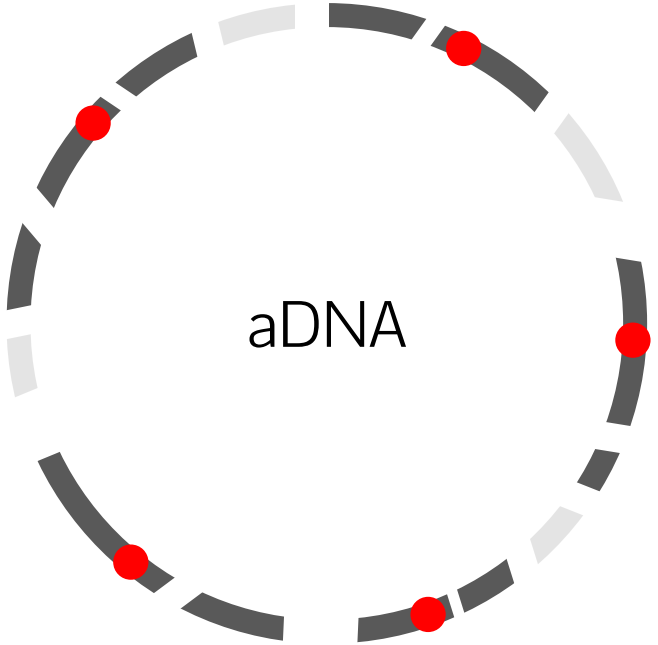
High taxonomic resolution!

Large databases!

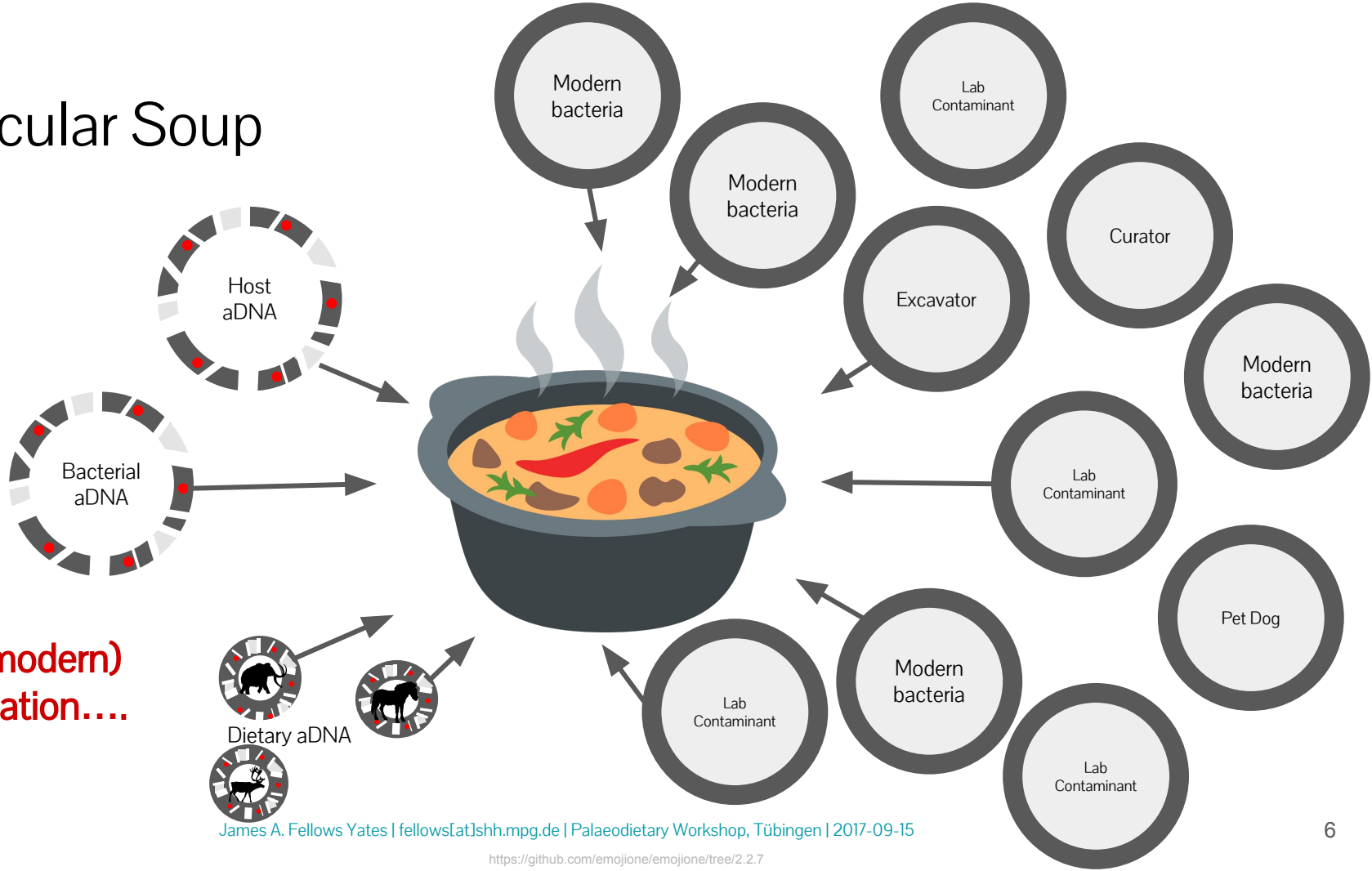
Ancient Dietary Metagenome



aDNA:
Fragmented (short)
Low yield (few)
Damaged (mistakes)



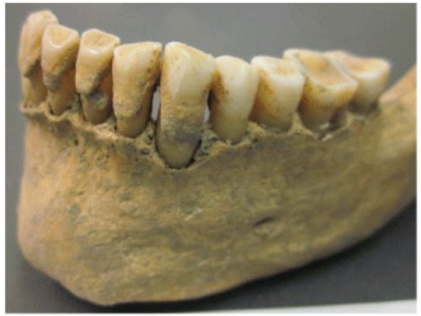
Molecular Soup



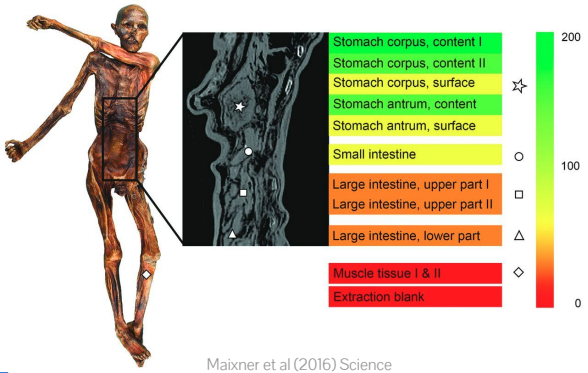
Lots of (modern) contamination….

Where to find?

Potential Sources of Dietary DNA



Warinner et al. (2014) Nat. Gen.



Maixner et al (2016) Science

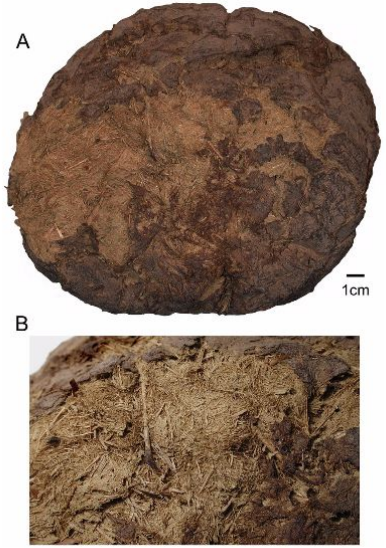
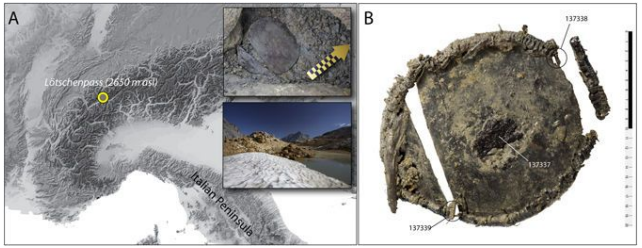


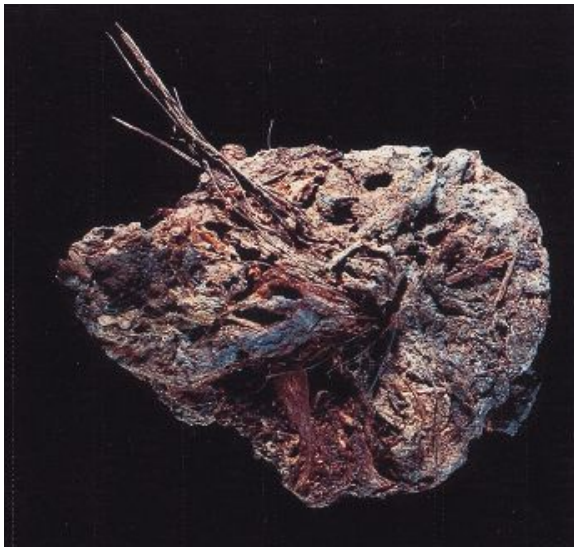
Fig. 2. **(A)** Single dung bolus presumed to be *Mammuthus* from Bechan Cave. **(B)** Close-up of bolus showing the size of poorly-chewed contents.

Karpinski et al. (2017) Quat. int

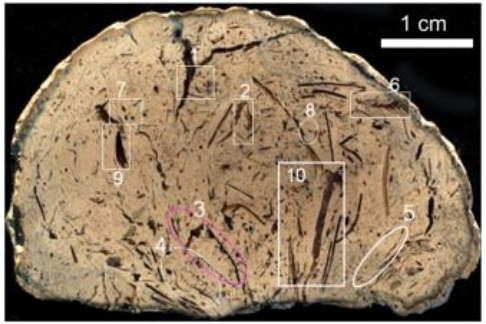


Colonese et al. (2017) Sci. Rep.

Palaeofaeces and Coprolites



Poinar et al. (2001) PNAS



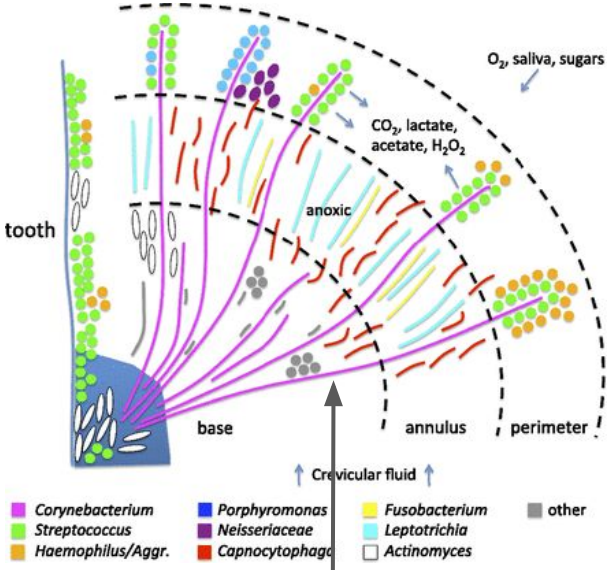
Kohsla et al. (2014) Palaeog., Palaeoc., Palaeoe.

Mineralised palaeofaeces == Coprolite

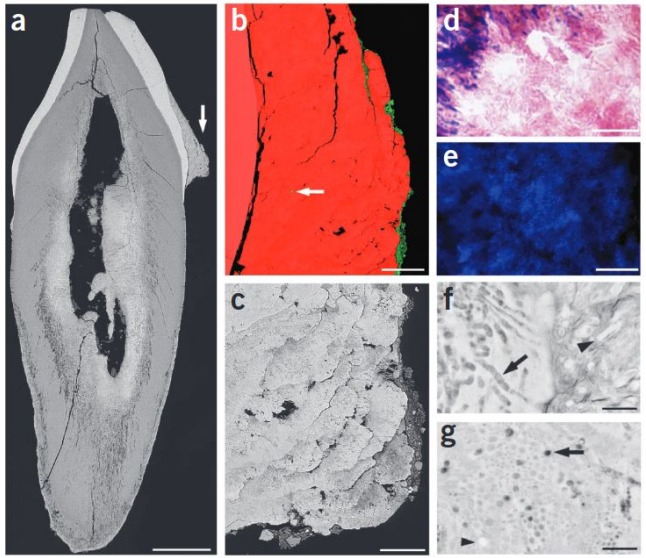
Direct waste of digestion

Rare

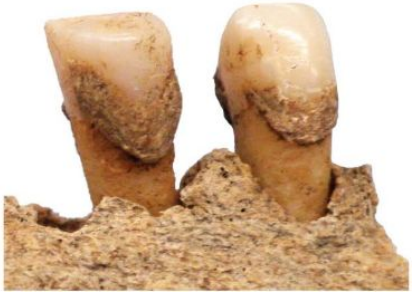
Dental Calculus



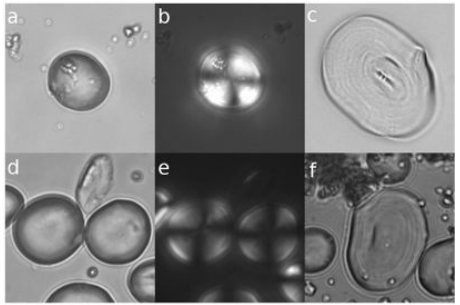
Welch et al (2016) PNAS



Warinner et al (2014) Nat. Gen.



Warinner et al (2014) Sci. Rep.



Henry et al. (2011) PNAS

Bacterial biofilm == Mineralised Plaque/Tartar == Calculus

Not directly related to consumption - *may* incorporate foodstuff

Common

A brief history of...

Early Days

Table 1. Number of clones matching families in the database at zero, one, or two differences. Numbers in parentheses after family names indicate number of genera in that family matching the assigned clones in the database. Sequence cluster designations refer to Fig. 4. Final column gives the assigned family or order [nomenclature according to (*28*)]. NID, no sequence in database matches clones.

Sequence cluster	No. of clones (n = total)	Zero differences	One difference	Two differences	Family or order
A	6	Capparaceae (2)	Capparaceae (1) Brassicaceae (1) Gyrostemonaceae* (2)	Capparaceae (2) Brassicaceae (1)	Capparales
	6	NID	Capparaceae (2)	Capparaceae (2) Brassicaceae (1) Gyrostemonaceae* (2)	
	7 (n = 19)	NID	NID	Capparaceae (2)	
B	2	Poaceae (3)	Poaceae (8)	Poaceae (6)	Poaceae
	1 (n = 3)	NID	Poaceae (3)	Poaceae (8)	
C	6	Liliaceae (2)	Liliaceae (7)	Liliaceae (14)	Liliaceae
	3	NID	Liliaceae (2)	Liliaceae (7)	
	9 (n = 18)	NID	NID	Liliaceae (2)	
D	1 (n = 1)	NID	Rubiaceae (3)	Rubiaceae (2) Buddlejaceae (2) Verbenaceae (3)	Rubiales Scrophulariales Lamiales
E	2 (n = 2)	NID	NID	Lamiaceae (13) Scrophulariaceae (2) Acanthaceae (2)	Lamiales/ Scrophulariales
F	1	Euphorbiaceae (1)	Euphorbiaceae (1)	Euphorbiaceae (4) Rosaceae (2) Onagraceae (2)	Euphorbiales Rosales Myrtales
	1 (n = 2)	NID	Euphorbiaceae (2)	NID	
G	1 (n = 1)	NID	Chenopodiaceae (2)	Chenopodiaceae (2)	Chenopodiaceae
Total clones	46	15	13	18	

*Family restricted to Australia.

17 JULY 1998 VOL 281 SCIENCE www.sciencemag.org

Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth *Nothrotheriops shastensis*

Hendrik N. Poinar, Michael Hofreiter, W. Geoffrey Spaulding, Paul S. Martin, B. Artur Stankiewicz,* Helen Bland, Richard P. Evershed, Göran Possnert, Svante Pääbo

DNA from excrements can be amplified by means of the polymerase chain reaction. However, this has not been possible with ancient feces. Cross-links between reducing sugars and amino groups were shown to exist in a Pleistocene coprolite from Gypsum Cave, Nevada. A chemical agent, *N*-phenacylthiazolium bromide, that cleaves such cross-links made it possible to amplify DNA sequences. Analyses of these DNA sequences showed that the coprolite is derived from an extinct sloth, presumably the Shasta ground sloth *Nothrotheriops shastensis*. Plant DNA sequences from seven groups of plants were identified in the coprolite. The plant assemblage that formed part of the sloth's diet exists today at elevations about 800 meters higher than the cave.

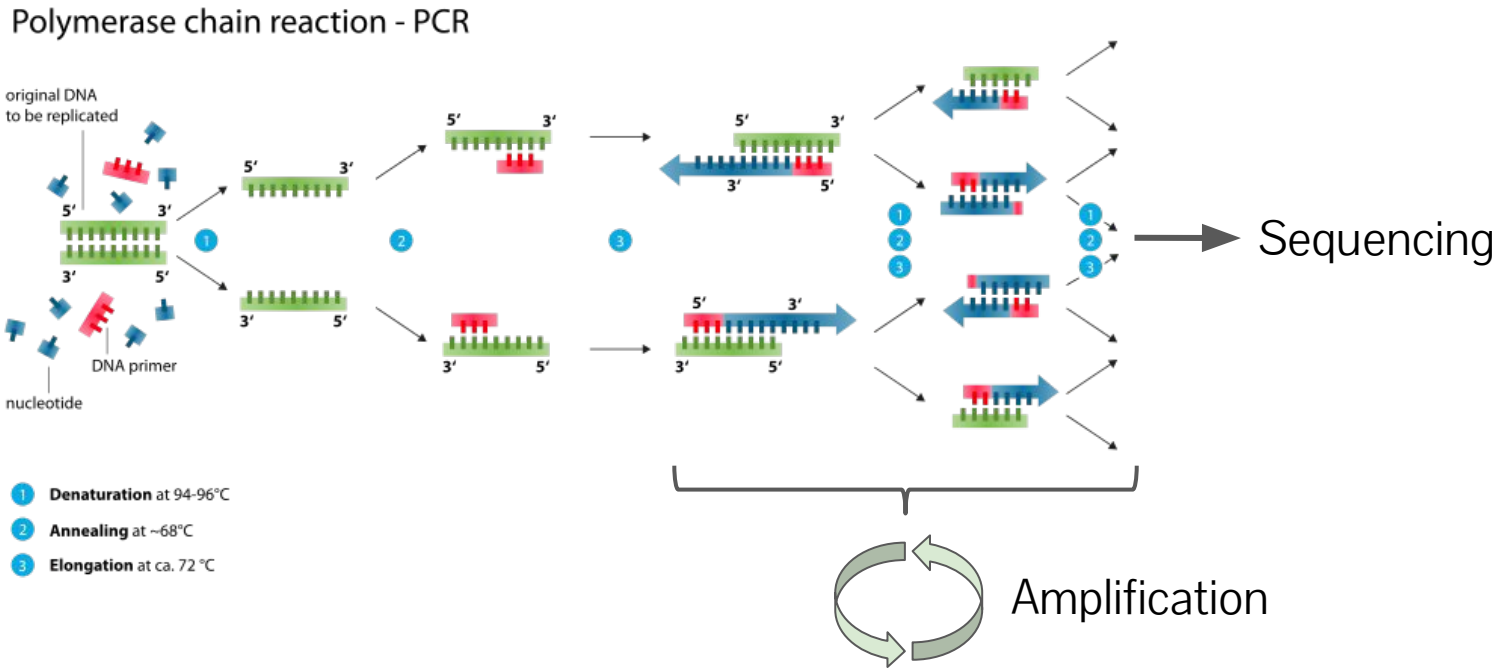
Order, families, and representative genera (common name)	Gypsum Cave coprolite		Flora present	
	Molecular evidence	Macroscopic evidence	Gypsum Cave	Spring Range
Capparales (capers, mustards)	45	—	X	X
Capparaceae				
Cleome, Cleomella, Polansia				
Brassicaceae				
Lepidium, Stanleya, Lesquerrella				
Cyperales (grasses)	10	X (2)	X	X
Poaceae				
(many genera)				
Liliales (agave, yucca)	35	X (1)	—	X
Liliaceae				
Agave, Yucca				
Gentianales (phacelia)	2	—	X	X
Hydrophyllaceae				
Phacelia, Nama				
Lamiales/Scrophulariales (borages, mints)	2	X (3)	X	X
Boraginaceae				
Cryptantha, Lithospermum				
Lamiaceae				
Monardella, Salazaria, Salvia				
Verbenaceae				
Verbena				
Buddlejaceae				
Buddleja				
Scrophulariaceae				
Castilleja, Mimulus, Penstemon				
Rhamnales (grape)	1	—	—	X*
Vitaceae*				
Vitis				
Malvales (mallows)	1	—	X	X
Malvaceae				
Sphaeralcea				
Caryophyllales (saltbushes)	1	X (4)	X	X
Chenopodiaceae				
Atriplex, Chenopodium				
Ephedrales (ephedra)	—	X (5)	X	X
Ephedraceae				
Ephedra				

*In the vicinity of springs and streams only.

Early Days

Primer 'locates' DNA sequence of interest in sample

Primer sequence = based on **known reference**



https://commons.wikimedia.org/wiki/File:Polymerase_chain_reaction.svg by 'Enzoklop'

Early Days

A molecular analysis of dietary diversity for three archaic Native Americans

Hendrik N. Poinar[†‡], Melanie Kuch[†], Kristin D. Sobolik[§], Ian Barnes[¶], Artur B. Stankiewicz[‖], Tomasz Kuder[**], W. Geofferey Spaulding[††], Vaughn M. Bryant[‡‡], Alan Cooper[¶], and Svante Pääbo[†]

PNAS | April 10, 2007 | vol. 98 | no. 8 | 4317–4322

PCR amplifications were performed as described (4) by using the primers listed below for three restriction sites (*Hae*III, *Hinc*II, and *Alu*I), the 9-bp repeat, hypervariable region I, 12S and 16S rRNA genes, and the chlorplast *rbcL* gene: L00635 5′-TGAAAATGTTTAGACGGCCTCACATC-3′; H00708, 5′-TAGAGGGTGAACTCACTGGAAC-3′; L13259, 5′-AATCG-TAGCCTTCTCCACTTCA-3′; H13377, 5′-TATCTTGT-TCATTGTTAACGTTGTGG-3′; L05054, 5′-TAGGAT-GAATAATAGCAGCTCTACCG-3′; H05184, 5′-GGGTG-GATGGAATTAAGGGTGT-3′; L09158, 5′-ATACTACGGT-CAATGCTCTG-3′; H09297, 5′-ATGCTAAGTTAGCTTTA-CAG-3′; L16131, 5′-CACCATGAATATTGTACGGT-3′; H16218, 5′-ATGTGTGATAGTTGAGGGTTG-3′; L16209, 5′-CCCCATGCTTACAAGCAAGT-3′; H16303, 5′-TGGCTT-TATGTACTATGTAC-3′; L16287, 5′-CACTAGGATACCAA-CAAACC-3′; H16379, 5′-CAAGGGACCCCTATCTGAG-3′; 12Sa′, 5′-CTGGGATTAGATACCCCACTAT-3′; 12So, 5′-GTCGATTATAGGACAGGTTCCTCTA-3′; 16S6, 5′-TT-TCGGTTGGGGCGACCTCGGAG-3′; 16S7, 5′-TTGCGCT-GTTATCCCTAGGGTAACT-3′; rbcLZ1, 5′ATGTCAC-CACAAACAGAGACTAAAGCAAGT-3′; rbcL19b, 5′CT-TCTTCAGGTGGAACTCCAG-3′, and rbcL19, 5′-AGATTC-CGCAGCCACTGCAGCCCCTGCTTC-3′.

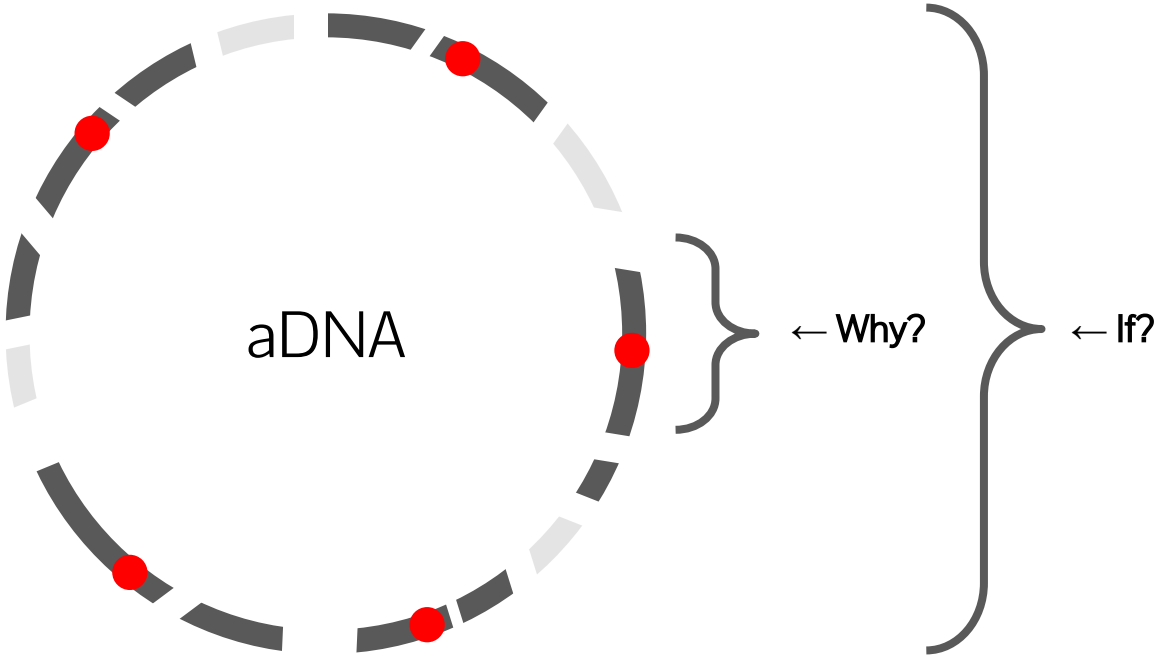
Lots work to design all primers!

Specificity

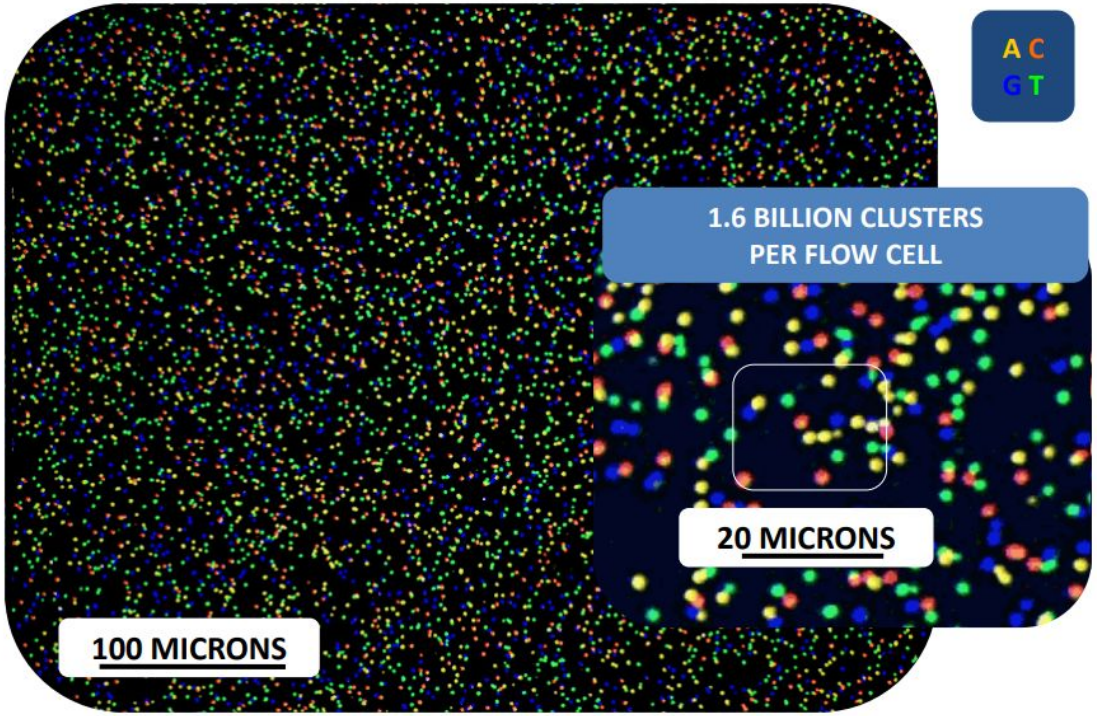
Known target needed

Biochemical requirements

Early Days



NGS Era



illumina.com

Next-Generation-Sequencing
High-throughput
Massively multiplexed
Metagenomic shotgun sequencing

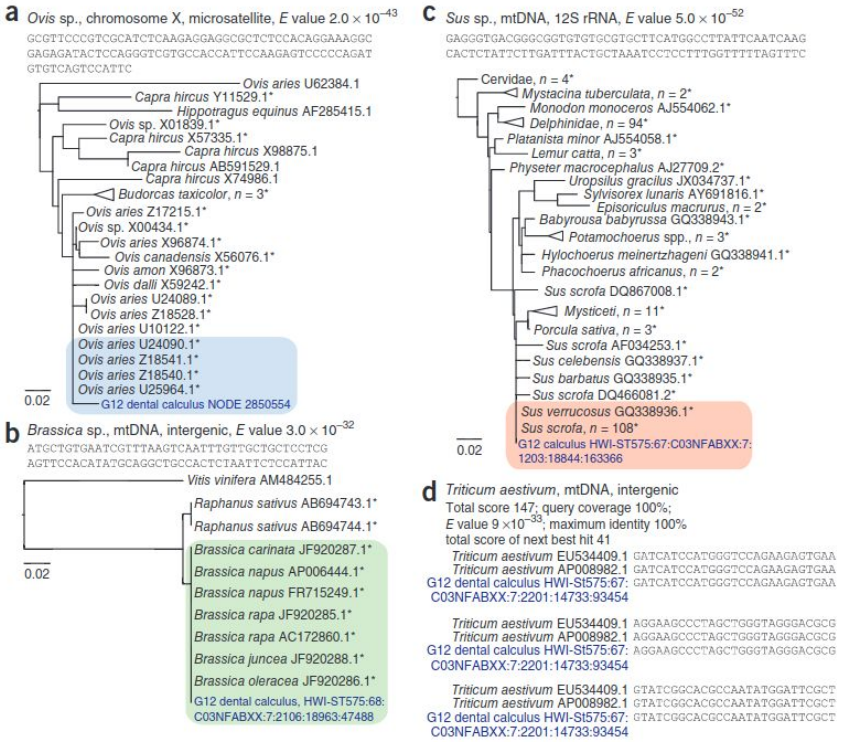
http://evomicsorg.wpengine.netdna-cdn.com/wp-content/uploads/2014/01/Cesky-Krumlov-Lecture-1-ver6.pdf

NGS Era

First 'shotgun' DNA study = Dental Calculus!

Warinner et al. (2014) *Nat. Gen.*

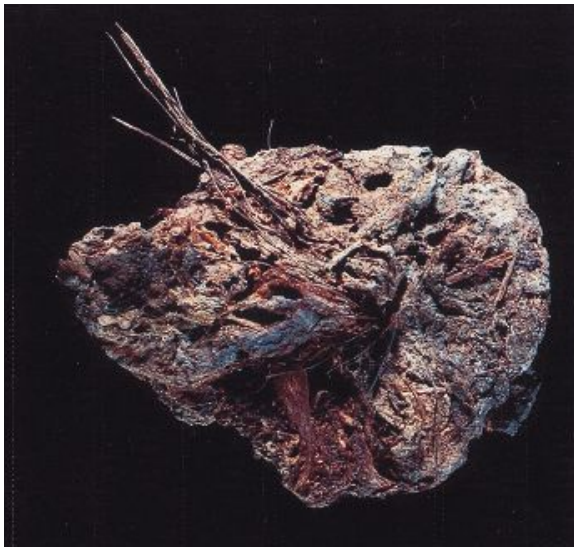
From our metagenomic sequence reads, a total of 487 reads (0.0003%) were confidently identified as eukaryotic organelle sequences; of these, 266 were assigned to the kingdom Viridiplantae, and 21 were assigned to the kingdom Animalia. Within these kingdoms, most of the organelle reads mapped ambiguously to multiple organisms or genera, leaving only 20 reads that could be positively identified at a subfamily level. Of these 20 reads, 17 were of host origin, and the remaining 3 reads matched diagnostic mitochondrial sequences for pig/boar (*Sus* species), crucifer (*Brassica* species) and bread wheat (*Triticum aestivum*). Analysis of assembled contigs additionally identified one putative sheep (*Ovis* species) and several human ($n = 326$) nuclear genomic sequences (**Fig. 4a–d**). Although previous studies have reported trace animal



...but is it real?

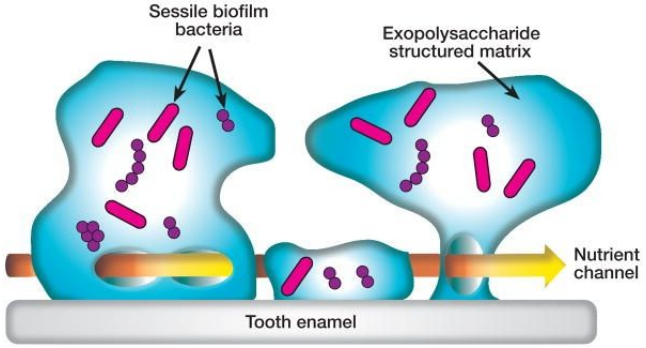
Formation processes

Formation



Poinar et al. (2001) PNAS

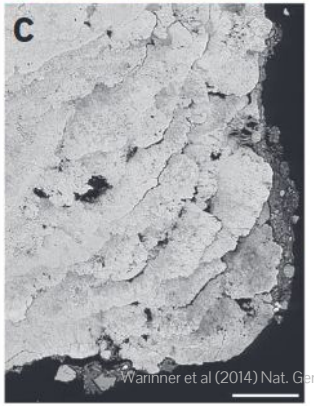
Palaeofaeces - direct waste



Nizet & Esko, 2009

Saliva flow **removes** free DNA?

1-2 months mineralisation

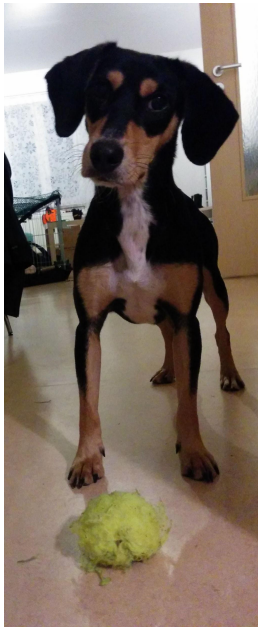


Warinner et al (2014) Nat. Gen.

hyderabadsmiles.com

Dietary DNA more likely **between teeth?**

Formation and Incorporation



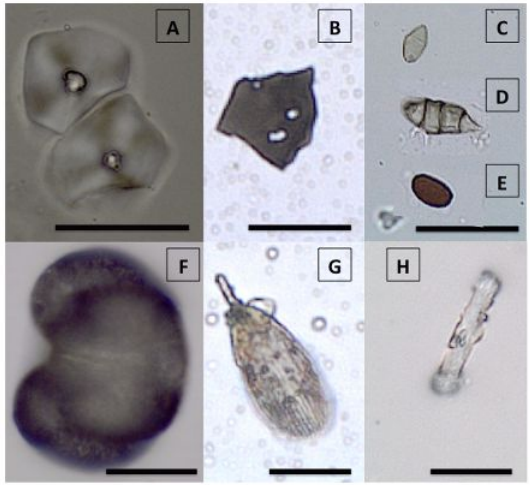


Fig. 3. Microfossils extracted from dental calculus. A = starch granules (G/9b 520–525), B = microcharcoal (G/22 705–710), C = fungi, type 1 spore (G/9b 520–525), D = fungi, type 3 spore (G/22 705–710), E = fungi, type 2 spore (G/9b 520–525), F = pine pollen grain (G/9b 520–525), G = Lepidoptera (moth and butterfly) (G/22 705–710), wing scale, H = long smooth long cell phytolith (G/9b 520–525), note fragments of calculus still adhering to it.

Hardy et al. (2015) Quat. Int.

Public domain

Consumed for sustenance?
Dental hygiene?

Post-depositional preservation

Biomolecular Preservation

Nature **410**, 771-772 (12 April 2001) | doi:10.1038/35071177

Neanderthal DNA: Not just old but old and cold?

Colin I. Smith[1], Andrew T. Chamberlain[2], Michael S. Riley[3], Alan Cooper[4], Chris B. Stringer[5] & Matthew J. Collins[1]

Is the burial environment suitable for aDNA preservation?

Preservation:
Applies to all aDNA
Still not well understood

Trend:
Cold, dry, and stable
Hot, wet, fluctuating

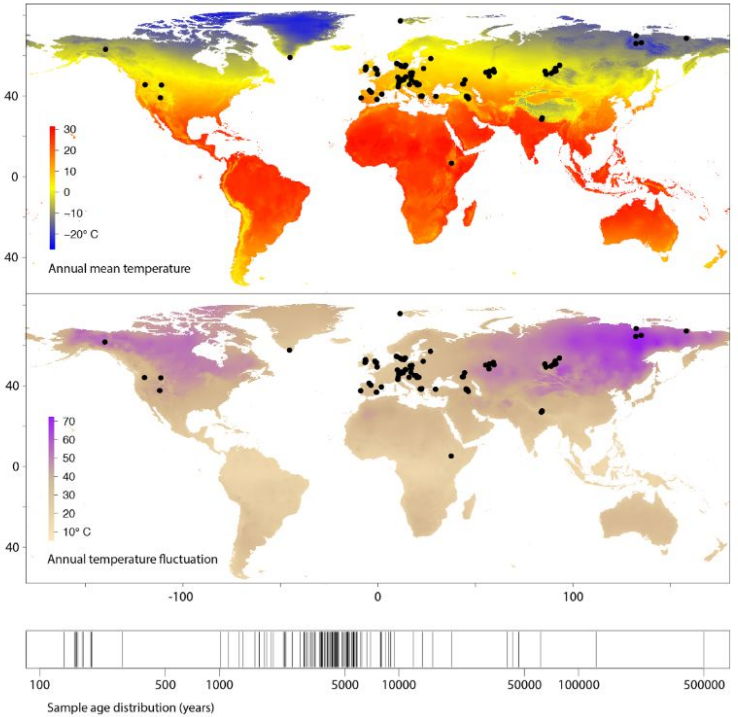


Figure 1. Locations of 185 samples (n = 94 unique sites) used in paleogenomic meta-analysis, global variation in mean temperature and temperature fluctuation, and timeline of sample ages. Note the absence of sites with annual mean temperature >20°C, reflecting known preservation bias toward cooler climates (25).

Burial Environment

http://www.abc.net.au/news/2017-06-06/megafauna-fossils-in-the-naracoote-caves/8593336

Possibility of ‘ancient’ eukaryotic contamination from burial environment?

Burial Environment

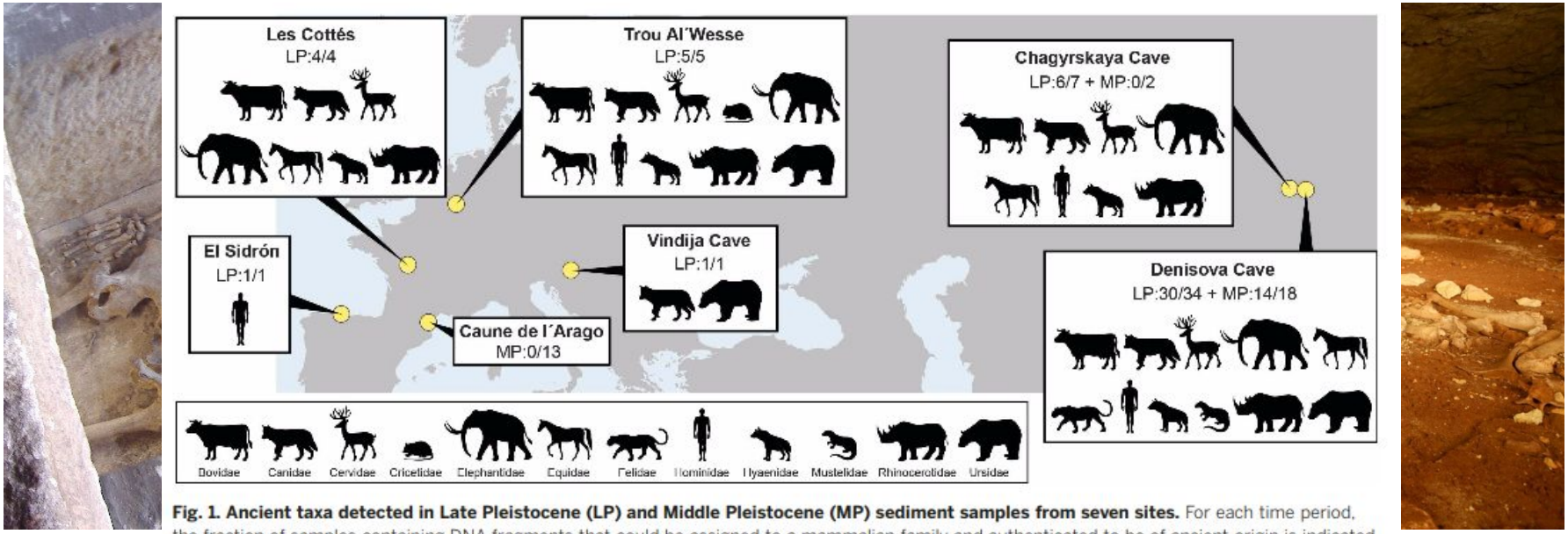


Fig. 1. Ancient taxa detected in Late Pleistocene (LP) and Middle Pleistocene (MP) sediment samples from seven sites. For each time period, the fraction of samples containing DNA fragments that could be assigned to a mammalian family and authenticated to be of ancient origin is indicated. The shaded symbols representing each family are not to scale.

Slon et al (2017) Science

Excavation

Sánchez-Quinto and Lalueza-Fox (2014) Proc. Roy. Soc. B.

Excavation under clean conditions?
Excavator handling?

Post-depositional storage

Storage and Conservation

Clean, cool dry?

Old, damp, rotting?

Organic or
Synthetic adhesive?

Schellmann et al. (2013) Studies in Conservation:

Bone, hide, rabbit skin, mammalian gelatin, fish swim bladders, fish skin/bone/cartilage

Lab protocols

Contamination

Build DNA library in **clean** lab

Contamination

ELSEVIER

Journal of Archaeological Science 34 (2007) 1361–1366

Journal of Archaeological SCIENCE

http://www.elsevier.com/locate/jas

Animal DNA in PCR reagents plagues ancient DNA research

Jennifer A. Leonard [a,b,c,*], Orin Shanks [d,1], Michael Hofreiter [e], Eva Kreuz [e], Larry Hodges [d], Walt Ream [d], Robert K. Wayne [b], Robert C. Fleischer [a]

3. Results

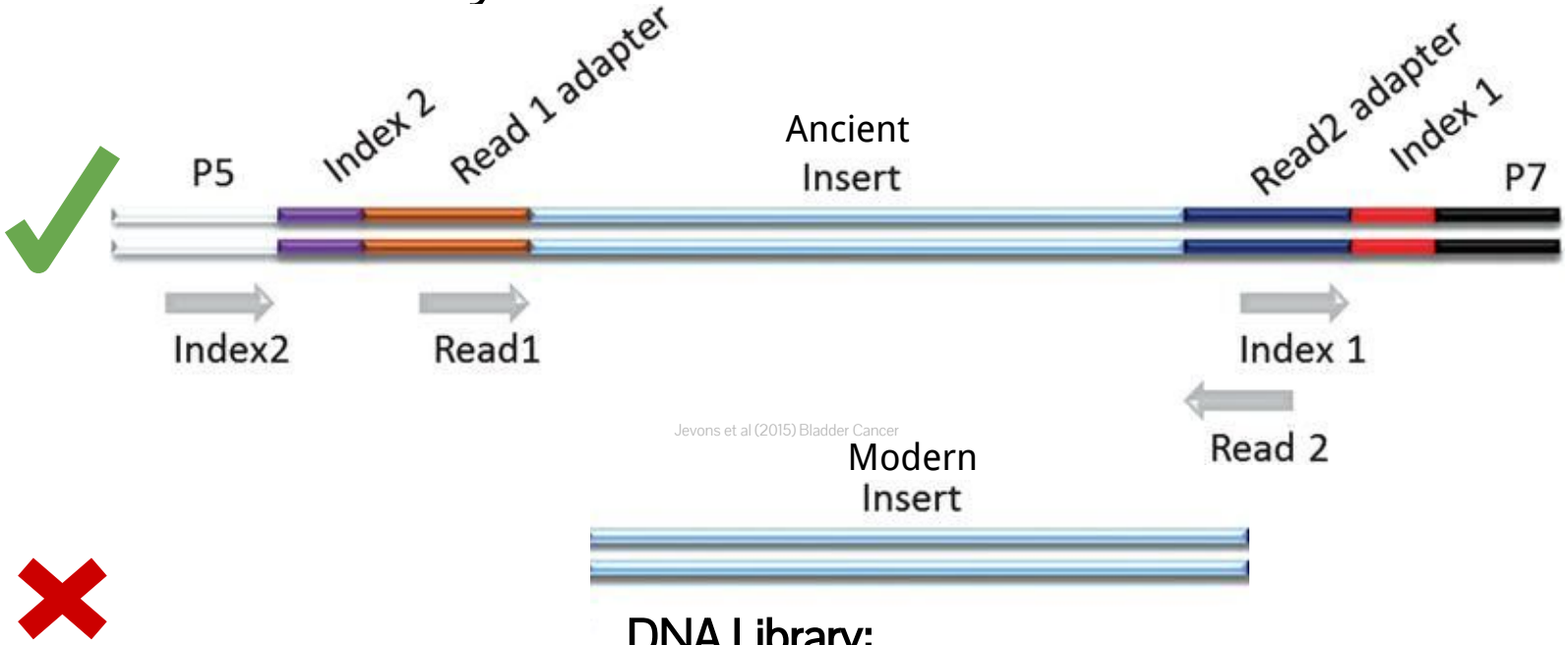
3.1. Assessing previous studies

The survey of anomalous PCR results from past ancient DNA projects on non-domestic species yielded a very strong pattern across laboratories. Cow, pig, and chicken were identified in independent experiments in all four laboratories. Laboratory animals were also identified, including mouse (*Mus musculus*), goat (*Capra hircus*), rabbit (*Oryctolagus cuniculus*) and guinea pig (*Cavia porcellus*), but at lower frequencies. Mouse was identified in two independent laboratories (MPI and SI), and goat, rabbit and guinea pig were each identified in a single lab (MPI, SI and OSU, respectively).

Reagents **contaminated**
(but **better** than before!)

Details of commercial dNTP preparation are proprietary. However, one supplier revealed that deoxynucleoside monophosphates are obtained by hydrolysis of animal DNA and then phosphorylated chemically to produce triphosphates.

Ancient DNA Library



DNA Library:

Add **unique sample-specific** barcodes

Only molecules with adapters sequenced (**contamination ignored**!)

Data analysis

Tools

Technology

The data deluge

Businesses, governments and society are only starting to tap its vast potential

Feb 25th 2010

http://www.economist.com/node/15579717

Is the analysis suitable for NGS data?

BLAST (PCR) vs. MALT (NGS)

Warinner et al. 2014: 100 million reads - **6 weeks**

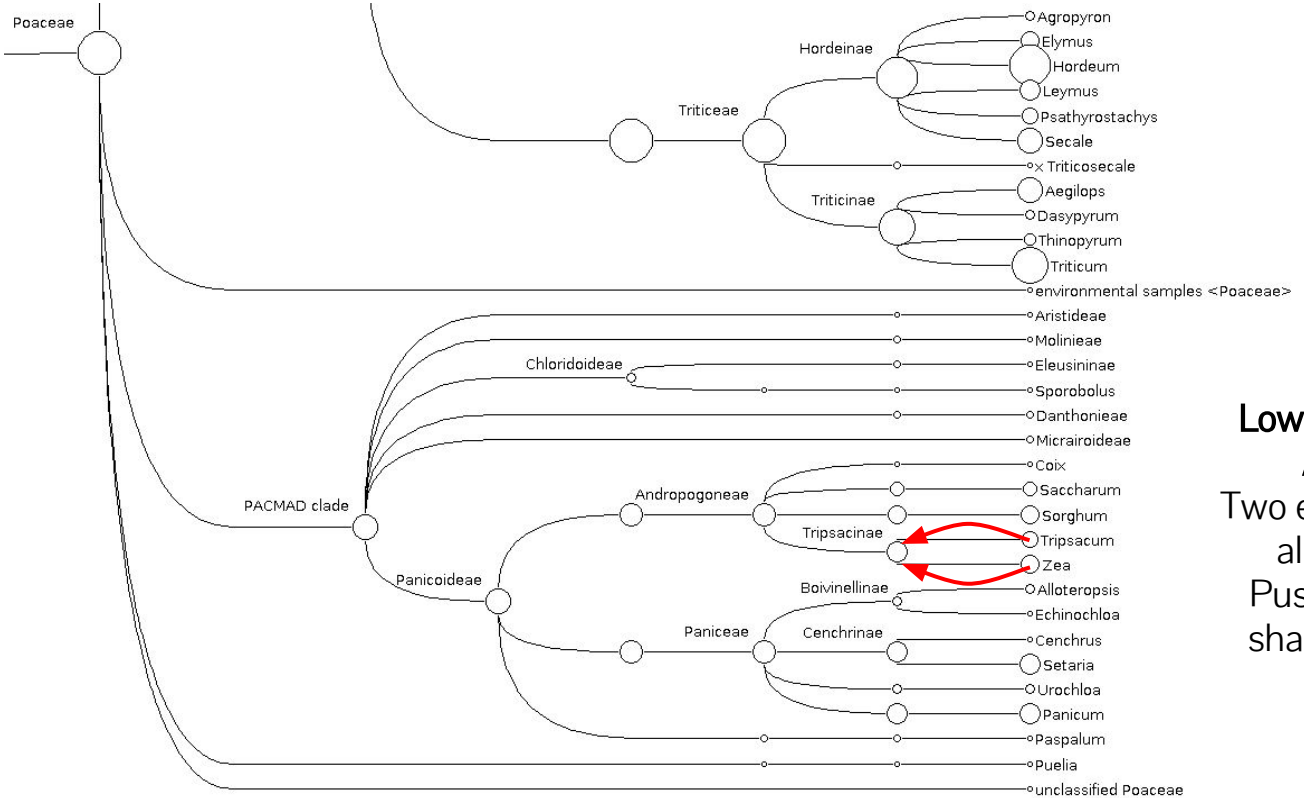
Herbig et al. 2016: 1 billion reads - **days**

(Read = representation of a nucleotide sequence from a DNA molecule)

MALT is substantially faster than *lambda* and achieves a runtime allowing for the processing of a complete HiSeq run within a few days. MALT performs the

Herbig et al. (2016) bioRxiv

Tools

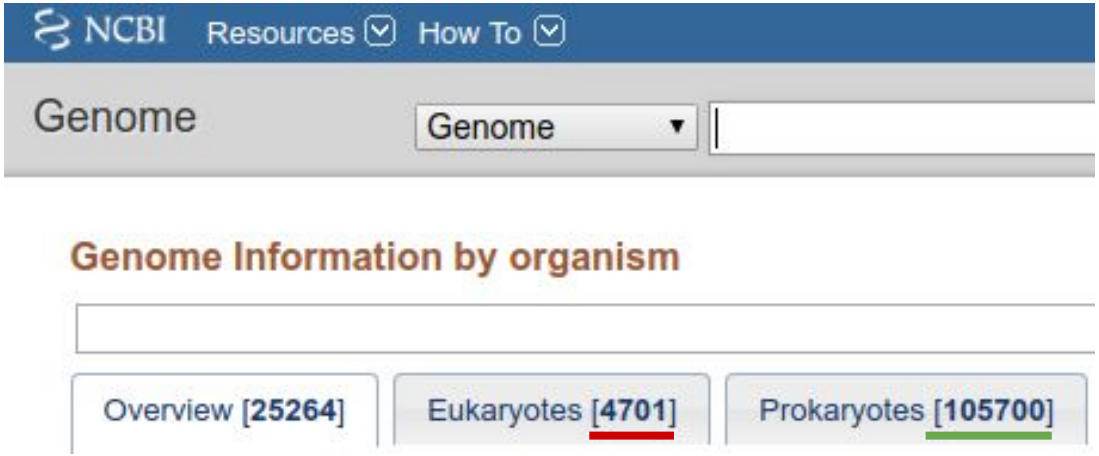


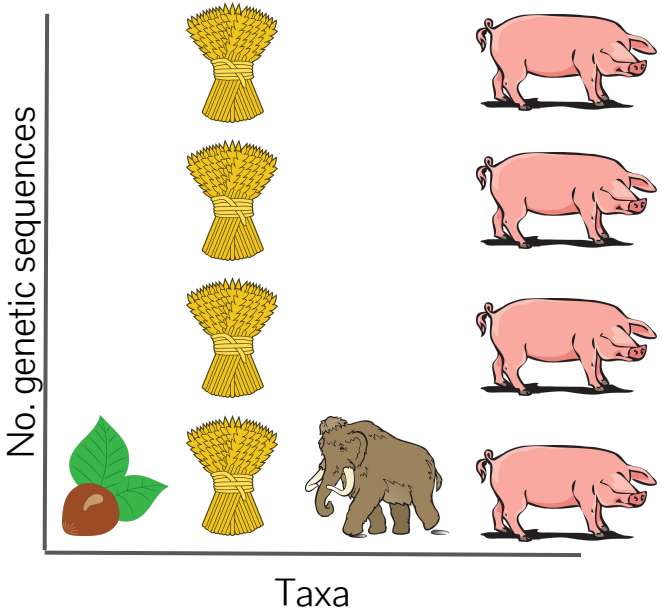
Lowest Common Ancestor

Two equal probable alignments?

Push read up to shared ancestor

Database Bias



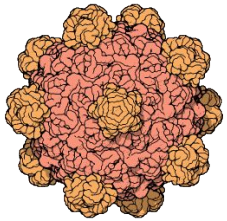


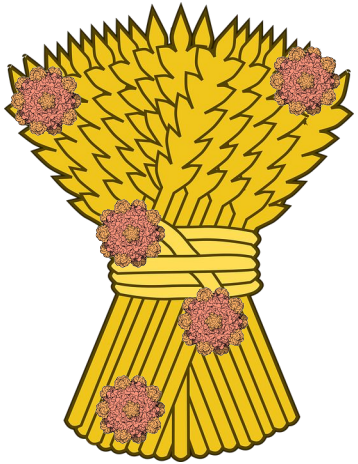
Bias in numbers of genomes? What references are missing?

Database Bias

Quality of the reference sequence?
Contamination?

Reagent



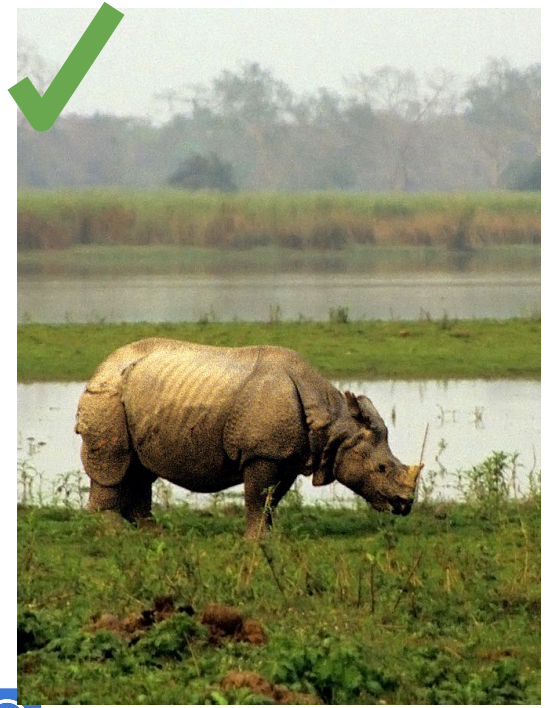


Reference sequence

Modern Sample

http://grahametherington.blogspot.de/2014/09/why-you-should-qc-your-reads-and-your.html

Database Bias



Do we have all the reference sequences we need?

Authentication

Search Parameters

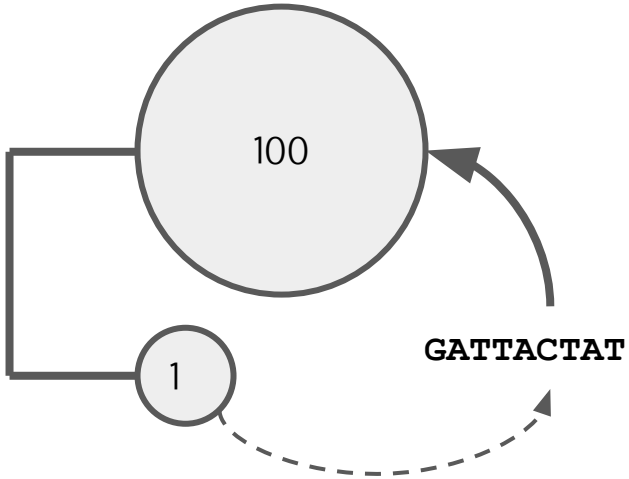
Sensitivity: more false positives?

- More species with greater risk of incorrect species identification

Vs.

Specificity: more false negatives?

- Fewer species with lower risk of incorrect species identification

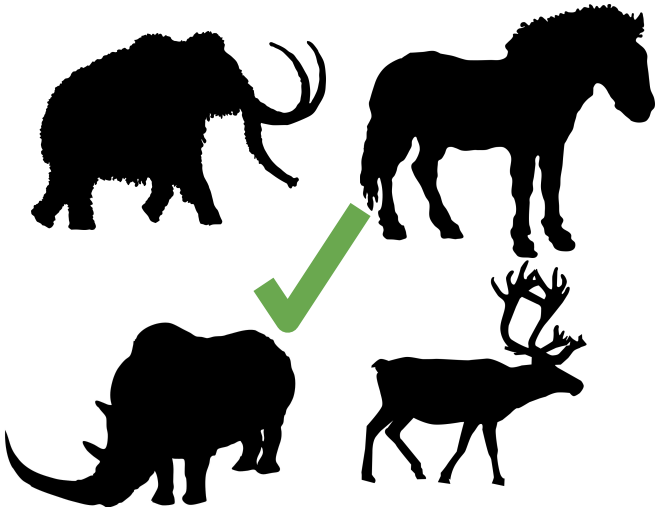


Read attraction away from true source, due to more (complete) genomes of other species?

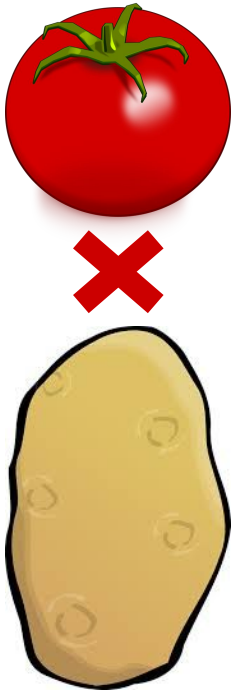
Dietary Composition

Tim Evanson, CC-SA commons.wikimedia.org

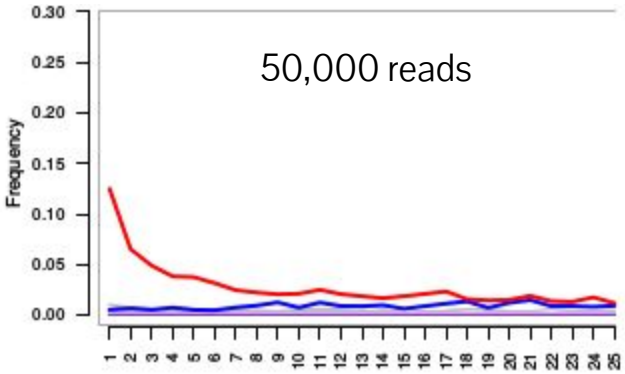
Does the overall dietary DNA signal **make sense**?

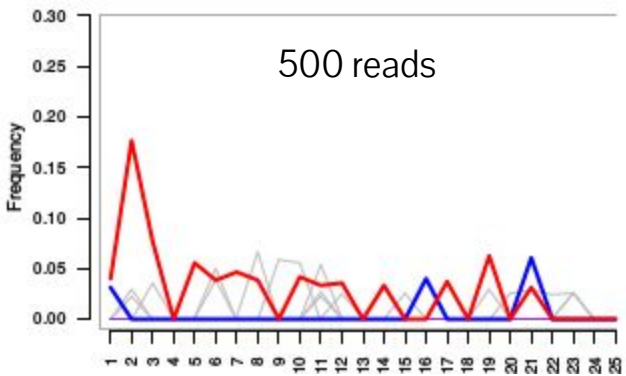


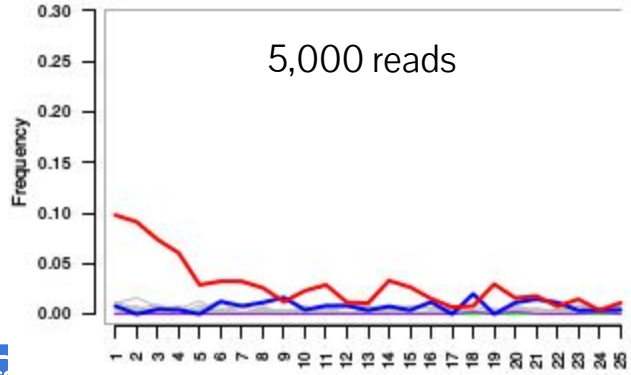
Dorothée Drucker
Pylopic.org - Steven Traver, Mercedes Yrayzoz

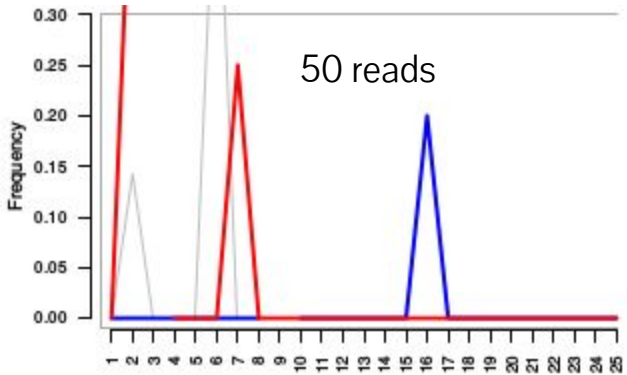


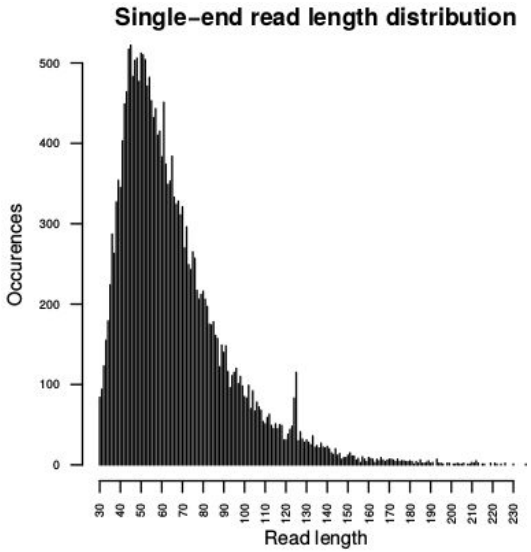
DNA Damage and Fragmentation







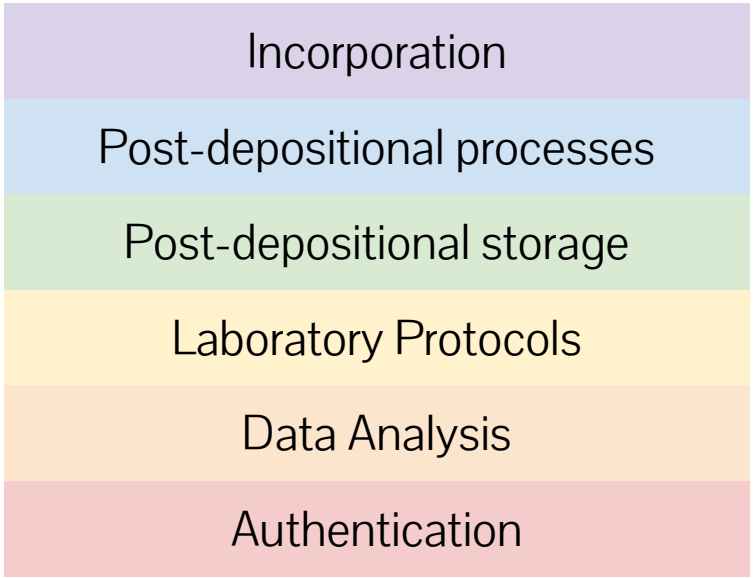




C>T

Are our reads old?
Fewer reads = harder to check

Review


Incorporation
Post-depositional processes
Post-depositional storage
Laboratory Protocols
Data Analysis
Authentication

Case Studies

Case Study

- Incorporation
- Post-depositional processes
- Post-depositional storage
- Laboratory Protocols
- Data Analysis
- Authentication

RESEARCH ARTICLE

Retroviral DNA Sequences as a Means for Determining Ancient Diets

Jessica I. Rivera-Perez[1]*, Raul J. Cano[2], Yvonne Narganes-Storde[3], Luis Chanlatte-Baik[3], Gary A. Toranzos[1]

PLOS ONE | DOI:10.1371/journal.pone.0144951 December 14, 2015

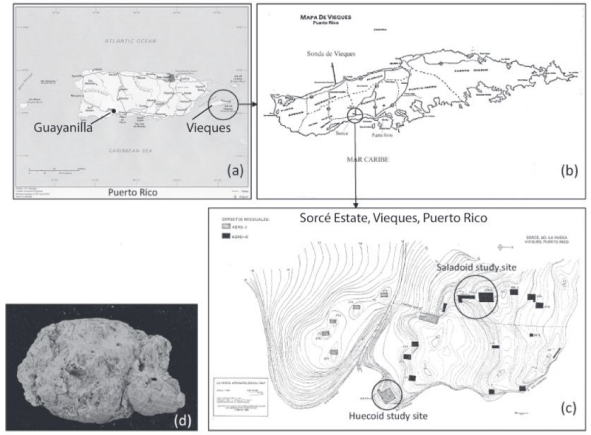


Figure 1. Location and obtainment of coprolites used in this study. Panels (a) and (b) show the sampling sites, located in Sorcé, Vieques, an island off the eastern coast of Puerto Rico. Panel (c) shows the Huecoid and Saladoid archaeological study sites (namely AGRO-I and AGRO-II, respectively). Panel (d) shows a coprolite extracted from these archeological sites.
doi:10.1371/journal.pone.0106833.g001

Cano et al. 2014 PLoS One

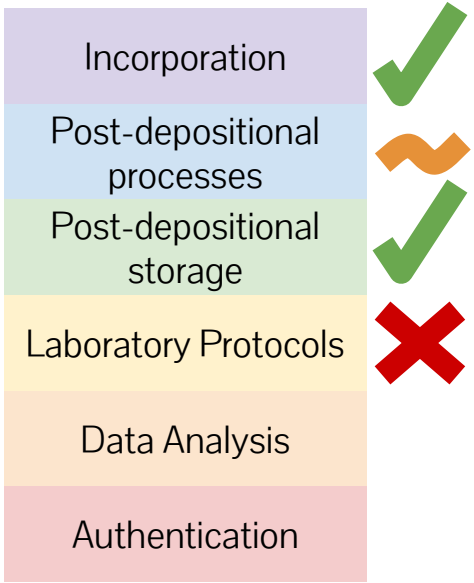
Case Study

- Incorporation
- Post-depositional processes
- Post-depositional storage
- Laboratory Protocols
- Data Analysis
- Authentication

Preventing contamination

Upon excavation, coprolites were individually stored in sterile sample bags until further use. Samples were individually processed inside a class II biosafety cabinet (BSC) used exclusively for ancient DNA analysis. Strategies used for the obtainment of reliable data were implemented [12]. The BSC was routinely cleaned with 70% ethanol and irradiated with UV for 30 min before and after use. Solutions used for DNA extractions were dispensed into single-use aliquots. Previously unused, sterile micropipettes were used; non-disposable equipment were sterilized by autoclave and baked overnight at >100°C. Baked and autoclaved stainless steel utensils were used to separate the inner and outer layers of the coprolites and were sterilized between samples using 70% ethanol. Previously published data comparing the microbial profiles detected in the inner and outer regions of the coprolites was used as an additional control [8].

Case Study



DNA extraction

First, the exterior layers of the coprolites were aseptically removed in a class II biosafety hood used strictly for aDNA, and subjected to controls to avoid extant DNA contamination. DNA was iso-

...

DNA sequencing

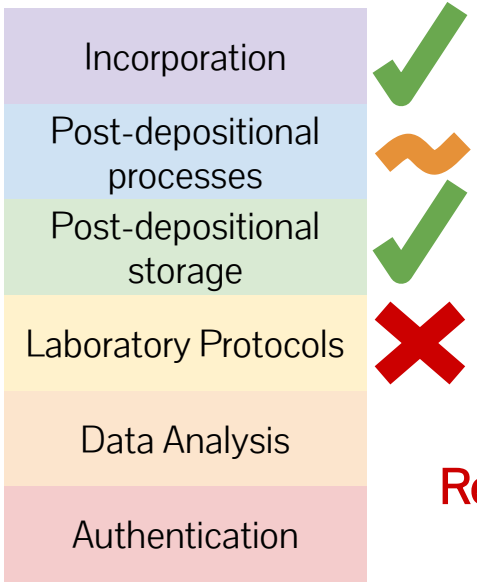
The two DNA composites were sequenced in a separate laboratory (MR DNA Research Lab, Shallowater, TX) using a non-targeted metagenomic approach. The library was prepared using

...

achieve the recommended DNA input of 50ng at a concentration of 2.5ng/uL. Subsequently, the samples underwent simultaneous fragmentation and addition of adapter sequences. These

Hood alone not enough to prevent air-borne DNA!
Sequencing in commercial lab OK if library built in **clean lab**

Case Study



S1 Table. DNA and library concentrations.

Sample	DNA concentration (ng/uL)[1]	DNA library concentration (ng/uL)	Average library size (bp)
MixS1	85.20	7.70	1230
MixH1	83.80	12.30	1550

[1]Indicates whole genome amplified and purified DNA. Reported size includes sequencing adaptors, measuring approximately 120bp each.

Reads so long couldn't sequence (not expected in hot/wet conditions)!
Fragment and add adapters same time?
Can't authenticate aDNA with short read lengths as all the same!

Case Study

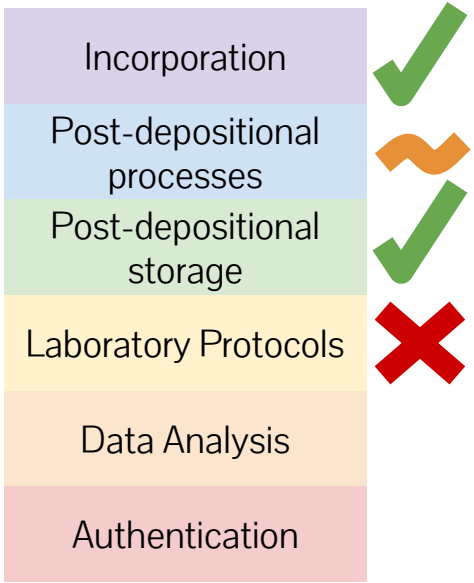


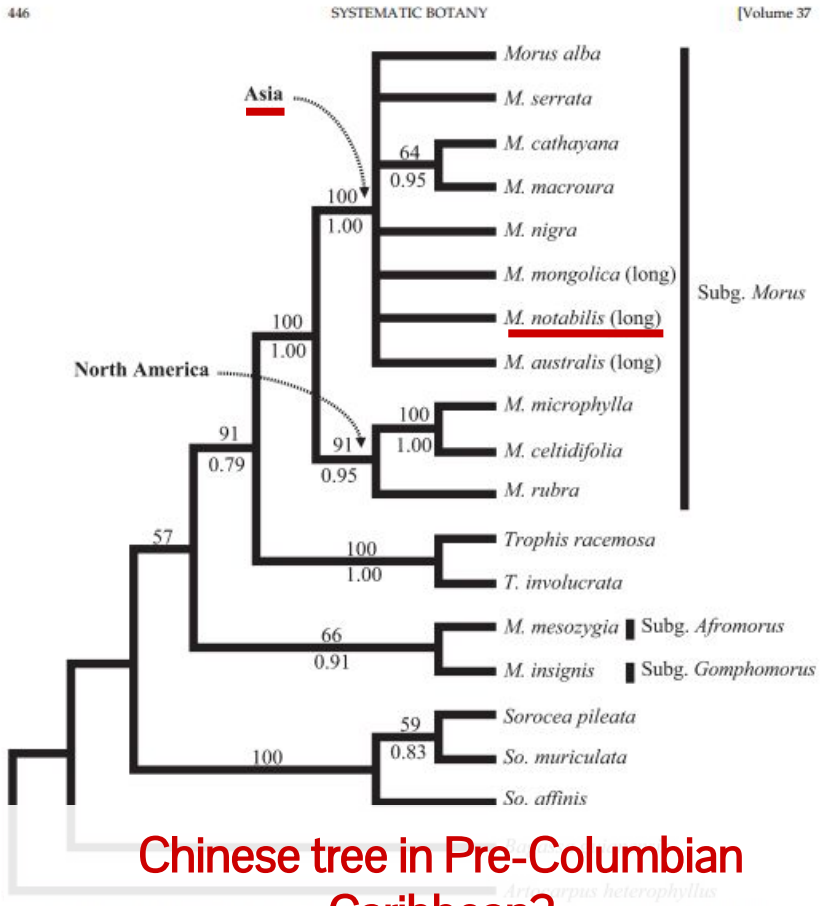
Table 3. Viral sequences detected as components of the Saladoid and Huecoid diets.

	Proviral sequence detected	Corresponding gene identification	Virus host organism	Previous osteological, archaeofaunal and botanical results
Vertebrates[a]	Moloney murine leukemia virus	reverse transcriptase	*Mus musculus (rodent)*	*Isolobodon portoricensis* (hutia)
	Murine endogenous retrovirus	retrotransposable element ORF2	*Mus musculus*	*Heteropsomys insulans* (spiny rat)
	Bat endogenous retrovirus	polymerase polyprotein	*Pteropus alecto*	ND
	Tiger frog virus	thymidylate synthase	Frog	ND
	Grouper iridovirus	unknown protein	*Epinephelinae* (fish)	*Epinephelinae*
	Avian pox virus	HAL3 domain	Fowl, parrots, and other birds	*Pelicanus occidentalis* (among many other birds)
	Monkey endogenous retrovirus	H element-like protein	Marmosets	ND
Invertebrates[a]	ND[c]			*Cardisoma guanhumi* (crab)
	ND			*Gecarcinus spp* (crab)
	ND			*Donax denticulatus* (oyster)
	ND			*Cittarium pica* (gastropod)
	ND			*Pelurodonte caracolla* (land snail)
?	Mulberry endogenous retrovirus	polymerase polyprotein, transposon TNT 1–94	*Morus notabilis* (berry fruit)	ND
	ND			*Psidium guajava* (guava fruit)
Plants[b]	ND			*Zea mays* (corn)[d]
	ND			*Ipomoea batatas* (sweet potato) [d]
	ND			*Cassava spp* (yucca)
	ND			*Capsicum spp* (pepper) [d]

Case Study

Incorporation	✓
Post-depositional processes	~
Post-depositional storage	✓
Laboratory Protocols	✗
Data Analysis	✗
Authentication	

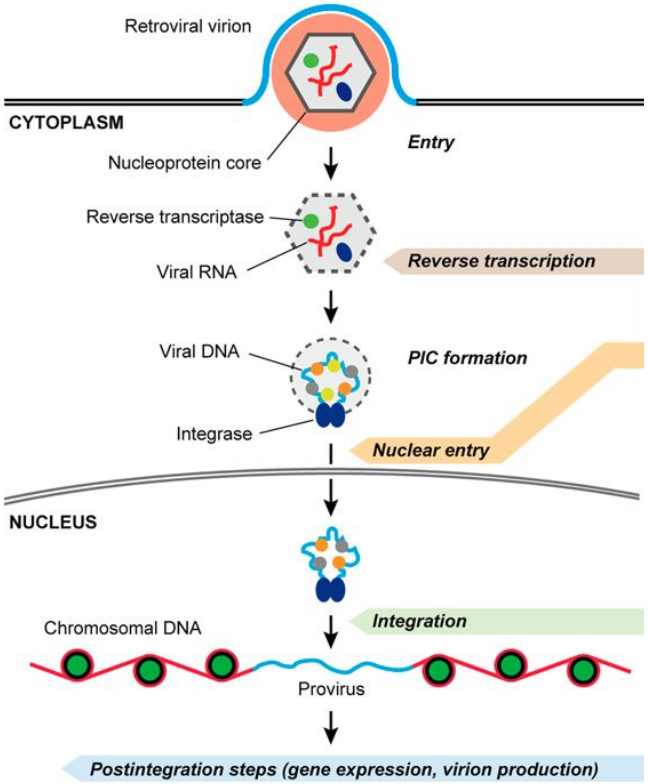
Nepal & Ferguson (2012) Systematic botany



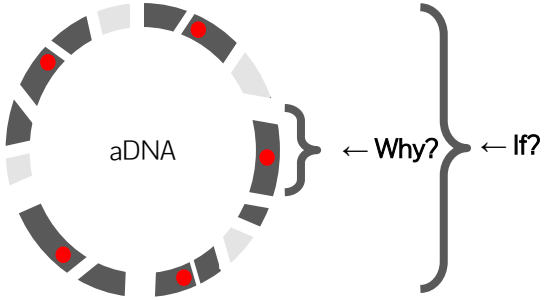
Chinese tree in Pre-Columbian Caribbean?

Case Study

Incorporation	✓
Post-depositional processes	~
Post-depositional storage	✓
Laboratory Protocols	✗
Data Analysis	✗
Authentication	

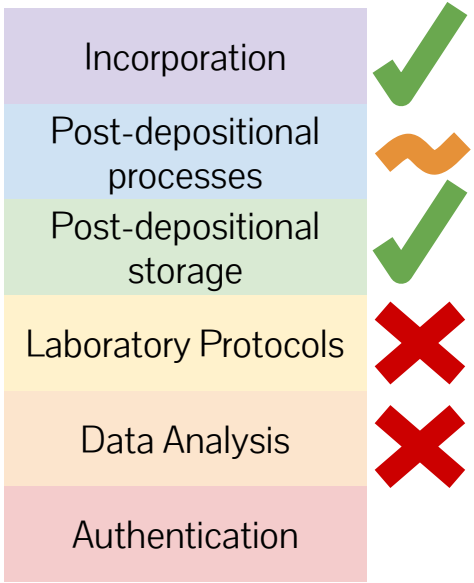


Suzuki et al. (2012) Frontiers in Microbiology



Case Study

DNA degradation patterns. One such example can be seen in **Figure 1A**, generated using Enterobacteria phage PhiX 174 as reference. Data used for this analysis is found in the Supplementary Materials included in this publication. It's important to point out that although a small segment of PhiX 174 DNA is used during Illumina sequencing, **Figure 1B** shows we detected the phage's entire genome, and thus this DNA could not be a result of PhiX controls remnant from sequencing.



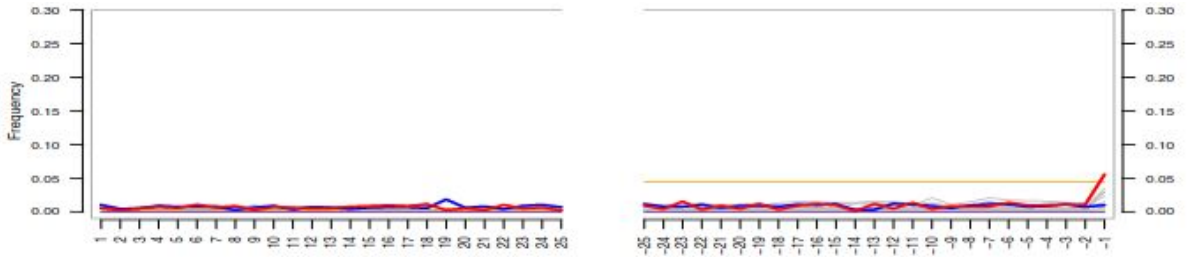
Case Study

Incorporation	✓
Post-depositional processes	~
Post-depositional storage	✓
Laboratory Protocols	✗
Data Analysis	✗
Authentication	✗ ✗

DNA degradation patterns. One such example can be seen in **Figure 1A**, generated using Enterobacteria phage PhiX 174 as reference. Data used for this analysis is found in the Supplementary Materials included in this publication. It's important to point out that although a small segment of PhiX 174 DNA is used during Illumina sequencing, **Figure 1B** shows we detected the phage's entire genome, and thus this DNA could not be a result of PhiX controls remnant from sequencing.

PhiX Control v3 is a reliable, adapter-ligated library used as a control for Illumina sequencing runs. The library is derived from the small, well-characterized PhiX genome, offering several benefits for sequencing and alignment.

illumina.com



Rivera-Perez et al. (2017) Unpublished [ResearchGate]

Case Study: Rivera-Perez et al. 2015

Summary

- Not in clean laboratory - free floating DNA
- Libraries built in commercial facility - 1000bp reads?
- Chinese tree in Pre-Columbian America?
- Retrovirus DNA vs rest of the genome?
- No damage pattern - use of known reagent as reference?

Suggestion

LETTER

doi:10.1038/nature21674

Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus

Laura S. Weyrich[1], Sebastian Duchene[2], Julien Soubrier[1], Luis Arriola[1], Bastien Llamas[1], James Breen[1], Alan G. Morris[3], Kurt W. Alt[4,5,6,7], David Caramelli[8], Veit Dresely[5,6], Milly Farrell[9], Andrew G. Farrer[1], Michael Francken[10], Neville Gully[11], Wolfgang Haak[1], Karen Hardy[12,13], Katerina Harvati[10], Petra Held[14], Edward C. Holmes[2], John Kaidonis[11], Carles Lalueza-Fox[15], Marco de la Rasilla[16], Antonio Rosas[17], Patrick Semal[18], Arkadiusz Soltysiak[19], Grant Townsend[11], Donatella Usai[20], Joachim Wahl[21], Daniel H. Huson[22], Keith Dobney[23,24,25] & Alan Cooper[1]

20 APRIL 2017 | VOL 544 | NATURE | 357

Table 1 | Dietary information preserved in calculus

Scientific name	Common name of probable source	Hominid pathogen or medicinal use
Zymoseptoria tritici	Plant (wheat) pathogen	
Phaeosphaeria nodorum	Plant (wheat) pathogen	
Penicillium rubens	Food fungus	MU
Myceliophthora thermophila	Cellulose fungus	
Coprinopsis cinerea	Edible mushroom (grey shag)	
Schizophyllum commune	Edible mushroom (split gill)	
Malassezia globosa	Human fungal commensal	
Enterocytozoon bieneusi	Intracellular parasite (microsporidia)	HP
Ovis aries	Sheep (wild mouflon)	
Ceratotherium simum	White rhinoceros (woolly rhinoceros)	
*Ixodes scapularis**	Tick	
Physcomitrella patens	Moss	
Pinus koraiensis	Pine tree	
Populus trichocarpa	Poplar tree	MU
Total eukaryotic reads		

Not all is lost!

Scaling up

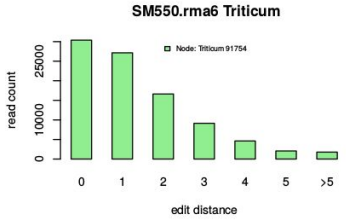
How to quickly identify and **authenticate** large amounts of data?

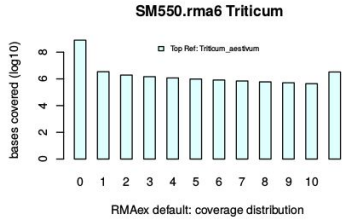
High-throughput pathogen detection in ancient metagenomic data

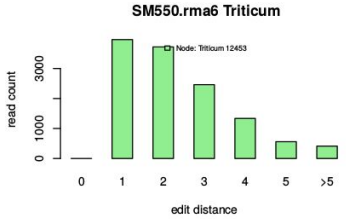
Felix M Key [1,*], Ron Huebler [1], Christina Warinner [1], Kirsten Bos [1], Wolfgang Haak [1], Johannes Krause [1], Alexander Herbig [1]

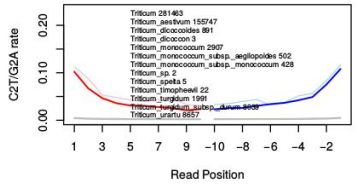
[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Analyses of pathogens associated with ancient human remains provide insights into the human disease burden of past populations. For a long time, such analyses were limited to characteristic skeletal lesions that can be caused by only a few pathogens. Today, the increasing availability of metagenomic data obtained from ancient human remains offers a new possibility for the detection of blood borne and oral pathogens, providing direct evidence for their presence in human remains. The genomic information from ancient pathogens provides insights into their evolution and adaptation to the human host. However, existing manual approaches for pathogen detection in ancient metagenomic data are laborious and error prone. Here, we present an in-silico pipeline to automate the detection of relevant bacterial species in ancient metagenomic data. We exploit several signatures indicative of ancient pathogen presence in shallow sequenced DNA libraries. We assess the sensitivity and specificity of our pipeline and optimize parameters using simulated data sets for different source material, and with varying endogenous DNA content for over thirty relevant pathogens and commensals. Lastly, we deploy our pipeline to a collection of over 2500 ancient metagenomic libraries from human remains to infer the presence of pathogens in more than ten thousand years of human evolution.





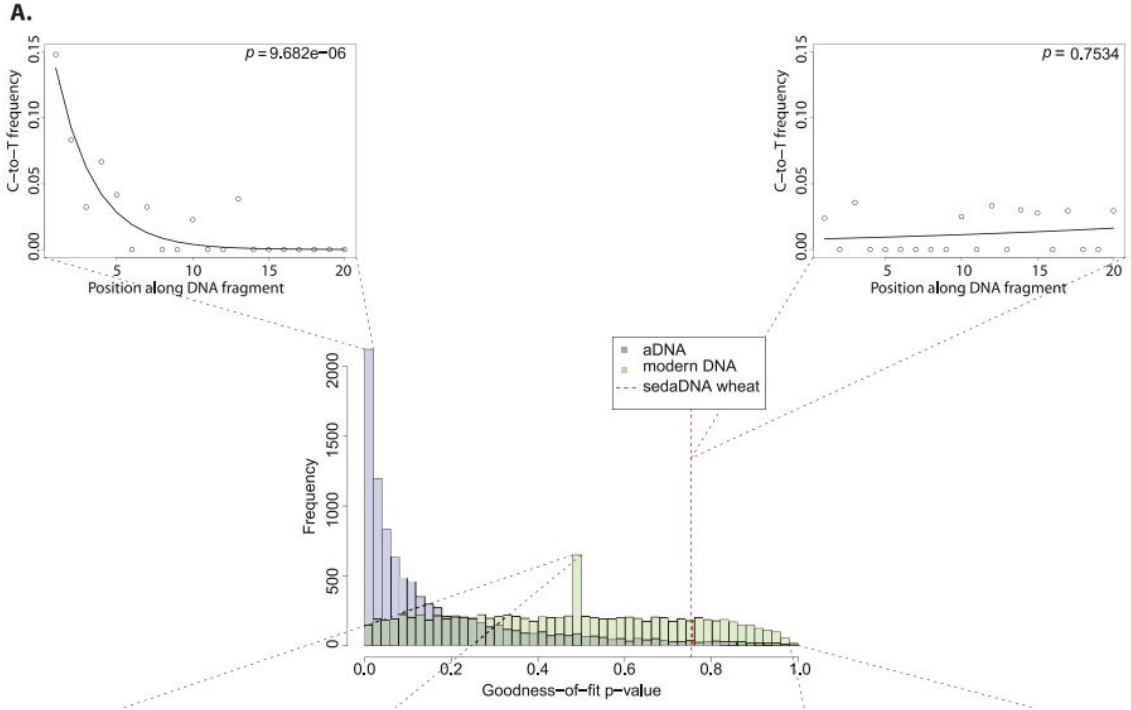




Node	Triticum
rdNodeRef	Triticum_aestivum
rdTotAlignments	330747
nonDup	327373
readDis	0.224
nonStacked	264312
C>T_1	0.1022
G>A_-1	0.108
mean length (sd)	49 (15.775)

Scaling up

How to authenticate damage patterns with **few reads?**



Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data

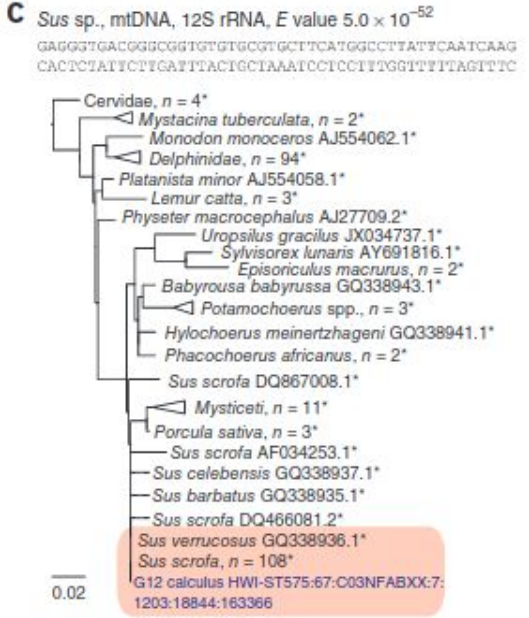
Clemens L Weiß[1], Michael Dannemann[2], Kay Prüfer[2], Hernán A Burbano[1]*

[1]Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany; [2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Weiß *et al.* eLife 2015;4:e10005. DOI: 10.7554/eLife.10005

Scaling up

How to authenticate identification **specificity**?



Pathogens and host immunity in the ancient human oral cavity

Christina Warinner[1,2], João F Matias Rodrigues[3,4], Rounak Vyas[3,4], Christian Trachsel[5], Natallia Shved[1], Jonas Grossmann[5], Anita Radini[6,7], Y Hancock[8], Raul Y Tito[2], Sarah Fiddyment[6], Camilla Speller[6], Jessica Hendy[6], Sophy Charlton[6], Hans Ulrich Luder[9], Domingo C Salazar-García[10–12], Elisabeth Eppler[13,14], Roger Seiler[1], Lars H Hansen[15,16], José Alfredo Samaniego Castruita[17], Simon Barkow-Oesterreicher[5], Kai Yik Teoh[6], Christian D Kelstrup[18], Jesper V Olsen[18], Paolo Nanni[5], Toshihisa Kawai[19,20], Eske Willerslev[17], Christian von Mering[3,4], Cecil M Lewis Jr[2], Matthew J Collins[6], M Thomas P Gilbert[17,21], Frank Rühli[1,22] & Enrico Cappellini[17,22]

Conclusions

- Dietary metagenomes has potential to complement other methods by providing high taxonomic resolution.

- But as fledgling field, need to be cautious:
  - Respect nature of sample and limits of data
  - Generation in clean lab
  - Checks for damage and fragmentation
  - Checks for reliability of taxonomic assignment

- Methods are currently in the pipeline to improve **authentication**

General aDNA Resources

Original Guidelines (and criticisms)

Cooper, A. & Poinar, H.N., 2000. Ancient DNA: Do It Right or Not at All. *Science*, 289(5482), pp.1139–1139. Available at: http://www.sciencemag.org/content/289/5482/1139.2.short.

Pääbo, S. et al., 2004. Genetic Analyses from Ancient DNA. *Annual review of genetics*, 38(1), pp.645–679. Available at: http://dx.doi.org/10.1146/annurev.genet.37.110801.143214.

Gilbert, M.T.P. et al., 2005. Assessing ancient DNA studies. *Trends in ecology & evolution*, 20(10), pp.541–544. Available at: http://dx.doi.org/10.1016/j.tree.2005.07.005.

NGS Era Lab Guidelines

Llamas, B. et al., 2017. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research*, 3(1), pp.1–14. Available at: http://dx.doi.org/10.1080/20548923.2016.1258824.

Knapp, M. & Hofreiter, M., 2010. Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. Genes, 1(2), pp.227–243. Available at: http://dx.doi.org/10.3390/genes1020227.

NGS Era Data Analysis Authentication

Warinner, C. et al., 2017. A Robust Framework for Microbial Archaeology. *Annual review of genomics and human genetics*. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-genom-091416-035526.

Key, F.M. et al., 2017. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends in genetics: TIG*, 33(8), pp.508–520. Available at: http://www.sciencedirect.com/science/article/pii/S0168952517300860.