Paleogenetics, Part 3

CHARLES J VELLA, PHD, 2022

Last Updates of 2021

Richard Leakey (1944-2022): trailblazing conservationist and fossil hunter, dies at 77

Richard Leakey: Turkana Basin mother lode

- Son of Louis and Mary Leakey: six when he found his first fossil, the jaw of an extinct giant pig at Olduvai
- Cheated death many times—a childhood skull fracture, kidney and liver failures that required transplants, public beatings, and a plane crash
- Flew over sedimentary rock on the shores of Lake Turkana. Leakey had a hunch about the basin below, and over the next 20 years, he and his "Hominid Gang" of Kenyan fossil hunters discovered many hominins. Made East Africa the central playing field for the study of human evolution
- Acheulean stone tools at 1.9 MA; Turkana Boy in 1984 – most complete fossil skeleton;
- director of the National Museum of Kenya & Kenya Wildlife Services – ban on ivory trade; conservation; then politics

E.O. Wilson, a Pioneer of Evolutionary Biology, Dies at 92

Carl Zimmer , 2021, NYT

Edward Osborne Wilson: Famed naturalist, 'Darwin's natural heir,' dies at 92

- "The Theory of Island Biogeography," in 1967: bigger islands could support more species than smaller ones; islands closer to the mainland would receive an influx of more species; destruction of habitats created island-like fragments leading to extinctions; foundation of conservation biology
- "Insect Societies", 1971; Inclusive fitness: a gene can also increase its evolutionary success by indirectly promoting the reproduction and survival of other individuals who also carry that gene; biology of altruism
- "Sociobiology: The New Synthesis" in 1975; got in trouble for extending sociobiology to humans; evolution-based account of human nature; genes determine human behavior. Critics = biological determinism; "Racist Wilson you can't hide, we charge you with genocide!"; but animal behavior studies today are "95 percent sociobiology,"

E. O. Wilson

- "On Human Nature" won the Pulitzer Prize for general nonfiction in 1979
- "The Ants," which Dr. Wilson wrote with his longtime colleague Bert Hölldobler, won him his second Pulitzer, in 1991; discovered how ants communicate through pheromones
- "The Diversity of Life." in 1992
- In 2008 he unveiled the Encyclopedia of Life website
- "The Social Conquest of Earth", 2012: rejects inclusive fitness; Richard Dawkins severely attacked this
- "Half-Earth: Our Planet's Fight for Life," 2016; we have catalogued less than 10% of life on earth; argued that the only way to avoid a mass extinction would be to leave half the earth wild.

AI analysis of dentition of Last Common Ancestor

- It is likely that the LCA had dental proportions more similar to gorillas than to humans
- Tesla Monson and her UCB team trained machine learning algorithms to identify different species of living apes on tooth data alone. Once the artificial intelligence was adequately trained, the scientists also programmed it to be able to identify extinct apes from fossil data. This culminated in a quantitively comparative analysis of the fossils to the living apes. These comparisons revealed that many of the oldest hominoid fossils, many dated to the Miocene epoch between 23 to 5 million years ago, look much more like gorillas in terms of their dental proportions than they do chimpanzees.
- The DNA difference between gorillas and humans is only about 1.6 percent; the DNA difference between humans and chimps is about 1.2 percent.

Baby teeth from 3 Neandertals

- Samples from Krapina, Croatia: dated 130 Ka
- We know that Neanderthals had **baby teeth,** just like humans.
- These teeth can reveal how quickly Neanderthal children began eating **solid food** — a developmental milestone that correlates with a period of accelerated brain growth.
- Human babies typically get their first teeth — incisors — in the front of their mouths between **6-10 months** old. Molars and canines grow more slowly.
- The researchers found that the rate of tooth growth in Neanderthals was similar to human babies, and in some cases even faster.

Neandertal teeth

- Past analysis suggests Neanderthal infants may have been able to start processing food besides their mother's milk at as early as ***4* months** old. The research at Krapina came to similar conclusions.
- Study co-author B. Holly Smith noted that faster growth can indicate higher rates of mortality in a species. As the risk of death before reproduction increases, pressure to grow, develop and reproduce earlier increases
- If teeth appeared earlier in Neanderthals, it may mean their brains and bodies may have developed — and aged — more quickly than humans, too.

Earliest Acheulean tools: 1.67 Ma in Algeria

- The oldest Acheulean lithic assemblage in North Africa: from the Oued Boucherit valley in Algeria, and dated to 1.67 Ma. This date is closely approximates the oldest Acheulean in eastern Africa (1.8 Ma).
- This work opens up two interesting possibilities: that of a rapid expansion of the Acheulean from the east to the north, or an almost contemporary and independent development of the Acheulean in both regions.
- A similar process was proposed in 2018 for the origin of the Olduvai Acheulean, accompanying the discovery of a set of tools for this technology in Ain Boucherit (Algeria) of 1.77 and 2.44 Ma.

Common basis for language and tool use

- Tool use and language are hallmarks of human evolution.
- Because of the similarity between the motor processes for tool use and those supporting language, it has been hypothesized that syntax and tool use may share brain resources.
- Using functional magnetic resonance imaging and multivariate pattern analysis, Thibault *et al*. found that small portions of the basal ganglia in the human brain act as common neural substrates for both tool use and syntax in language.
- In a behavioral experiment, they showed that learning a novel task that involves the use of a tool also improves performance in a complex language task.
- These results further support the hypothesis of a coevolution of tool use and language.

Simon Thibault et al., 2021

Language and stone tool making

- This indicates common neural resources for the two abilities.
- Indeed, learning transfer occurs across the two domains:
  - Tool-use motor training improves syntactic processing in language and,
  - reciprocally, linguistic training with syntactic structures improves tool use

Oviraptorid dinosaur

Oviraptorid dinosaur embryo laid 70 million years ago:
Living bird embryos are known to move into the best position, known as tucking behaviors, to help them hatch from their eggs. In essence, birds inherited these pre-hatching

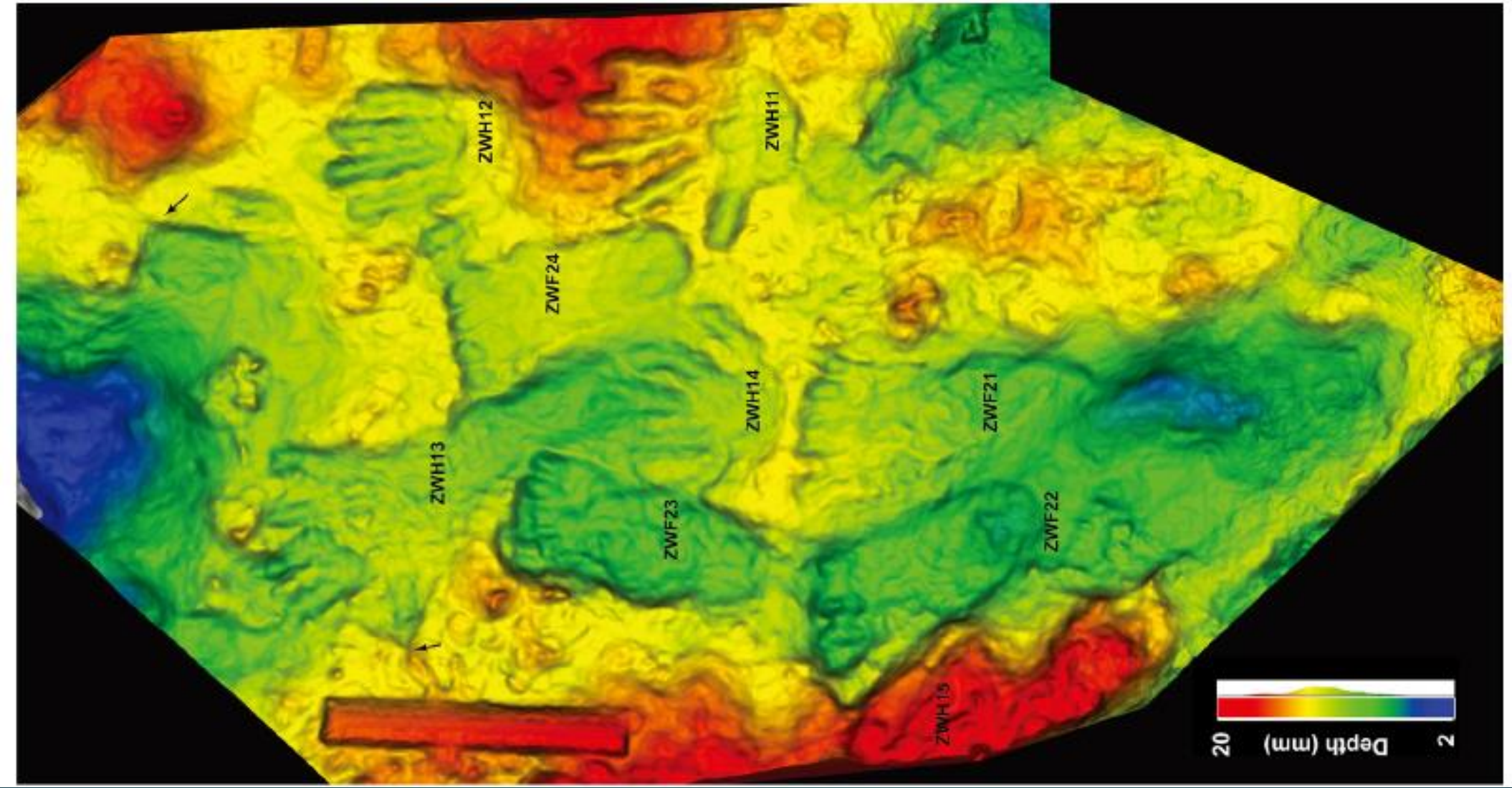
A reconstruction of Baby Yingliang, an oviraptorid dinosaur embryo laid 70 million years ago in what is now China. (Image credit: Shoulin Animation)

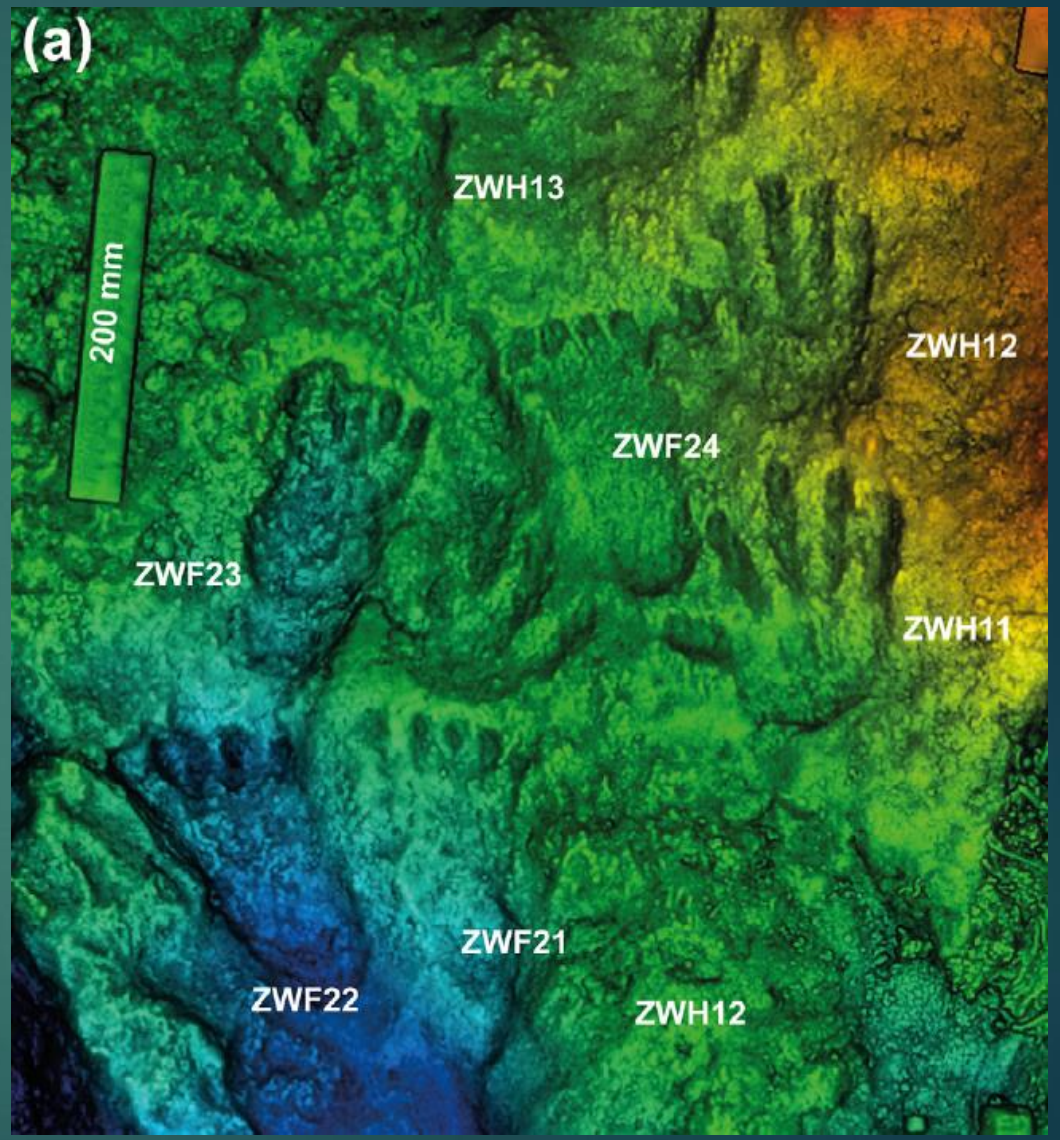
1.5 Ma *H. erectus* footprint: highly similar to MH foot

A pronounced arch, a big toe that lay parallel to the other toes, and a round heel, similar to human feet. Analysis of the prints revealed that they were created by stepping motions that were identical to those used by *Homo sapiens* — with the inner ball of the foot and the first two toes

Quesang, Tibet: 2 kids messing around

Oldest parietal art: N or D Child handprints in Quesang Cave, Tibet, dated to 226-169 kya




(a)
200 mm
ZWH13
ZWH12
ZWF24
ZWF23
ZWH11
ZWF21
ZWF22
ZWH12

A red-billed chough is an elusive species to hunt during the day. But its nighttime roosting habits could have made it easy prey for Neandertals to catch with their bare hands,

Predictable cloughs: Ns could plan

- This idea to role-play started with butchered bird bones. Piles of ancient tool- and tooth-nicked choughs bones have been found in the same caves that Neandertals frequented, evidence suggesting that the ancient hominids chowed down on the birds. But catching choughs is tricky. During the day, they fly far to feed on invertebrates, seeds and fruits. At night though, their behavior practically turns them into sitting ducks. The birds roost in groups and often return to the same spot, even if they've been disturbed or preyed on there before.
- Teams of two to 10 researchers silently snuck into caves and other spots across Spain, where the birds roost to see how many choughs they could catch. Using flashes of light from flashlights to resemble fire, the "Neandertals" dazzled and confused the choughs. The birds typically fled into dead-end areas of the caves, where they could be easily caught, often bare-handed. Hunting expeditions at 70 sites snared more than 5,500 birds in all.
- The regular catchment of choughs by Neandertals implies a deep knowledge of the ecology of this species, a previous planning for its obtaining, including procurement techniques, and the ability to plan and anticipate dietary needs for the future

Neandertal spear throwing ability

Why modern javelin throwers hurled Neandertal spears at hay bales

- Six javelin throwers who approached the physical strength of Neandertals. The weapon: Two replicas of a 300,000-year-old wooden spear, one of nine ancient hunting tools discovered at Germany's Schöningen coal mine. The test: Could Neandertals, the likely makers of the Stone Age weapon, have hurled the spears at prey with any power, accuracy and distance?
- Athletes threw the two wooden spear replicas a total of 102 times at bales of hay, hitting bales five meters away 58 percent of the time. That figure fell to 25 percent for throws from 10 meters and 15 meters, and 17 percent for 20-meter throws.
- The results are the first measurements of the Schöningen projectiles' flight characteristics when hurled at a target.

Milks et al., 2019

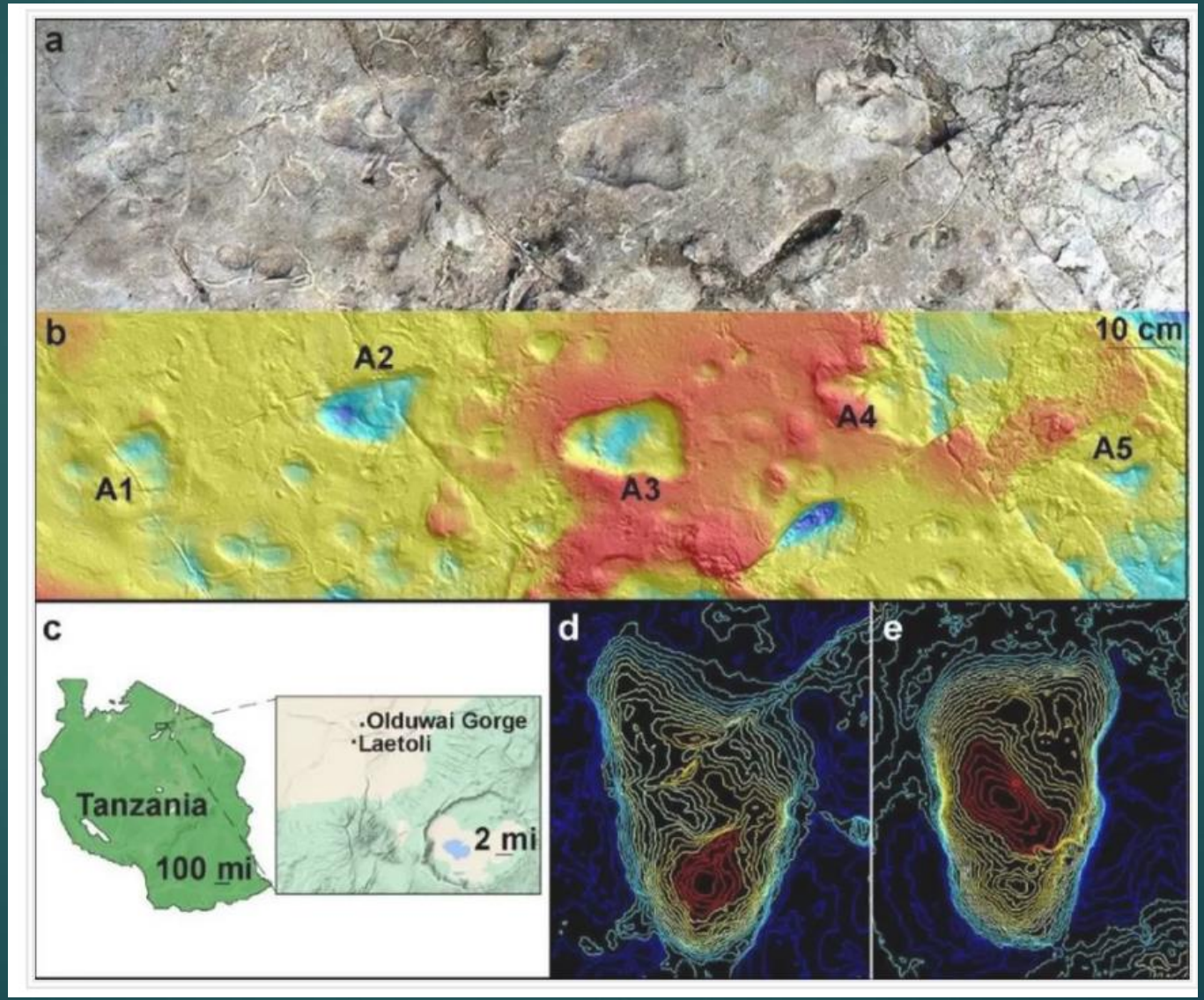
- Classic bipedal trackways discovered in 1978 at Laetoli site G, Tanzania and dated to 3.66 million years ago are widely accepted as the oldest unequivocal evidence of obligate bipedalism in the human lineage

New footprints at Laetoli

- 1978, Laetoli footprints, Mary Leakey: a set of about 70 ichnites (fossilized footprints) assigned to three individuals of *Australopithecus afarensis* that walked 3.66 Ma (millions of years ago) on a soil of volcanic ash. These footprints left a 27 meter trail (called site G) confirmed the totally bipedal walking of these hominin
- 26 years later, in 2014, another set of footprints was found at site S, just 150 meters from the previous one, consisting of 14 tracks left by two other individuals (13 of one and 1 from another), along 32 meters.
- These are virtually contemporaneous with those left by Australopithecus at Site G. Comparing footprint characteristics, it appears that the walkers were also *Australopithecus afarensis* . Many other animals left their tracks in Laetoli (mammals such as bovids, giraffes, equids, rhinos, etc. as well as birds and insects), up to 33 different sites that had been studied since the early 1970s.

McNutt, EJ et al. (2021).

New footprints at Laetoli



New footprints at Laetoli

New footprints at Laetoli: but from a different hominin

- 2 years before the first discovery, in 1976 a path of ichnites had been excavated in another of those sites, the A, among which there were 5 in a row that had also raised suspicions about a possible hominin authorship. First theory was of bears standing upright. Then discovery of site G and A was forgotten.
- **What hominin was it?**
- Proposed that the footprints of A were left by hominins other than *Australopithecus afarensis*. For example, while the *afarensis* at site G measured 111-116 cm, and those at site S are161-168 cm, the estimated height for hominins at site A is 101-104 cm.
- This conclusion would indicate the coexistence of more than one hominin species at the site almost 3.7 million years ago.

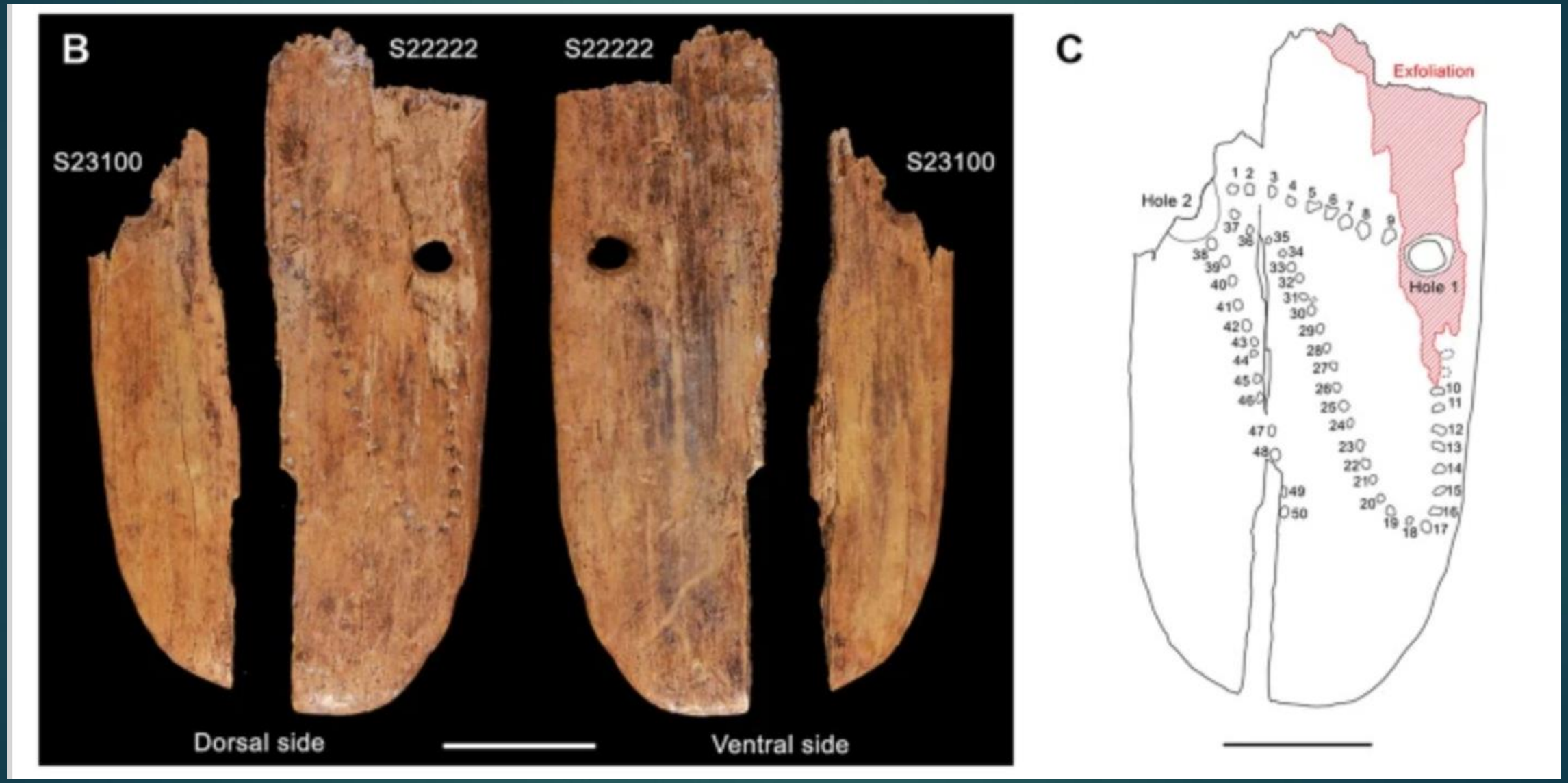
New footprints at Laetoli

- The MRD skull of *Au. anamensis* is 3.8 Ma, and is contemporary with some specimens associated with *afarensis*
- Could *Australopithecus anamensis be* the author of the footprints at Site A?
- Jeremy DeSilva opines that “if *A. afarensis* is a descendant of *A. anamensis*, I would expect their feet to be more similar, even to the point that they are indistinguishable. But we won't know until we find new fossils at Laetoli (which we hope to do this summer). '

41,500-year-old ivory pendant may be oldest human-decorated jewelry in Eurasia at Bacho Kiro

Sahra Talamo, et al., 2021

Pendant


B
S22222
S22222
S23100
S23100
Dorsal side
Ventral side
C
Exfoliation
Hole 2
Hole 1

Awl from same cave

The direct radiocarbon date makes the Stajnia ornate pendant (41,730–41,340 cal BP) the earliest punctate ivory object known to date to the Early Upper Palaeolithic record in Eurasia

for La Chapelle-aux-Saints. The "Old Man of La Chapelle“

New diagnosis for La Chapelle-aux-Saints. The "Old Man of La Chapelle“

- Thought to be in his late 50s or 60s when he died about 50,000 years ago, had advanced osteoarthritis in his spinal column and hip joint; not all the changes in the bones could be explained by the wear and tear of osteoarthritis.
- New study: diagnosis of brucellosis. Brucellosis, also known as 'Mediterranean Fever,' is a contagious disease that develops from exposure to a bacteria called Brucella. Acquire the disease through direct contact with infected animals, by eating or drinking contaminated animal products. Most cases are caused by unpasteurized milk or cheese from infected goats or sheep.
- A zoonotic disease -- illnesses that are transmitted from animals to humans; i.e. Covid

Martin Haeusler, et al., 2019

Brucellosis in Neandertals

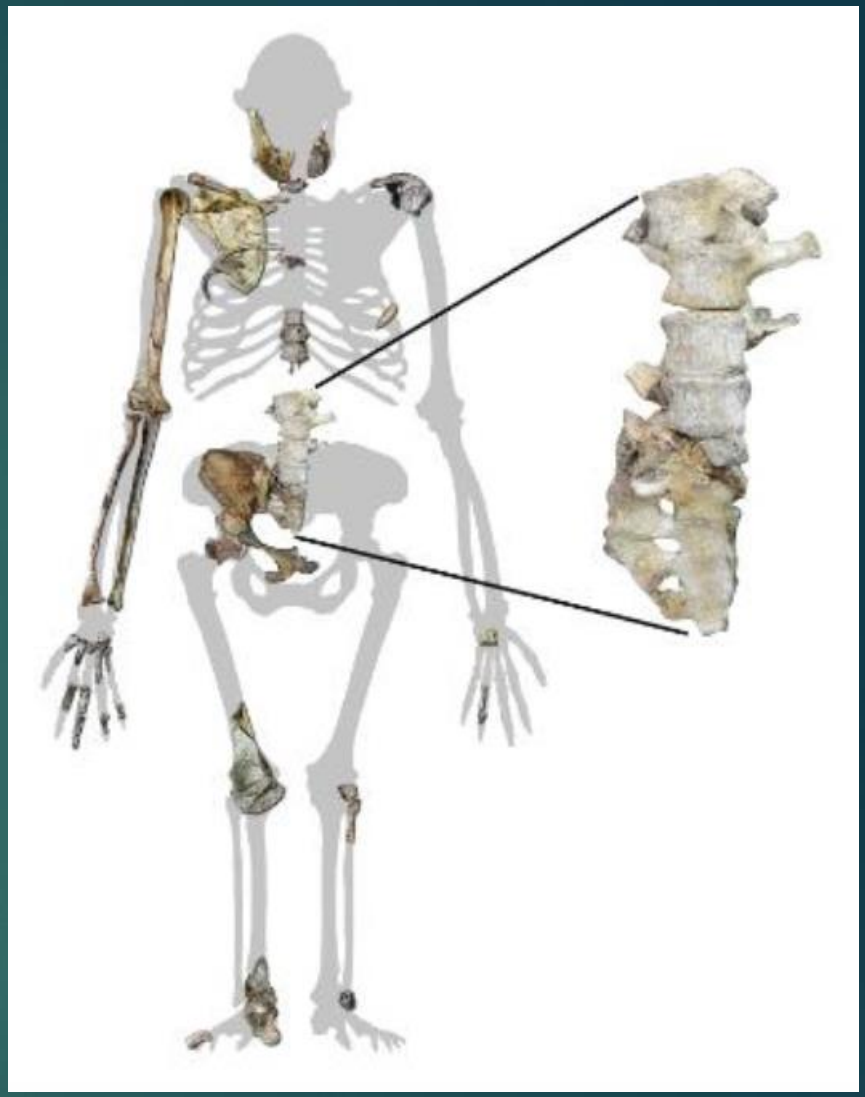
- Long-term problems resulting from the disease are variable but can include arthritis pain, back pain, inflammation of the testes -- which can lead to infertility -- and inflammation of the heart valves known as endocarditis

- Neanderthal man likely caught the disease from butchering or cooking an animal that had been hunted as prey. Possible sources include wild sheep, goats, wild cattle, bison, reindeer, hares and marmots.

A. *sediba*

Australopithecus sediba were able to use their upper limbs to climb and swing like apes.

Scott A Williams, et al., 2021



Near-complete lower back of Malapa Hominin 2 (MH2)

Discovery of near-complete lower back of Malapa Hominin 2 (MH2)

A. sediba: walk like a man, swing like a chimp

- Analysis of 2 Ma A. sediba's newly found lower back bones: able to walk the ground on two legs and use their upper limbs to climb and swing like apes sediba had only five lumbar vertebrae like humans, .
- Presence of lordosis (the inward curve of the lumbar spine in sediba's spine was more extreme than any other Australopithecus yet discovered -- that kind of spine curvature is typically seen in modern humans and demonstrates strong adaptations to bipedalism.
- There are other features, such as the large and upward oriented transverse processes, that suggest powerful trunk musculature, perhaps for arboreal behaviors.
- The study concluded that Australopithecus sediba was a transitional form of ancient human relative and its spine is clearly intermediate in shape between those of modern humans and great apes -- meaning that the species would have possessed both human and ape-like traits in its movements.

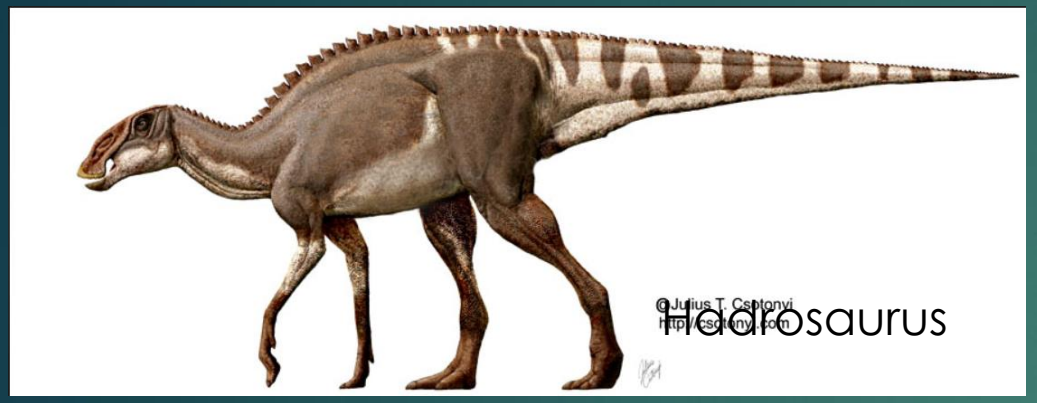
Ancient Tusk 150 Miles From Land, 10,000 Feet Deep?

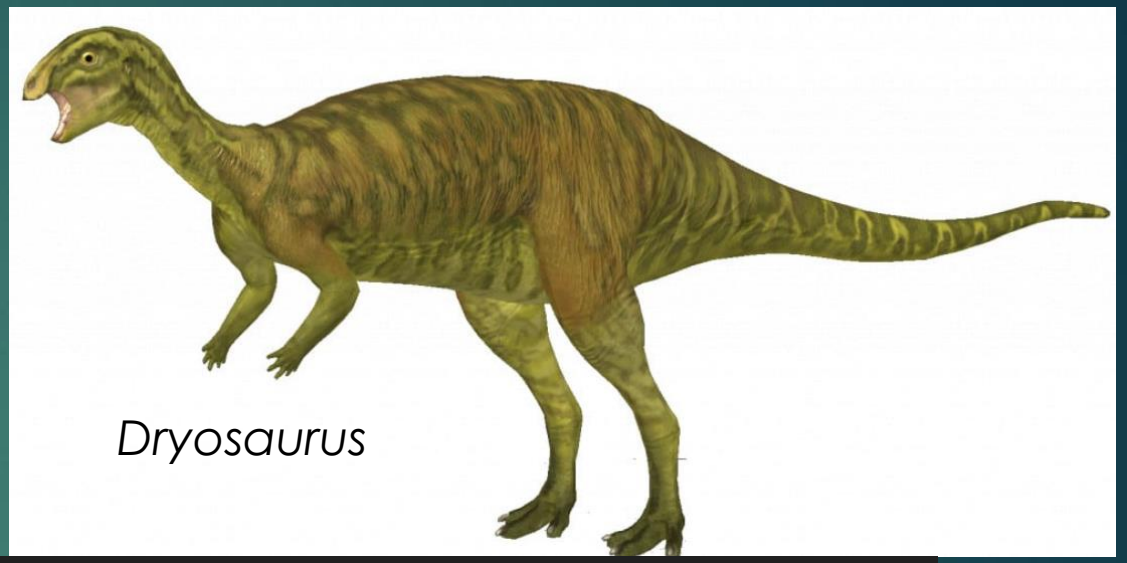
- A young female mammoth was wandering long ago near what would become the Central Coast of California, when her life came to an untimely end.
- Although she died on land, her massive body found its way into the Pacific Ocean. Carried by currents, her remains drifted more than 150 miles from shore before settling 10,000 feet beneath the water's surface on the side of a seamount.
- Scientists from the Monterey Bay Aquarium Research Institute stumbled upon one of her tusks while using remotely operated vehicles to search for new deep-sea species off the coast of Monterey, Calif.
- Possible date of 200 Ka; very rare fossils

New Spike-tailed ankylosaur

New ankylosaur *Stegouros elengassen;* tail like Aztec sword

What is commonality of these Mesozoic dinosaurs?





Your average dinosaur

- Most non-avian dinosaurs had a mass of about 7,700 pounds. That's a pretty big animal, somewhere between modern-day rhinos and elephants; exemplified by duckbilled dinosaurs
- These dinosaurs suffered terrible mortality rates during their first year of life.
- Often were the food our favorite carnivores relied on.
- Dinosaurs were better protected from predators when they reached a certain size. Studies of dinosaur growth have indicated that hadrosaurs rapidly packed on the pounds as a defense against carnivores, so an adult hadrosaur represents the size threshold when a potential lunch became too much of a bother.

New Bone pieces from Denisova Cave

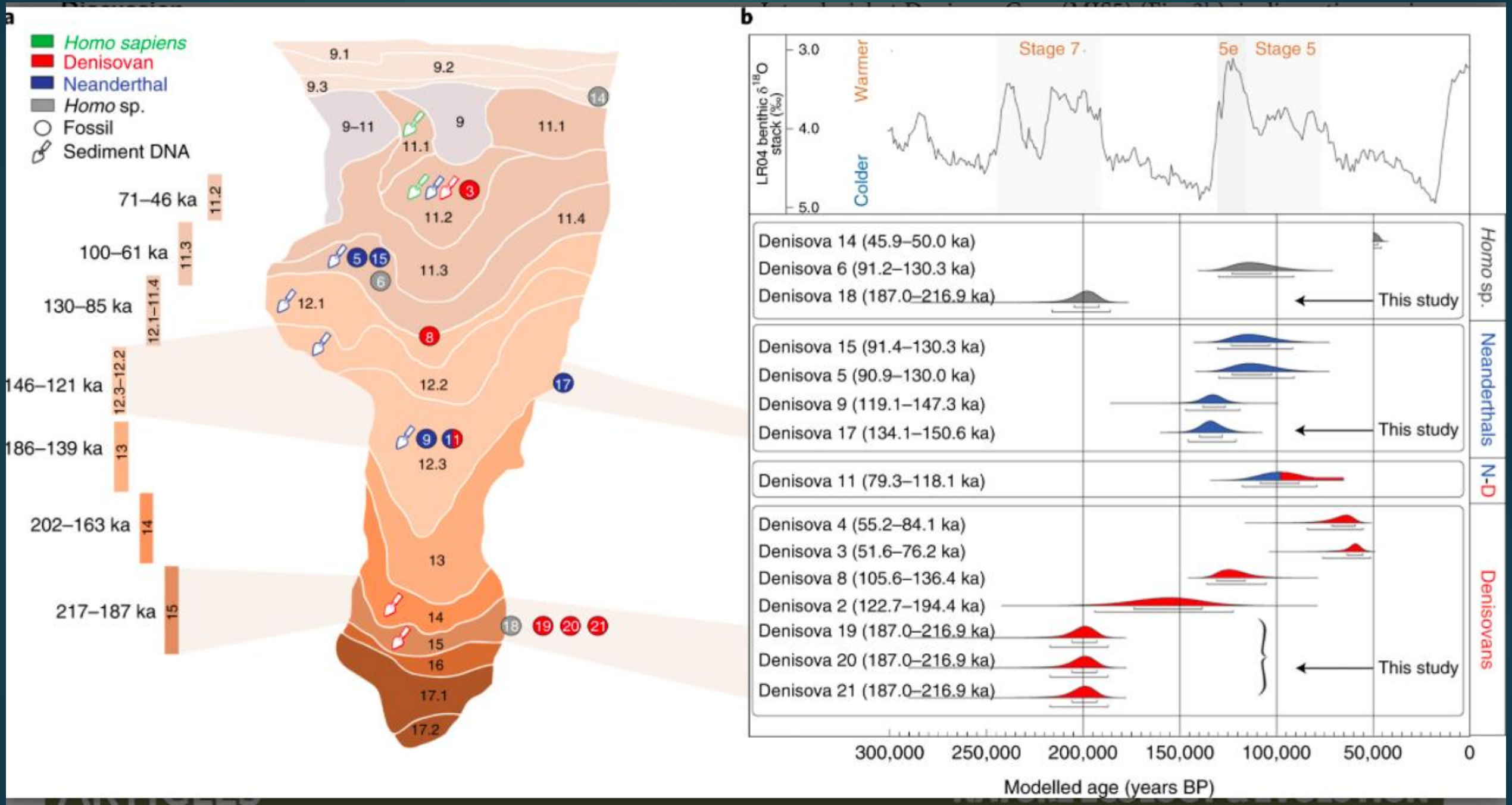
Earliest Denisovans: 200 Ka

- 5 prior fossil bones of Denisovans, dated 122 to 194 Ka
- Discovery of 3 new Denisovan bones dated to 200 Ka
- Protein analysis of 3,791 bone scraps from Denisova Cave
- Found five human bones. Four of these contained enough mtDNA to reveal their identity — one was Neanderthal, and the other three were Denisovan. Based on mtDNA similarities, two of these fossils may either come from one person or from related individuals.
- Also contained first contiguous stone artifacts (scapers) and animal remains; Denisovans may have fed on deer, gazelles, horses, bison and woolly rhinoceroses.

Samantha Brown, et al., 2021

New Denisovans

- Bones of carnivores such as wolves and wild dogs suggest Denisovans may have actively competed with these predators over prey and perhaps the cave itself.
- Used peptide mass fingerprinting (or ZooMS; Zooarchaeology by Mass Spectrometry) of bone collagen


a
Homo sapiens
Denisovan
Neanderthal
Homo sp.
Fossil
Sediment DNA
71–46 ka
100–61 ka
130–85 ka
146–121 ka
186–139 ka
202–163 ka
217–187 ka
11.2
11.3
12.1–11.4
12.3–12.2
13
14
15
9.1
9.2
9.3
9–11
9
11.1
11.4
12.1
12.2
12.3
13
14
15
16
17.1
17.2
b
LR04 benthic δ18O stack (‰)
3.0
4.0
5.0
Warmer
Colder
Stage 7
5e
Stage 5
Denisova 14 (45.9–50.0 ka)
Denisova 6 (91.2–130.3 ka)
Denisova 18 (187.0–216.9 ka)
This study
Homo sp.
Denisova 15 (91.4–130.3 ka)
Denisova 5 (90.9–130.0 ka)
Denisova 9 (119.1–147.3 ka)
Denisova 17 (134.1–150.6 ka)
This study
Neanderthals
Denisova 11 (79.3–118.1 ka)
N-D
Denisova 4 (55.2–84.1 ka)
Denisova 3 (51.6–76.2 ka)
Denisova 8 (105.6–136.4 ka)
Denisova 2 (122.7–194.4 ka)
Denisova 19 (187.0–216.9 ka)
Denisova 20 (187.0–216.9 ka)
Denisova 21 (187.0–216.9 ka)
This study
Denisovans
300,000
250,000
200,000
150,000
100,000
50,000
0
Modelled age (years BP)

Neandertals at Denisova Cave

- The presence of Neanderthals in the Altai was originally identified in Okladnikov Cave, a site located 50 km to the north of Denisova Cave, on the basis of mtDNA evidence.
- Further archaeological and genetic data suggest that Neanderthals were in Siberia on several separate occasions. They appeared at Denisova Cave (layer 12, East Chamber) at least ~150–130 ka.
- Five Neanderthal fossils have been found in the East Chamber so far, of which three are from layer 12 (Denisova 9, 11, 17) and two are from the overlying layer 11.4.
- A single sediment sample from layer 14 of the East Chamber yielded Neanderthal DNA

Ns and Ds at Denisova Cave

- The molecular age of the mtDNA of the newly identified Neanderthal (Denisova 17) to ~134 ka (94–177 ka).
- Phylogeny inferences show that the mtDNA of Denisova 17 is more distantly related to the mtDNAs of the two other Neanderthals from Denisova Cave, Denisova 5 and Denisova 15, who are more closely related to one another.
- In contrast, Denisova 11 mtDNA is more closely related to the mtDNAs of Neanderthals from western Eurasia and to other Siberian Neanderthals, such as those from Okladnikov Cave and Chagyrskaya Cave.
- Gene flow between Neanderthals and Denisovans provides additional indirect evidence of earlier interactions between the two groups.
- Analysis of the genome of a female Denisovan individual (Denisova 2), for example, has revealed that she had Neanderthal ancestry deriving from an introgression ~1,500 years before she lived, as early as 250–200 ka.

Ns and Ds at Denisova Cave

- Two other Denisovans from higher up the stratigraphic sequence (Denisova 8 and 3) also show Neanderthal introgression from two different Neanderthal populations.
- Although it is not possible to tell where these interbreeding events occurred, they provide evidence for potential cohabitation and frequent interactions between the two hominin groups from >200 ka (Denisova 2) until their disappearance from the Altai around 50 ka (Denisova 3).
- Neanderthal presence, while more pronounced during the Last Interglacial at Denisova Cave (MIS5), is discontinuous in the Altai region and may reflect occasional eastward migration of Neanderthal groups across large tracts of Eurasia.
- Gene flow between them most likely to have occurred in northeastern Eurasia.
- The Altai, in particular, appears to be an overlapping zone for both Denisovan and Neanderthal groups for over 150,000 years, witnessing and possibly facilitating population admixture as well as sustaining distinct hominin populations over this long period.

Ns and Ds at Denisova Cave

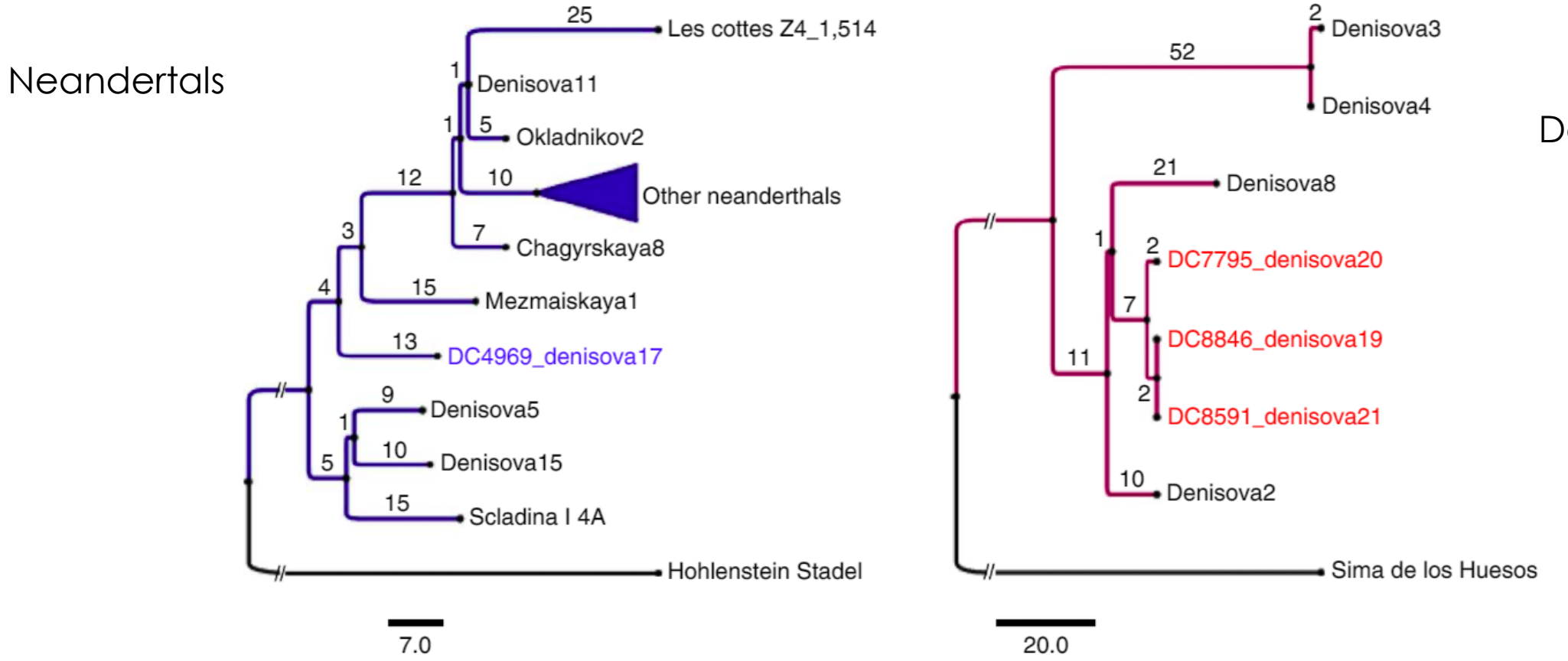


Fig. 2 | mtDNA maximum parsimony phylogenetic trees for the newly identified hominin bones. a, Neanderthal mtDNA parsimony phylogeny including

Ns and Ds at Denisova Cave

- The earliest Denisovan (Denisova 2) was estimated to date to 122–194 ka; the new fossils reported here were excavated in 2012–13.
- Layer 15 is the oldest archaeological layer of the East Chamber and is estimated to date to ~200 ka (205–192 ka)
- Inferred a divergence date for the mtDNAs of the three new and the four previously published Denisovans to ~229 during the Interglacial period MIS 7. Both the mtDNA age estimates and the established chronology for layer 15 render Denisova 19, 20 and 21, or their maternal relatives, the oldest Denisovans currently documented.
- The presence of individuals carrying Denisovan mtDNA in the lowermost archaeological layer 15 of the East Chamber offers an opportunity to consider the wider archaeological and subsistence context of this group of hominins. So far, this has not been possible because previous Denisovan fossils were either derived from layers impoverished in archaeological material or from layers where Neanderthal cohabitation could not be excluded.

Ns and Ds at Denisova Cave hunted during interglacial period

- Denisova 19, 20 and 21 date to the Penultimate Interglacial (MIS 7), a warm climatic period with comparable conditions to today that would have rendered the Altai a favorable location for hominin expansion and intensified occupation.
- During this phase, a mosaic of landscapes can be detected in the vicinity of the cave, including both broad-leaved forests and open steppe landscapes.
- Both traditional zooarchaeological and ZooMS analyses revealed that the inhabitants of the cave targeted a variety of taxa living in these environments, including interglacial forest and forest-steppe species, such as roe deer, Siberian red deer and giant deer, as well as species typical of more open country, such as horse, bison, woolly rhinoceros and Mongolian gazelle. Frequent anthropogenic impacts on bones, including splitting, burning and butchery cut-marks, confirm that these species were procured regularly.

Competition with wolves

- Humans appear not to have been the only occupants of Denisova Cave during this period, however. About a quarter of the macroscopically identified faunal assemblage from layer 15 comprised carnivore remains, predominantly Canis lupus and Cuon alpinus. This high proportion of carnivore taxa suggests that humans may have been actively competing with these predators over resources and perhaps the cave itself.
- Archaeologically, layer 15 (and layer 14) of the East Chamber contain the highest frequency of stone artifacts in the entire sequence of the cave. The lithic assemblage comprises discoidal, Levallois, and parallel cores to produce flakes using primary reduction techniques. Scrapers are the dominant tool type

Ns and Ds at Denisova Cave

- The closest parallel is the Acheulo–Yabrudian cultural complex from the Near East. The AYCC has been identified at several cave (mostly) and open-air sites such as Tabun, Qesem, Hayonim and Misliya, dating to between 400/350 and 250 ka. There are no Acheulean bifacial tools in the Denisova assemblage.
- The Denisovan DNA introgressed in present-day humans from Siberia and East Asia, and indigenous Americans share the highest similarity with the high quality genome of Denisova.
- However, the mtDNAs of the three older Denisovans we identified here—Denisova 19, 20 and 21—belong to a different mtDNA lineage from that of Denisova 3. Characterization of the nuclear DNA of these individuals is required to determine whether these early Denisovans are more closely related to the Denisovans that admixed with the ancestors of present-day humans living in island Southeast Asia and New Guinea.

Denisovans were highly adaptable

- The challenges encountered by Denisovans while living in extremely diverse and changing environments, from the Altai mountains to the high altitudes of the Tibetan Plateau, and possibly from north China to island Southeast Asia, would have required adaptation in novel ways to survive.

13,000 genes active in cerebral cortex

- Genetic code in brain more complex than previously thought.
- Some genes there encoded tens or even hundreds of different proteins.
- All of the mRNA in brain cells characterized: from 13 K genes, discovered 33 K different mRNAs
- More than 200 genes produce between 10 and 100 different mRNAs

Internet control of neuroscience animals

- A new study shows that researchers can remotely control the brain circuits of numerous animals simultaneously and independently through the internet.

- Wireless implantable devices and IoT could manipulate the brains of animals from anywhere around the world due to their minimalistic hardware, low setup cost, ease of use, and customizable versatility

- As long as researchers have internet access, they are able to trigger, customize, stop, validate, and store the outcomes of large experiments at any time and from anywhere in the world. They can remotely perform large-scale neuroscience experiments in animals deployed in multiple countries.

Vaccines

- The goal of a vaccine is to stimulate the adaptive immune system to create antibodies that precisely target that particular pathogen. The markers on the pathogen that the antibodies target are called antigens; these are disease-specific proteins.
- Traditional vaccines stimulate an antibody response by injecting either antigens, an attenuated (weakened) virus, an inactivated (dead) virus, or a recombinant antigen-encoding viral vector) into the body.
- These antigens and viruses are prepared and grown outside the body. Most vaccines against viral diseases are made from viruses grown in chicken eggs or mammalian cells.
- The process of collecting the viruses, adapting them to grow in the lab, and shipping them around the world can take months and is complex.

mRNA Vaccines: teach immune system to recognize a virus

- In contrast, messenger RNA (mRNA) vaccines introduce a short-lived[synthetically created fragment of the RNA sequence of a virus into the individual being vaccinated.
- No virus is needed to make a batch of an RNA vaccine.
- Only small quantities of virus are used for gene sequencing and for the vaccine testing. The DNA can be synthesized from an electronic sequence that can be sent across the world in an instant by computer.
- Currently it takes about a week to generate an experimental batch of an mRNA vaccine.

mRNA Vaccines

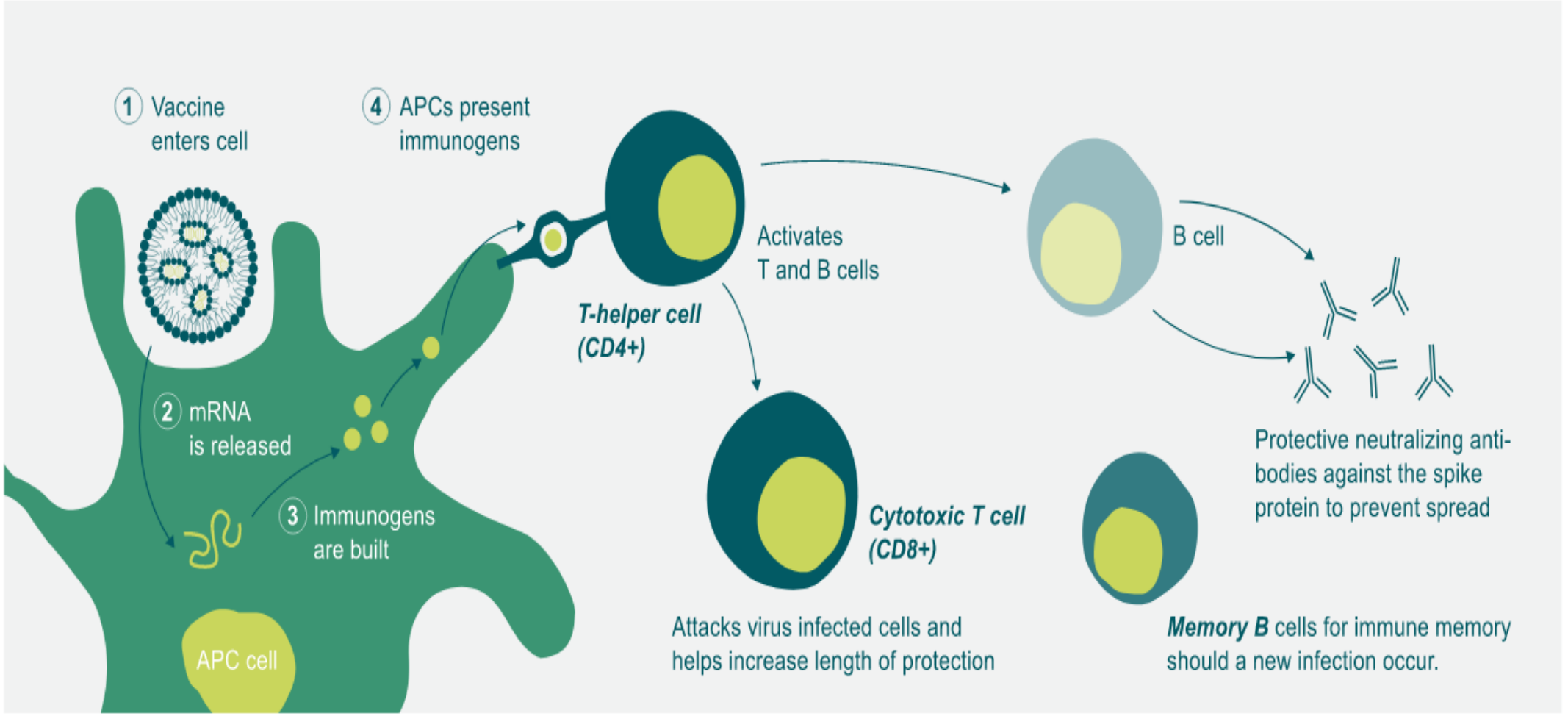
- An mRNA vaccine is a type of vaccine that uses a copy of a molecule called messenger RNA (mRNA) to produce an immune response.
- mRNA create protein molecules.
- The vaccine delivers molecules of antigen-encoding mRNA into immune cells, which use the designed mRNA as a template to build foreign protein that would normally be produced by a pathogen (such as a virus) or by a cancer cell.
- Deliver genetic code, not a virus, into the cell.
- Antigen is a molecule or molecular structure that can bind to a specific antibody or T-cell receptor. The presence of antigens in the body may trigger an immune response.

mRNA vaccines

- Briefly explained, mRNA vaccines deliver RNA to our body's cells that encode harmless fragments of a viral protein.
- Cells expressing the mRNA into proteins display these viral protein fragments for the immune system to recognize.
- Exposing the immune system to the viral protein triggers an immune response, leaving our bodies prepared to fight infection by the actual virus in the future.
- The key to mRNA vaccines is that, instead of immunizing with pre-made viral protein fragments like in traditional vaccines, it is our own cells that translate viral genetic information from RNA into protein.

mRNA vaccines

- These vaccines contain information from messenger RNA, including the “blueprint” or code of a specific virus spike protein trait (virus antigen). The information enables the body to produce this antigen on its own: mRNA transfers the information for the production of the antigen to our cell machinery that makes proteins.
- Cells in our body then present the antigen on their surface and thus trigger the desired specific immune response. When the body comes into contact with the virus, the immune system recognizes the specific antigen and can fight the virus and thus the infection quickly and in a targeted manner.
- mRNA vaccines against COVID-19 are designed to provide our bodies with the code to produce the non-infectious virus spike protein to instruct the cell’s machinery to help stimulate a natural immune response
- Thus, in contrast to conventional vaccines, a mRNA vaccine does not contain any viral proteins itself, but only the information that our own cells need to produce a virus trait that triggers the desired immune response.



The vaccine encodes the presentation of **small, harmless fragments** of the COVID-19 virus to the immune cells, so they **"learn" how to recognize and attack the virus**. This allows for **a quick and specific immune response upon exposure** with the actual virus, thereby preventing its replication and spread in the human body and to other individuals.

mRNA Vaccines: Pfizer, Moderna

- These protein molecules stimulate an adaptive immune response that teaches the body to identify and destroy the corresponding pathogen or cancer cells.
- The mRNA is delivered by a co-formulation of the RNA encapsulated in lipid nanoparticles that protect the RNA strands and help their absorption into the cells.
- The advantages of mRNA vaccines over traditional vaccines are ease of design, speed and lower cost of production, the induction of both cellular and humoral immunity, and lack of interaction with the genomic DNA.

mRNA Vaccines

- mRNA vaccines offer specific advantages over traditional vaccines:
- Because mRNA vaccines are not constructed from an active pathogen (or even an inactivated pathogen), they are non-infectious.
- In contrast, traditional vaccines require the production of pathogens, which, if done at high volumes, could increase the risks of localized outbreaks of the virus at the production facility.
- Another biological advantage of mRNA vaccines is that since the antigens are produced inside the cell, they stimulate cellular immunity, as well as humoral immunity.
- mRNA vaccines have the production advantage that they can be designed swiftly. Moderna designed their mRNA-1273 vaccine for COVID-19 in 2 days. They can also be manufactured faster, more cheaply, and in a more standardized fashion (with fewer error rates in production), which can improve responsiveness to serious outbreaks.
- Pfizer requires very cold storage

mRNA Vaccines

- The COVID-19 mRNA vaccines from Moderna and Pfizer–BioNTech have efficacy rates of 90 to 95 percent.

- ***Can mRNA vaccines change your DNA?***
  No, the mRNA in the vaccine does not enter your cell's nucleus, and it does not interact with your DNA at all.
- In fact, RNA is very short-lived inside your body, and one of the challenges of vaccine development is making the RNA stable enough to do its job before your body degrades it.
- Additionally, the flow of genetic information from DNA to RNA to protein is unidirectional, and our cells cannot use mRNA to create DNA.

21 hominins: From 1864 to 1949, paleoanthropologists named at least nine human-like genera and 12 species

Reward neurons of social brain system

- What is the mechanism behind the motivation we feel to engage with others?
- Discovery of the reward system that are responsible for motivating us to interact with our fellow human beings.
- Study of neurobiological mechanisms at stake when two mice come into contact through learning a task.
- The motivation to invest in a social interaction is closely linked to the reward system, via the activation of dopaminergic neurons. These neurons release dopamine – the so-called pleasure prediction molecule
- Dopamine (DA) neurons of the ventral tegmental area (VTA) increase their activity during interactions with another person. VTA DA neuron activity encodes social prediction and drives social reinforcement learning

Lady Sapiens: role of women in human evolution

- Evidence that:
- Homo erectus: breast fed for up to 4 years (dental evidence)
- Black skin & blue eyes (sexual selection) in Europe until after 8 Ka
- Peruvian women hunters
- Women hunted rabbits
- Collected medicinal plants
- Not just male long lasting spears and stone tools; 90% of ancient goods were nondurable and made by women (clothing, bedding, knots, netting, carrying bags, small animal bones, plant material, etc.) – based on imprints in clay that have been preserved

Lady Sapiens

- Kristen Hawkes – Hadza modern hunter-gatherer tribes: grandmothers:
  - dug for deeply buried tubers:
    - provided 30% of tribe's tuber supply;
    - primarily for children's food supply;
    - weaned kids weight correlated with grandmother's foraging effort;
  - Foraging correlated with doubling of homo sapiens lifespan; subsidized fertility of younger females – could have next baby sooner because weaned kids were subsidized by grandmothers;
  - latter also passed on their knowledge
- Humans are only species with large numbers of post-menopausal females – benefit the group

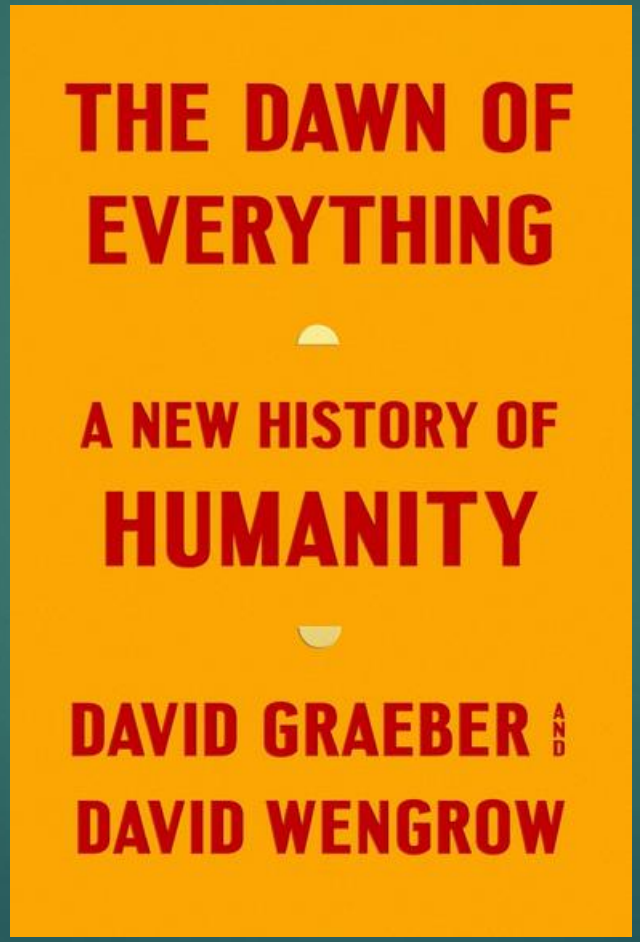
Firm upper age limit for Sima de los Huesos Neandertals

- Sima de los Huesos (SH) hominin fossil site, Atapuerca, Spain, have established a close minimum age of at least 430 ka for sedimentary material immediately overlying the human remains.
- Four single-grain TT-OSL depositional ages of 453 ± 56 ka, 437 ± 38 ka, 457 ± 41 ka and 460 ± 39 ka were obtained for the red clay lithostratigraphic units (LU-5 and LU-6) found underlying and encasing the SH hominin bones.
- The combined modeled ranges reveal that the hominin-bearing layer (LU-6) was deposited between 455 ± 17 ka and 440 ± 15 ka, with a mean age of 448 ± 15 ka.

Firm upper age limit for Sima de los Huesos Neandertals

- The SH fossils represent the oldest reliably dated hominin remains displaying Neandertal features across Eurasia.
- These Neandertal features are first observed in the facial skeleton, including the mandible and teeth, as well as the temporomandibular joint, and appear consistently across the SH collection.
- These chronological findings suggest that the appearance of these Neandertal traits may have been associated with the climatic demise of MIS 12 and the ecological changes that occurred in Iberia during this period.

New book challenges some historical assumptions



** Paleogenetics, Part 3:

aDNA characteristics & DNA Sequencing Technology

Charles J Vella, PhD, 2022

Genes from dinosaurs saved in amber?

Problem = DNA degrades

- 1990's Dinosaur DNA peer reviewed papers have never been retracted.

1995: Major Error: no Dinosaur DNA

- The first extraction of ancient DNA from dinosaur fossils was claimed in 1994-1995 in two papers.
- No phylogenetic tree produced. Was not replicable.
- Turned out to be contaminated plant, fungal, & human DNA (from a Y-chromosome)

S. R. Woodward, et al. 1994; Li et al. 1995; S. Blair Hedge and Mary H. Schweitzer, 1995

No Jurassic Park

- DNA begins as one very long strand.
- DNA Degradation: Sunshine (UV radiation) breaks down DNA in our skin, but proof-reading enzymes correct it in us.
- Once death occurs, it begins to degrade (break down into ever smaller fragments).
-
- UV radiation, oxygen, water, enzymes in gut, microorganisms in soil, etc. degrade DNA in dead cells.

DNA damage

- Postmortem, DNA strands are cut into ever smaller pieces.
- Greatest DNA degrader is water. DNA fragments may survive if cells dry out postmortem.
- Bones and teeth survive longest.

No Dino DNA; but what about proteins?

- DNA lasts longer in cold, dry places; permafrost is best preservative (Permafrost is ground that continuously remains below 0 °C (32 °F) for two or more years)

- Pääbo's earlier opinion: no DNA preserved after 1 million years

- No dinosaur DNA is possible
- But: 2017 - A 195 Ma Chinese *Lufengosaurus* — a long-necked herbivore = Collagen proteins discovered via a synchrotron to create high-power beams of infrared light and molecular mass spectrometry

Ancient DNA degradation

- Pääbo's lab: no replicable DNA from ancient amber
- His conclusion -- No dinosaur DNA: can't extract DNA from specimens that no longer have any.
- When organisms die, their DNA decomposes into minute fragments; the older the specimen, the smaller the DNA fragments.
- How long this takes depends on factors like temperature, burial conditions and the number of microbes making a meal of it.
- Eventual calculations then predict that in the optimal conditions — very cold ones — DNA could survive for around 1 million years.

Contamination

- By the end of the 1990s, aDNA researchers were concerned about contamination in a literal sense.
- This referred to unintentional and problematic exposure to modern DNA.
- Ancient DNA sequences, for example, were easily contaminated by environmental, bacterial, or recent human DNA introduced to a specimen over time or through handling in a museum collection or lab.
- This issue was heightened because of the degraded and damaged composition of aDNA, which resulted in fragmented genetic sequences. Therefore, it was difficult for researchers to determine what DNA sequences belonged to the actual specimen under study.
- This question of aDNA authenticity was a major problem for scientists.

Ancient DNA: fragmentation

- Direct quantitative comparisons of aDNA fragmentation in a large number of bone samples have revealed that:
  - the number of aDNA fragments exponentially increases with the decrease of their length
  - as the random breakage of long molecules results in an accumulation of shorter ones.
- While the rate of the fragmentation depends on different environmental factors, e.g., temperature, pH, and water availability:
  - initially rapid, most likely due to high enzymatic activity, and
  - followed by reduced rates over the long term.

Very Little DNA Survives Death

- Crucially, the quagga study noted what remains the most pervasive problem in the field of ancient DNA:
  - that very little DNA survives postmortem.
- 10 K to 1 M less DNA than when fresh: .0000001 mg of DNA per gram
- Worse enemy of DNA: Release of water and oxygen which are destructive and break apart the DNA
- Dust particle (in labs, usually skin particles, full of DNA) contamination

DNA degradation

- Symbiotic and external bacteria degrade DNA postmortem
- End up with small fragments from 10 to several 100 basepairs long
- Ancient DNA is always damaged,
  - This fact is a marker that you have the right type of DNA (a way to exclude contamination which is better preserved)

Types of DNA damage

- Some damage results in changes to the DNA sequence.
  - Cytosine can change to uracil, which is read by copying enzymes as thymine, resulting in a C to T transition. This is the most common type of damage
  - Normally, in spontaneous chemical damage in functioning cells in the body, 10,000 C's per cell morph into U's each day, and are removed and correctly replaced with a T. Enzymes replace incorrect nucleotides before a bond rupture can occur.
  - Changes from G to A also occur.
  - Some of these DNA modifications are problematic because although they allow the amplification of the template molecules, they cause incorrect bases to be incorporated during the PCR.

- DNA errors are very common at the ends of molecules.

Setbacks: Contamination

- Setbacks and occasional disasters: paleogeneticists have discovered to their chagrin how easily ancient DNA samples can become degraded and contaminated with modern DNA, giving rise to erroneous and misleading conclusions.

- Contaminating DNA. DNA introduced into an experiment from the
  - preservation environment,
  - from excavation,
  - sample handling,
  - sample processing,
  - during the experiment itself;
  - DNA on dust particles

Contamination horror

- Pääbo describes his horror at watching a curator at the Natural History Museum in London lick an ancient bone to detect whether it had once been chemically treated, thus coating it with his own DNA.

- Quite different is Pääbo's savoring of the smell of burnt bone when a Neanderthal arm bone was being cut to provide a sample for analysis—the smell suggested that collagen within the bone had survived and hence DNA would be found.

DNA damage and DNA contamination almost killed the field

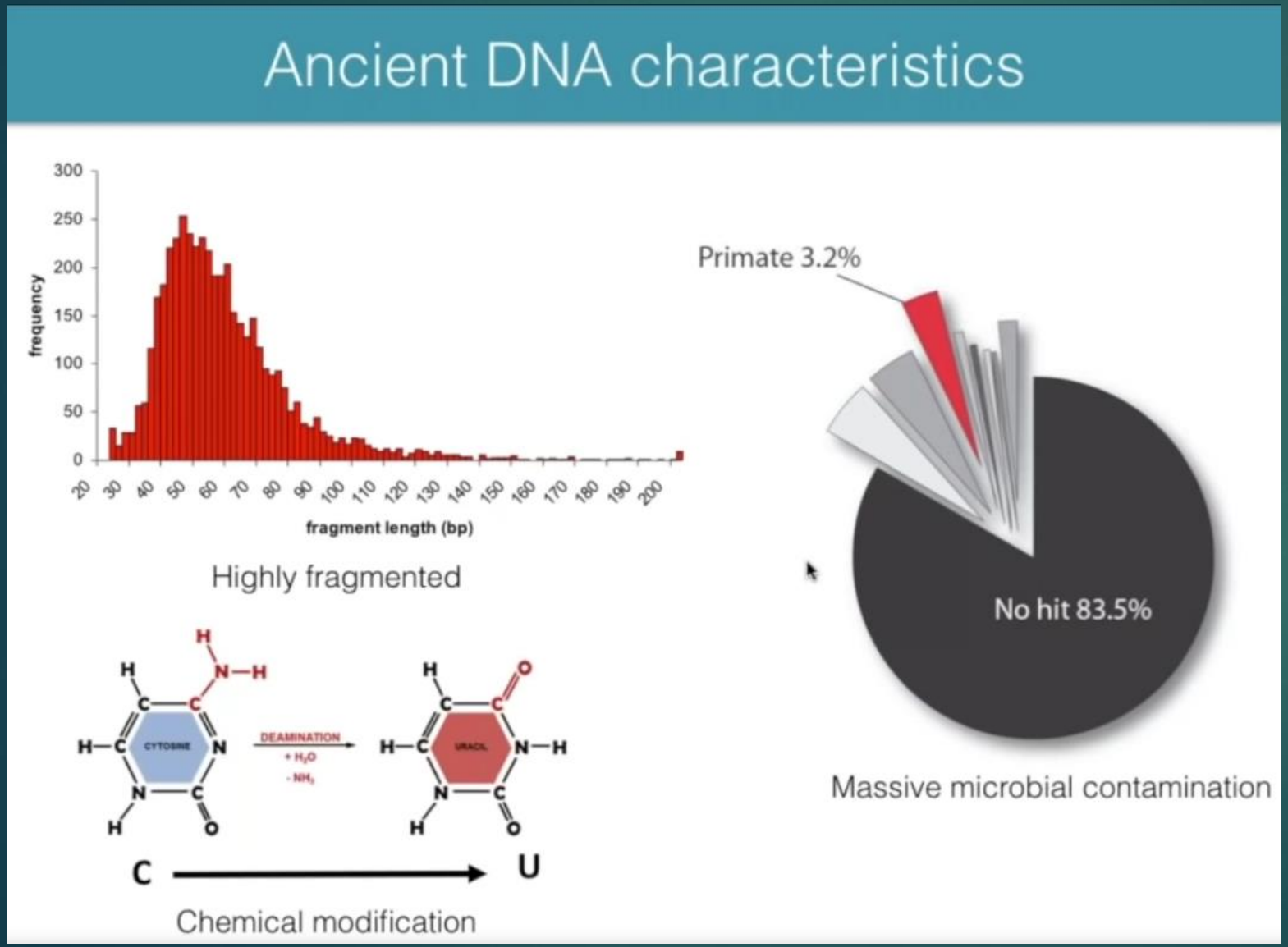
- Just 3 years after extracting mtDNA from the quagga, Wilson's team discovered that some of its sequences had undergone chemical alteration after the animal died, complicating attempts to figure out how closely related the extinct animal was to living horses and zebras.

- And in 1994, the claims of a research team to have sequenced dinosaur DNA—later discovered to be human contamination—nearly led to the premature death of paleogenetics.

Contamination is huge problem

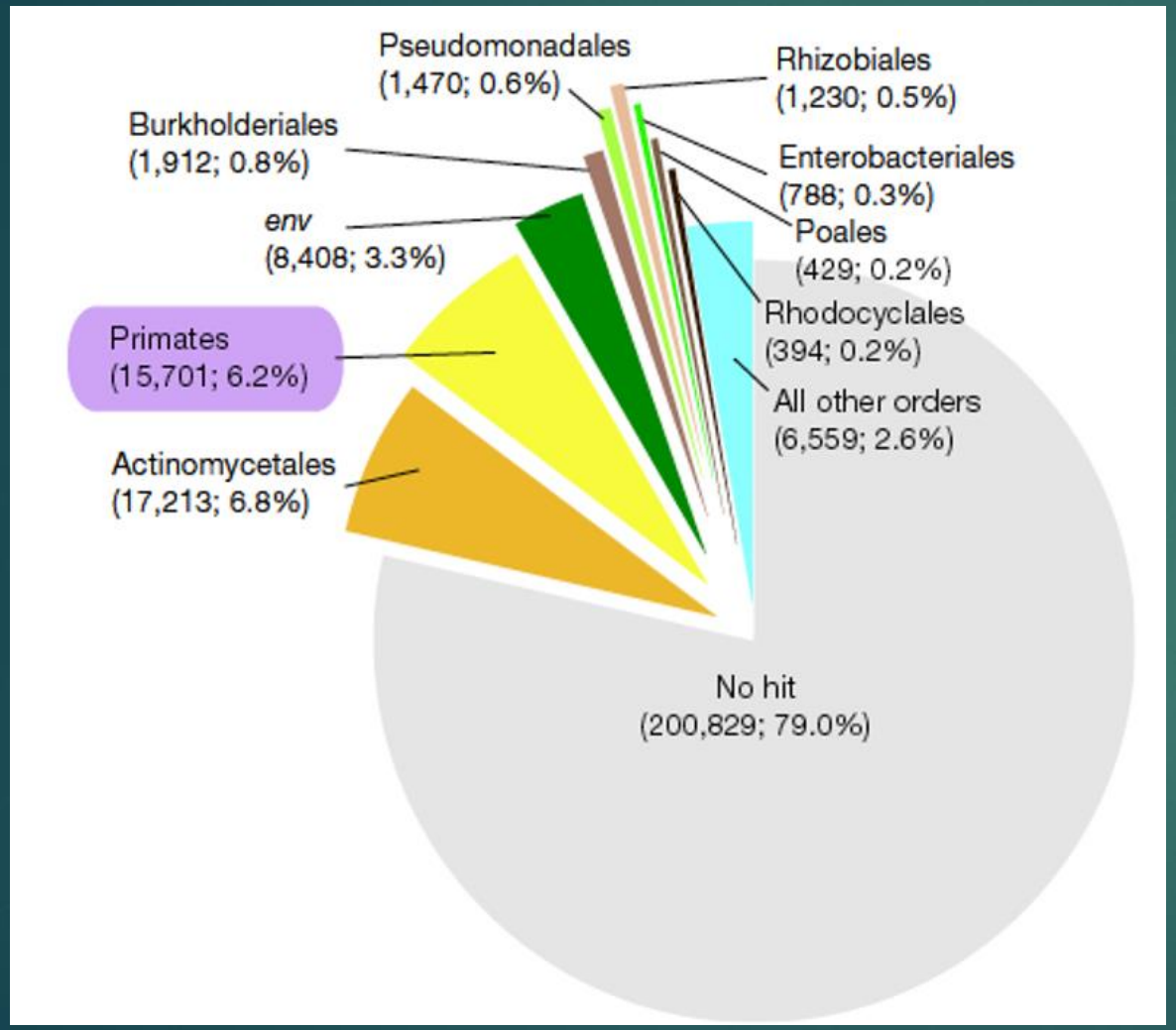
- Contamination by modern DNA is a particularly difficult problem to solve.
- Labs (and chemicals) may be contaminated by the DNA of the people working in them, while many fossils have been handled by researchers for years.
- Contamination is difficult to detect because Neanderthals and humans share much of their genetic material, making some DNA sequences indistinguishable.

Ancient DNA characteristics:



- aDNA = is very fragmented (40-50 bps),
- It is damaged (Cytosine to Uracil modification),
- It is mostly nonhuman (only 3.2% primate)

aDNA contamination = Mostly unknown soil-living microbes; Very little hominin DNA in bones: typically 3.5%



Mostly contaminating microbes

In first 12 years of work on aDNA, Pääbo found human DNA in every animal DNA sample he worked on

DNA survival

- Pääbo has warned future practitioners of the specific challenges associated with working with ancient material.

- Ancient DNA sequences contain
  - chemical modifications including
  - strand breaks,
  - DNA crosslinks,
  - modified bases,
  - that make their recovery challenging.

- He proposed an inverse relationship between fragment length and the number of surviving molecules of that length (i.e. smaller the fragment length, greater the number of fragments of that length) vs contaminants which are better preserved

From errors to success

- DNA preservation is not determined by specimen age but by the environment in which the specimen was preserved.
- Contamination by modern DNA is likely to be the most serious challenge of working with ancient specimens
- Early days of ancient DNA were marked by a few spectacular but flawed results, the field has matured into a robust, internally rigorous scientific pursuit with the potential to provide real insight into the mechanisms of evolution at both the species and the population level

Expected damage patterns consistent with authentic aDNA

- Computer R package decontam to detect and remove laboratory and environmental contaminants prior to downstream computer analysis.

- Use classic aDNA damage type to identify what is actually aDNA

- Examine each dataset and confirmed the presence of DNA damage characteristics of ancient samples, including
  - short fragment lengths and
  - elevated levels of cytosine to thymine deamination

Ancient DNA: Miscoding lesions

- High-throughput sequencing analyses of miscoding lesions have confirmed that
  - (*a*) cytosine deamination (enzyme breakdown) to uracil is the most common base modification.
  - (*b*) this deamination increases toward fragment ends, promoting the formation of single-stranded overhangs
  - (*c*) depurination drives postmortem DNA fragmentation.
    - Depurination = DNA bond is cleaved releasing a nucleic base, adenine or guanine; loss of bp site

Ancient DNA

- Favorable environmental conditions for aDNA preservation include
  - low temperatures,
  - rapid drying
  - high salt concentration.

- These factors facilitate destruction and/or inactivation of bond cleaving enzymes and reduce bacterial metabolic activity and water effects.

- Consequently, the oldest genomes sequenced to date come from specimens preserved under such conditions, for instance,
  - a 700 Ka horse bone excavated in Yukon, Canada, and
  - a 430 Ka Neandertal hominin fossil at Sima de los Huesos in Spain.

aDNA: effect of temperature on preservation

- The oldest environmental aDNA has been sequenced from ice and permafrost ranging between 400 Ka and 1.6 Ma in age.

- In contrast, the age of the oldest environmental aDNA reads from the tropics is ~2 orders of magnitude (100 x) lower.

- While future studies may succeed in retrieving DNA sequences older than one million years current technological and methodological limitations make it hard to imagine such practice ever becoming routine.

Precautions

- In response, two aDNA researchers – Alan Cooper and Hendrik Poinar (the amber insects guy) – co-authored a paper in Science, titled "Ancient DNA: Do it Right or Not at All" (Cooper and Poinar, 2000). In it, they argued for the adoption of nine criteria for doing aDNA in order to avoid contamination.

- *Physically isolated work area*. To avoid contamination, it is essential that, prior to the amplification stage, all ancient DNA research is carried out in a dedicated, isolated environment. Physical separation of the "Ancient DNA Lab" from other labs.
- *Control amplifications*. Multiple extraction and PCR controls must be performed to detect sporadic or low-copy number contamination.
- *Appropriate molecular behavior*. PCR amplification strength should be inversely related to fragment size (large fragment sizes (300+ bps sized fragments are unusual). Sequences should make phylogenetic sense.

Precautions

- *Reproducibility*. Results should be repeatable from the same, and different, DNA extracts of a specimen.

- *Cloning*. Direct PCR sequences must be verified by cloning amplified products to determine the ratio of endogenous to exogenous sequences, damage-induced errors, and to detect the presence of nurnts. Overlapping fragments are desirable to confirm that sequence variation is authentic and not the product of errors introduced when PCR amplification starts from a small number of damaged templates.

- *Independent replication*. Intra-laboratory contamination can only be discounted when separate samples of a specimen are extracted and sequenced in independent laboratories.

Contamination: garbage in, garbage out

- However, contamination was even more of an issue when it came to working on ancient humans since authentic sequences were difficult to distinguish from contaminants.

- PCR exacerbated the issue of contamination because of its high sensitivity, often amplifying exogenous DNA in the PCR reaction because it was often better preserved than endogenous DNA. The obtained sequences could appear reliable but in fact be false positives

- PCR is preferentially prone to recover chemically unaltered, modern contaminant DNA from excavators and laboratory workers.

PCR & NGS

- By 2005 an innovation of a new technology called next-generation sequencing (NGS).
- While PCR required specific DNA fragments of sufficient length to be present in a sample, NGS was capable of sequencing every DNA fragment present.
- As a result of NGS, researchers could now recover millions of sequences at higher yield and in a fraction of the time.
- This allowed better estimation of the percentage of endogenous and exogenous DNA by searching for signatures of molecular degradation or post-mortem damage characteristic of authentically ancient DNA.

Neanderthals and too much data

- Paabo and his team finally finished the Neanderthal Genome Project (Green et al., 2010). The effort, conducted by over 50 scientists at a cost of approximately 5 million euros, successfully sequenced 4 billion base pairs of Neanderthal DNA (much duplication; more than a third of the genome remained unsequenced).

- However, it was not just the data that was important but the ability to analyze it that was critical.

- The combination of this genomic data and statistical methods, developed by Harvard University geneticist David Reich and his lab, allowed them to detect signals of admixture between early humans and Neanderthals.

Too much data? Need for bioinformatics

- This race for the first, the oldest, or the most human genomes shared striking similarities to the 1990s' hunt for the first or the oldest DNA.

- However, one key difference between the PCR and NGS eras was that practitioners went from having too little to almost too much data.

- Machine sequencing technologies could produce large amounts of data that required researchers to seek or learn computational and statistical skills to interrogate it.

- From a field dominated by the laboratory scientist, aDNA research was moving into the realm of the bioinformatician/statistician.

Biometric analysis

- New rule: ‘Grab as much data as you possibly can, hire a great bioinformaticist, and then start asking questions in the resulting datasets’.

- aDNA research, scientists argued, can be seen as data mining in terms of producing data and describing its patterns without a specified hypothesis (Millar and Lambert, 2019).

- This has been criticized as a deviation from the normal scientific hypothesis-based approach.

Lack of hypothesis testing

- Given the nature of aDNA research, aDNA scientists often adopt a material-based rather than inductive approach as hypotheses rarely can be built a priori and then be tested on the results.

- For a few, this approach was far from desirable and even problematic.

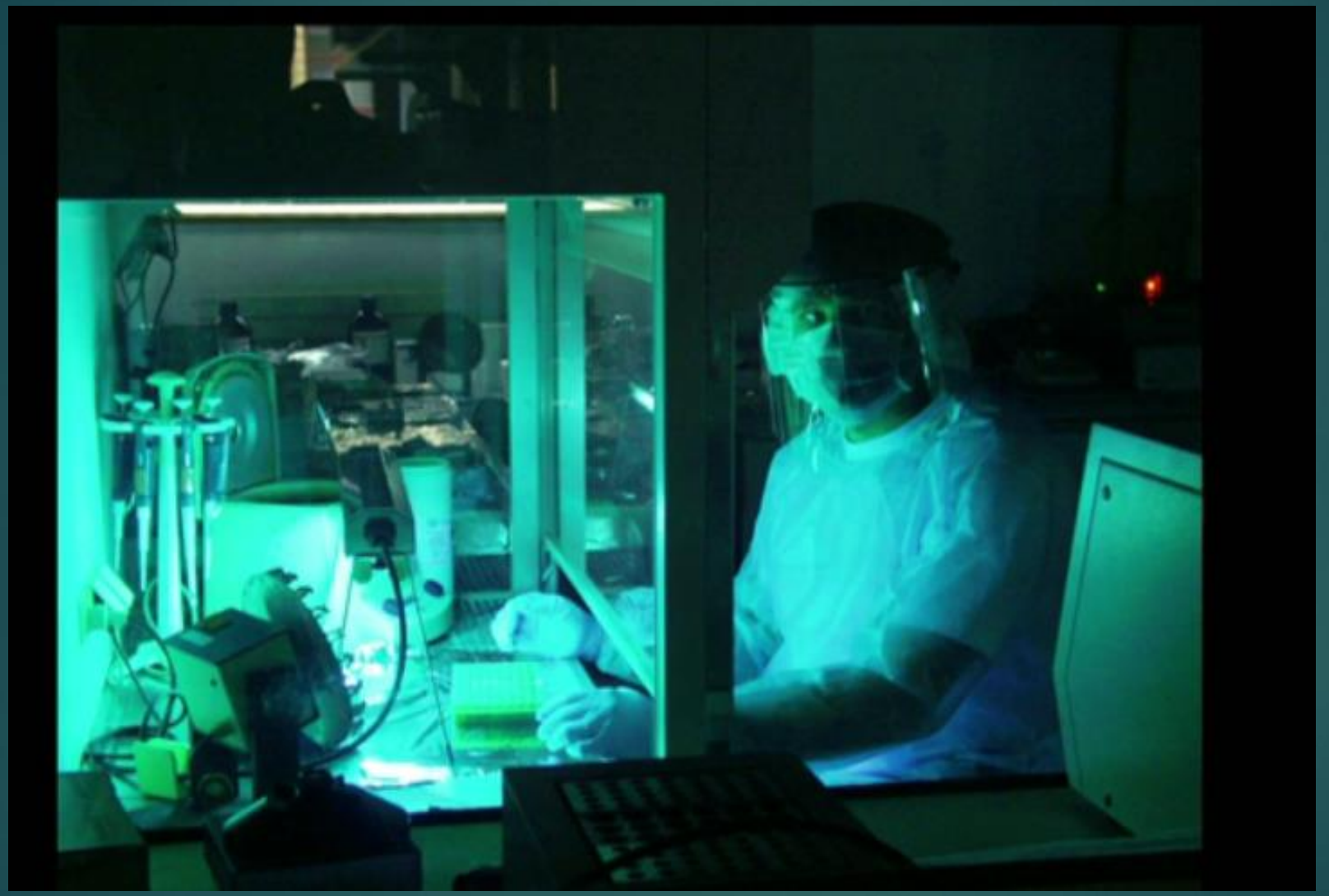
- The current hype cycle is focused on the interpretation and presentation of DNA from ancient humans that sometimes offers the public an inaccurate impression that genetic evidence can conclusively and consistently solve archaeological problems.

Ancient DNA

- To date, the title for 2nd oldest complete genome belongs to a horse unearthed from frozen ground in Yukon, Canada, and dated to be 780-560 Ka old.
- The oldest current genome is from three 1.6 Ma old mammoths.
- The oldest DNA from a member of our *Homo* genus is a fragment of genetic code from 448-430 Ka old Neanderthal ancestors found in Spain's *Sima de los Huesos* cave, which stays at a cool 50 degrees Fahrenheit.
- Recently, discovery that DNA preserves best — with up to 100 times higher recovery rates — in the petrous bone of temporal lobe (bone which houses inner ear)

aDNA research needs Ultraclean Rooms

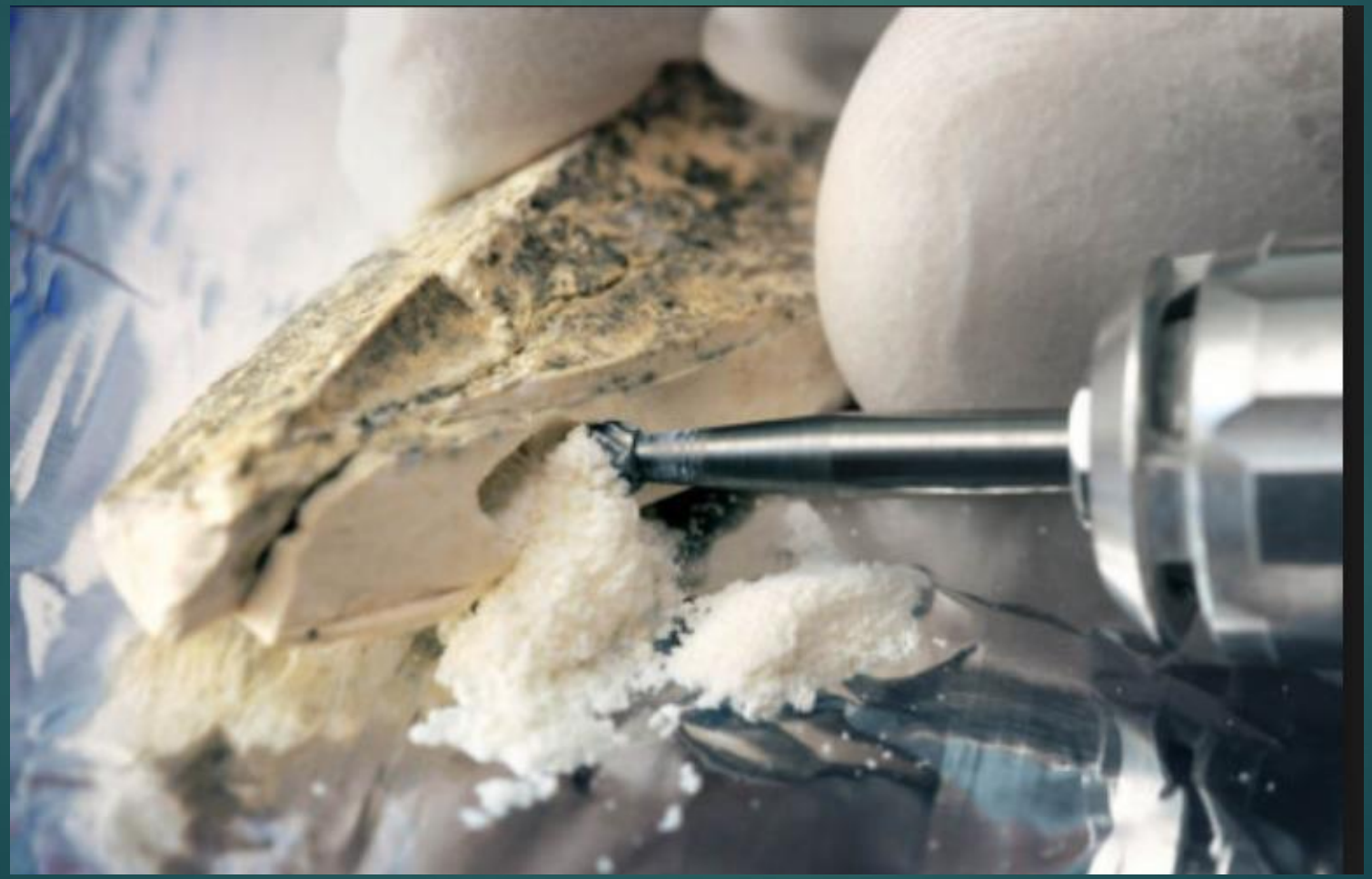
Ancient DNA Extraction



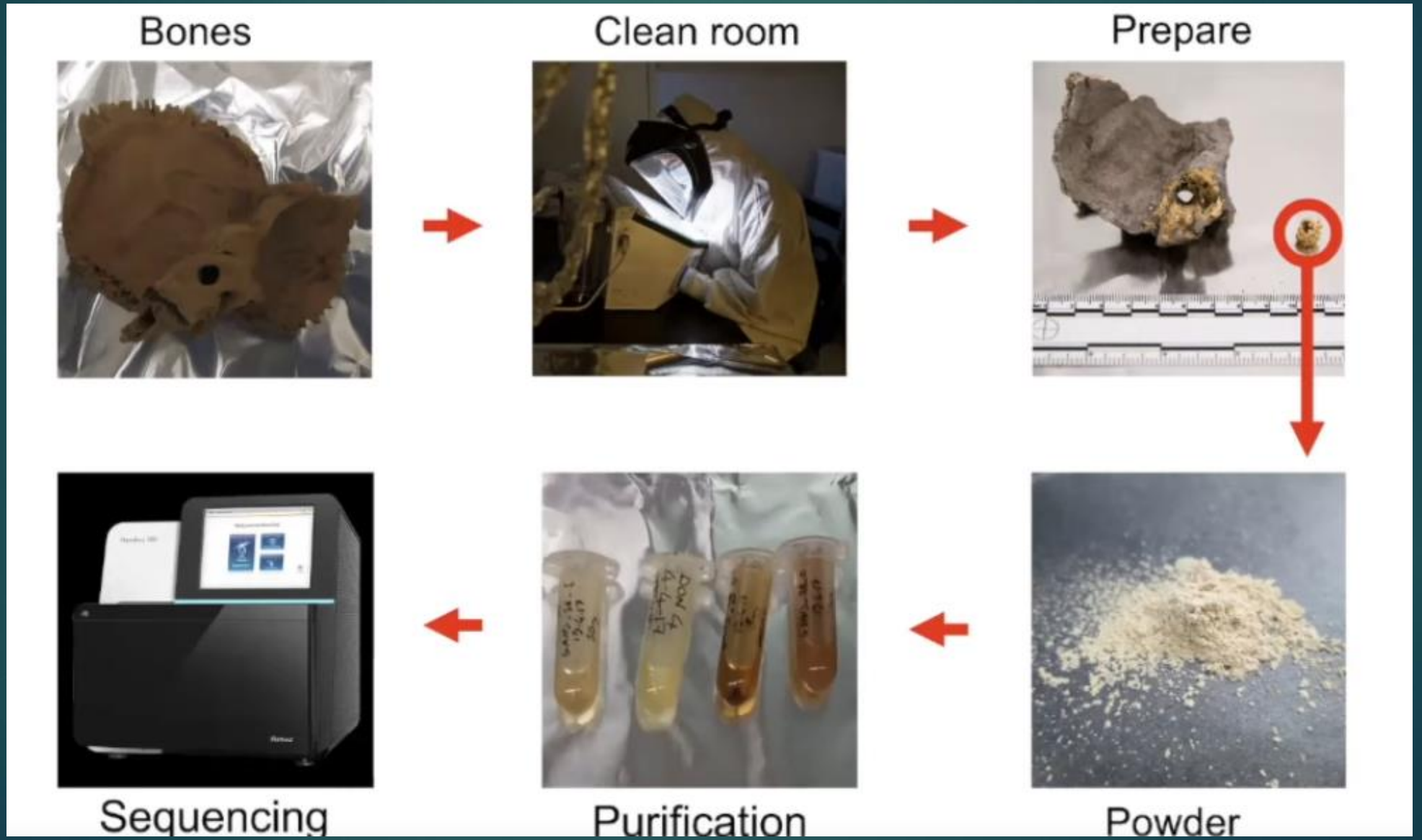
Like
silicon chip
factory
production

Contamination reduction: Isolation, high reverse pressure air flow, UV li

Accessing DNA in Bone



Steps in analyzing aDNA



1. Extraction

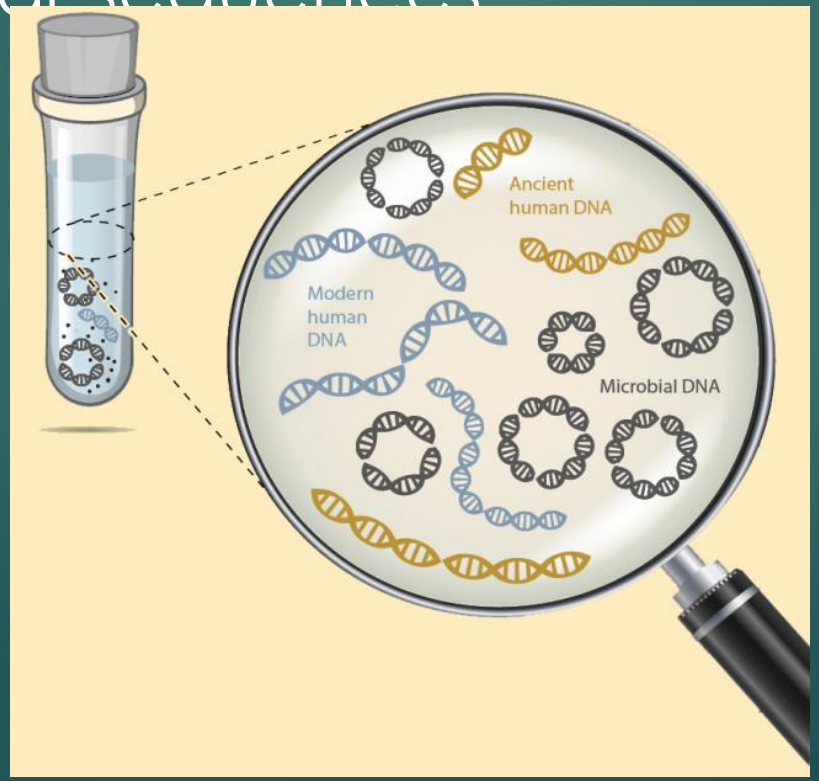
- In a sterile lab, bones and other surviving tissues are cleaned, crushed into powder and dissolved with chemicals that isolate short DNA strands.

aDNA: Extraction

- The characteristics of each sample pose specific requirements for the extraction procedures, related to the effective digestion of the material and the solubilization of the DNA.
- For mineralized samples (e.g., bones, teeth), this is often undertaken with the aid of chemical buffers: demineralizing agents; detergents; surfactants
- Collectively, these compounds break down cell walls, degrade proteins, and release the DNA into solution
- Dissolved DNA is then purified. The use of multiple purification steps, however, represents a trade-off between removing enough inhibiting substances to allow downstream analysis and maintaining workable quantities of DNA.

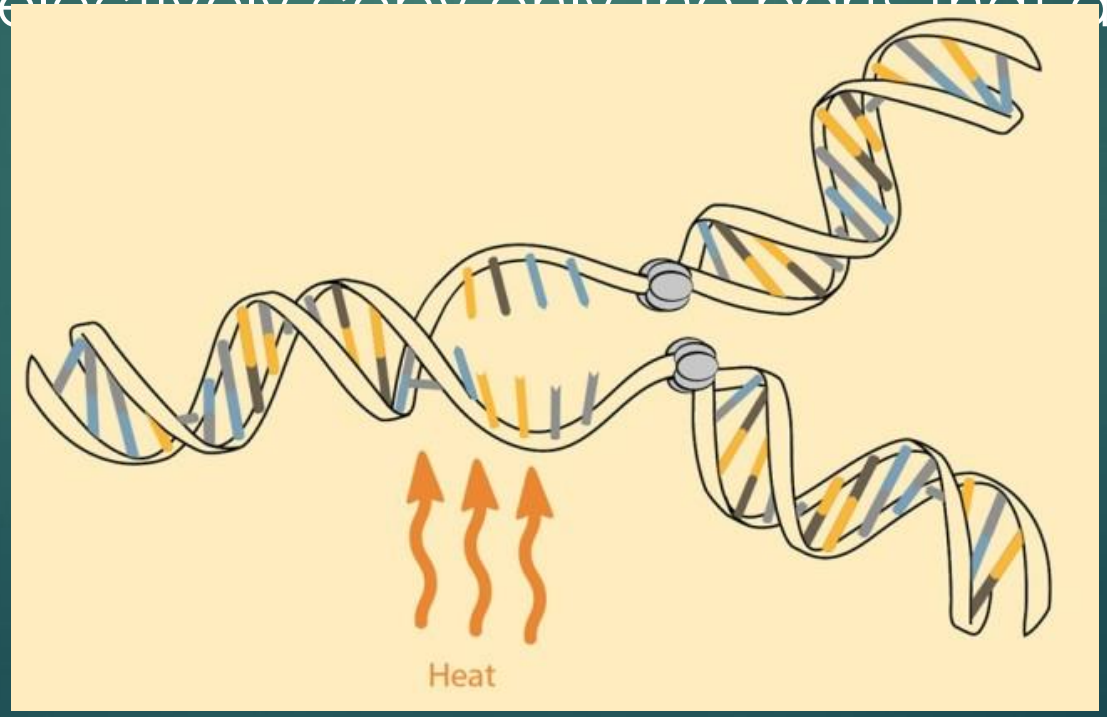
2. DNA soup

- Despite cleaning, the extract is a soup of DNA from the sample and contaminated material, mostly microbes from soil where the remains were buried. Researchers add molecular tags that will later work like barcodes, selectively binding to DNA to inventory and find particular sequences



3. Copies

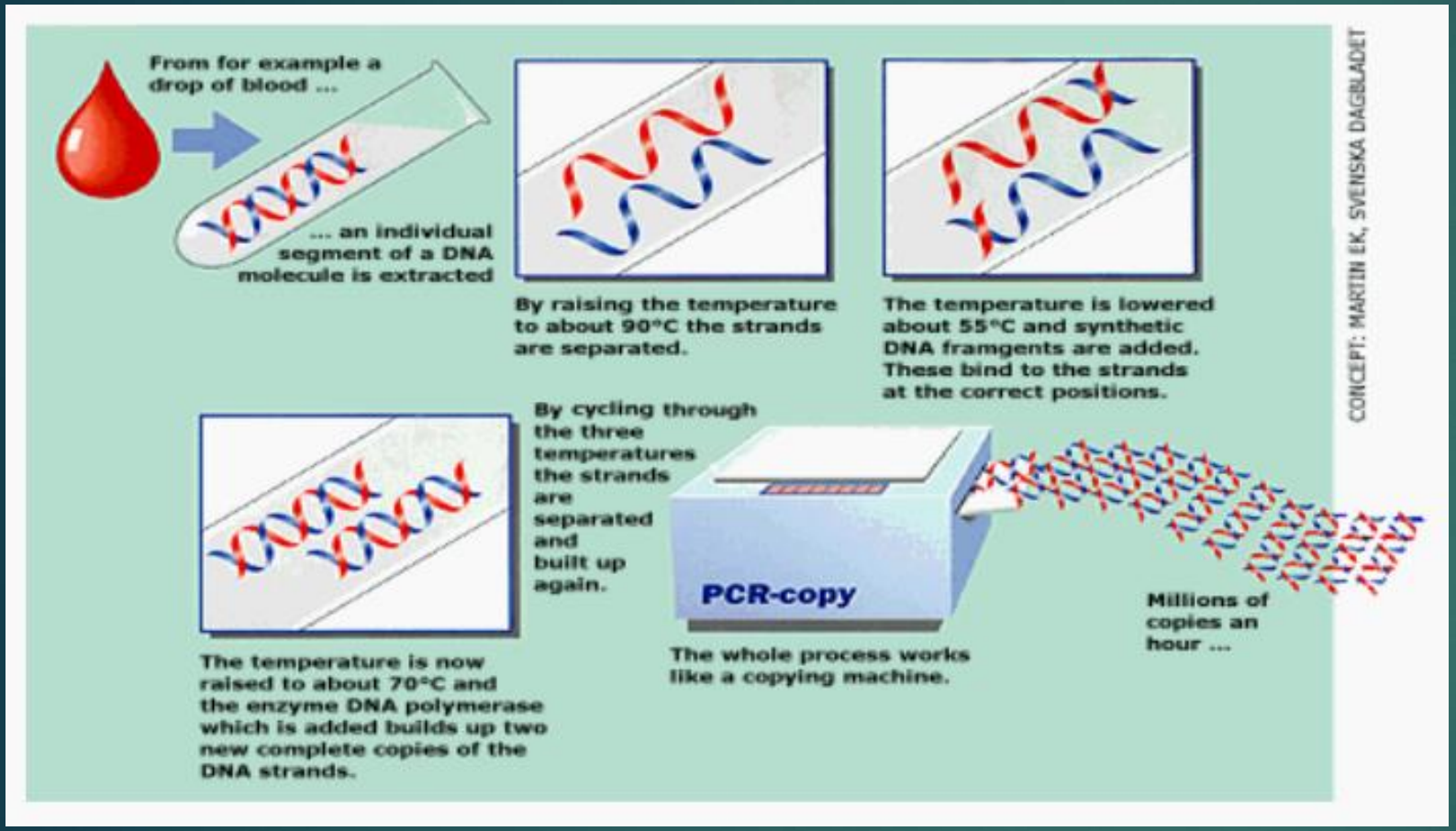
- **To** read the DNA quickly and accurately, computers must analyze millions of copies simultaneously. Geneticists make these copies by heating double-stranded DNA, causing it to separate; enzymes then build new double-strands from each half. Repeating the procedure makes two strands into four, then four into eight, and so forth, until millions of strands have identical DNA code and tags. Because human genomes are over 99 percent identical, researchers often selectively copy only the parts that differ.



PCR

- Polymerase Chain Reaction (PCR): An enzymatic & thermal technique for amplifying from one to a few copies of DNA by several orders of magnitude.

PCR: DNA Xerox Machine



1 - Example: From a drop of blood, an individual segment of a DNA molecule is extracted.
2 -By raising the temperature to about 90*C the strands are separated.
3 - The temperature Is lowered about 55*C and synthetic DNA fragments are added . These bind to the strands at the correct positions.
4 - The temperature is now raised to about 70 *C and the enzyme DNA polymerase which is added builds up two new complete copies of the DNA strand.
5 - By cycling through the three temperatures, the strands are separated and built up again.
6 - The whole process works like a copying machine. Millions of copies an hour.

4. Sequencing & Authentication

- **4. Sequencing:** In the final round of copying, geneticists chemically color the different nucleotide bases — better known by their letters A, T, C and G. Computers then read the code based on the order in which the colors appear, analyzing all the identical strands with matching tags at the same time to weed out any errors.
- **5. Authentication:** Researchers use a number of clues to discard contaminants in the ancient code: For example, aDNA strands are usually shorter than 100 base letters, with predictable degradation patterns

DNA damage types and ways to combat them

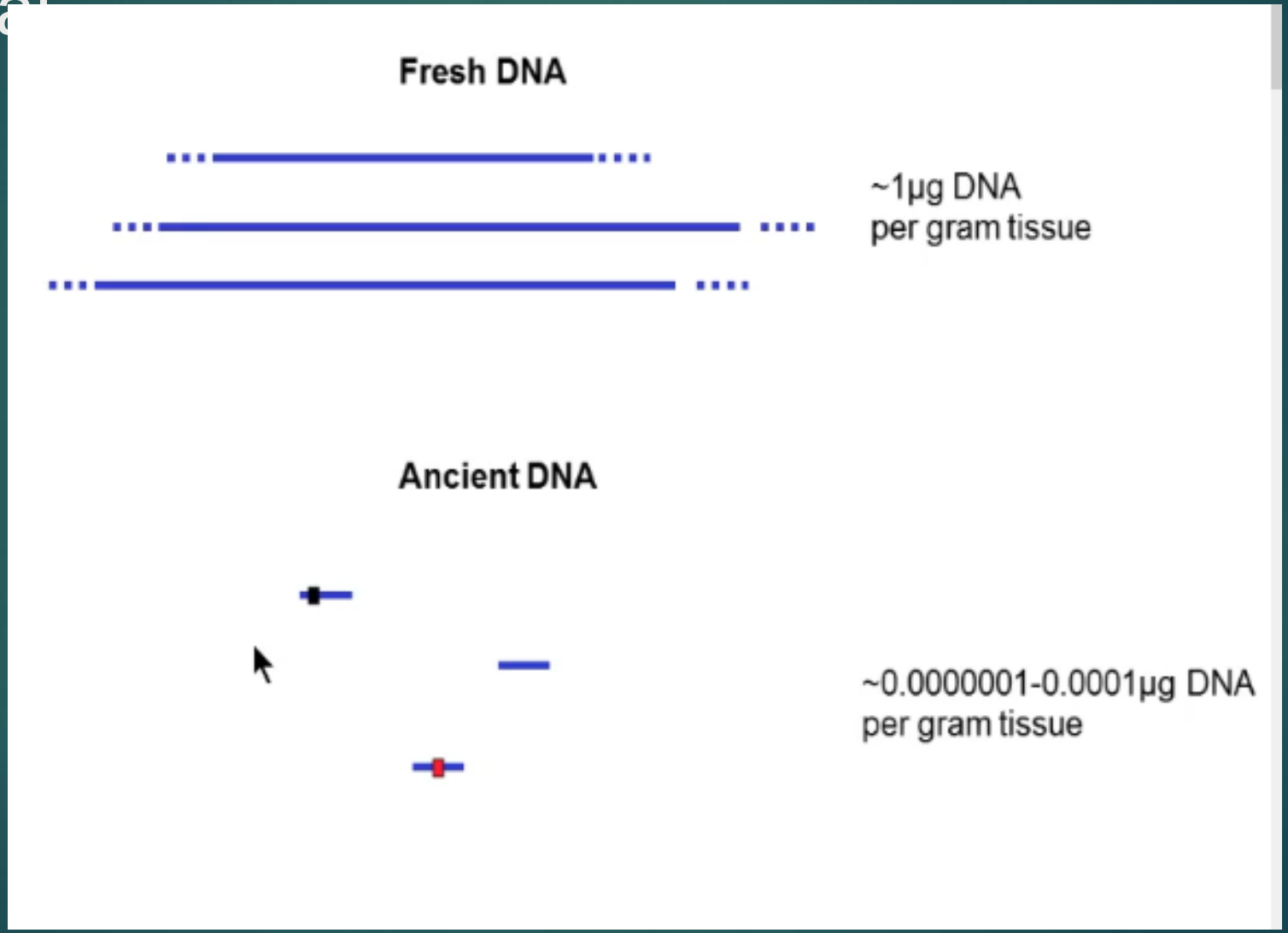
Damage	Process	Effects on DNA	Possible solutions
Strand Breaks	Degradation by Microorganisms Nucleases in postmortem cell Other chemical processes	Reduction of overall DNA amounts Size reduction	PCR of overlapping fragments of short length
Oxidative lesions	Damage to bases Damage to deoxyribose residues	Base fragmentation Sugar fragmentation Nucleotide modification	PCR of overlapping fragments of short length Multiple independent PCRs Cloning and sequencing of several clones
DNA crosslinks	Reactions between DNAs as well as DNA and other biomolecules		PTB (N-phenylacyl thiazolium bromide)
Hydrolytic lesions	Loss of amino groups 1. adenine => hypoxanthine 2. cytosine => uracil	Change of Coding potential	Multiple independent PCRs

6. Alignment

- The short strands need to be arranged into their proper positions across the full genome.

- Software lines them up based on overlapping stretches of code and comparisons with previously sequenced reference genomes.

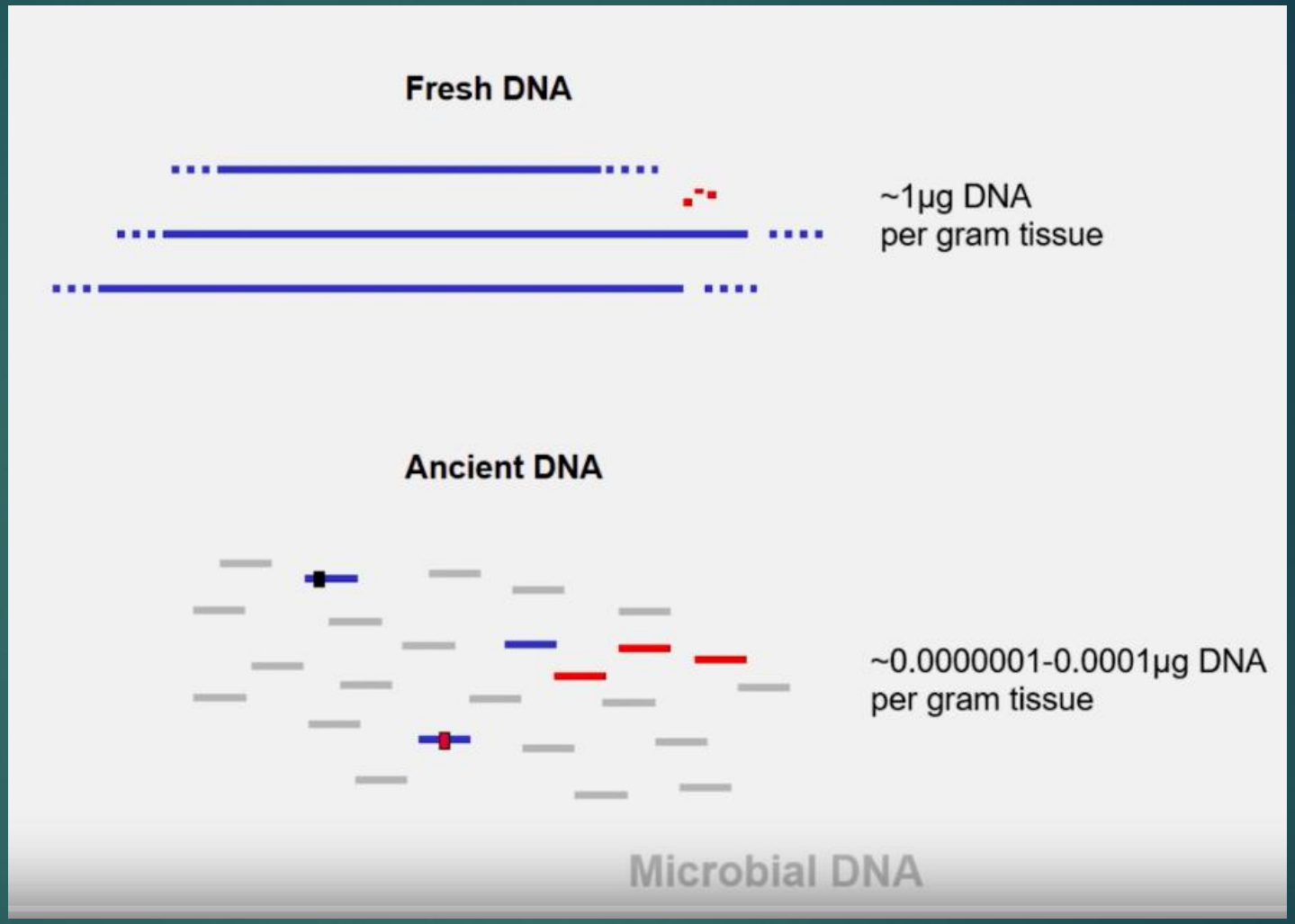
aDNA is very fragmented: 1 Million times less DNA and is chemically modified; and is mostly bacterial/fungal



DNA extraction

10,000 basepair lengths from modern human fresh tissue

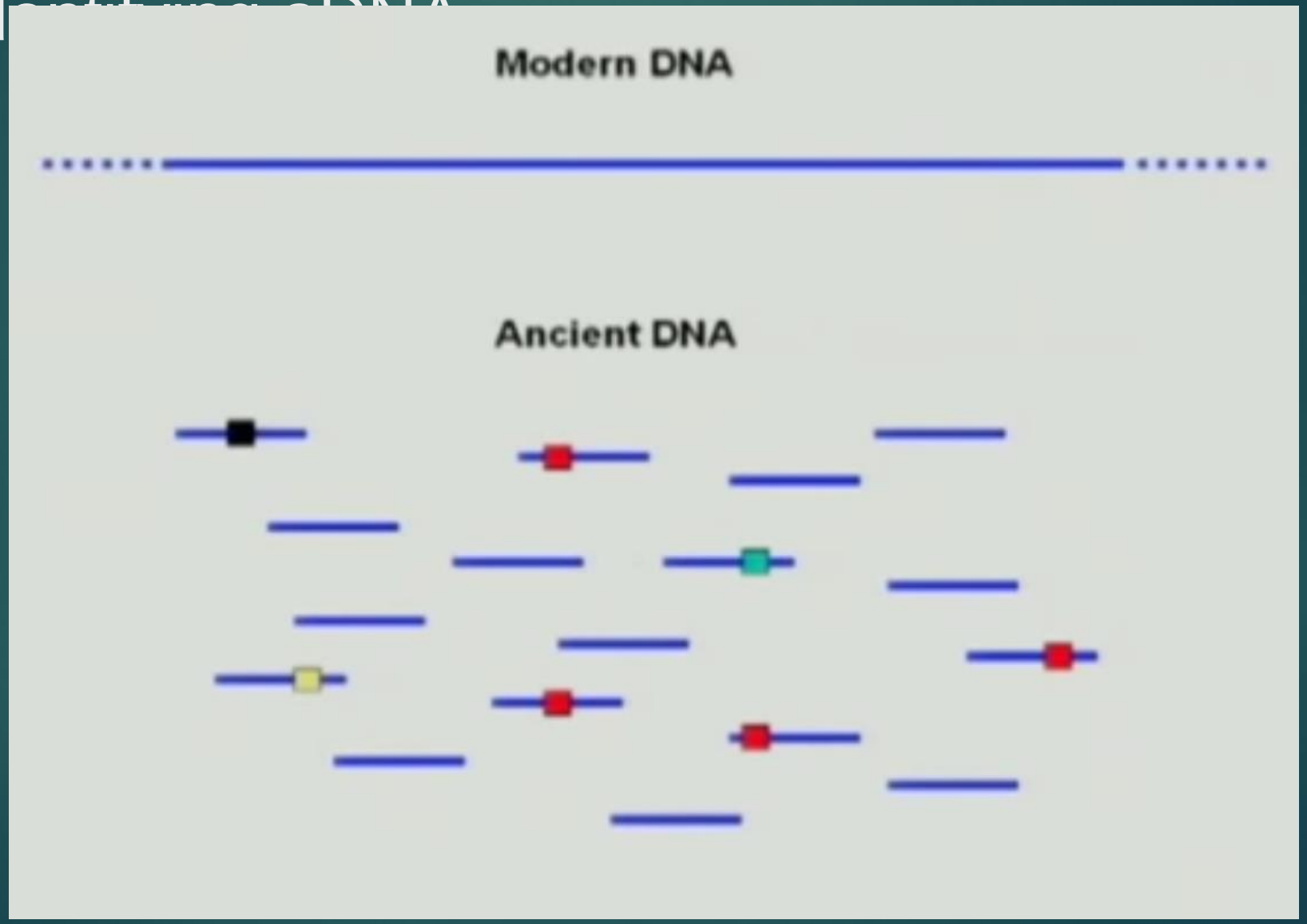
30-40 basepairs of Ancient DNA



1 millionth of gram

Always massive contamination:
In ancient DNA, mostly microbial DNA (grey) and potentially modern researcher DNA (red)

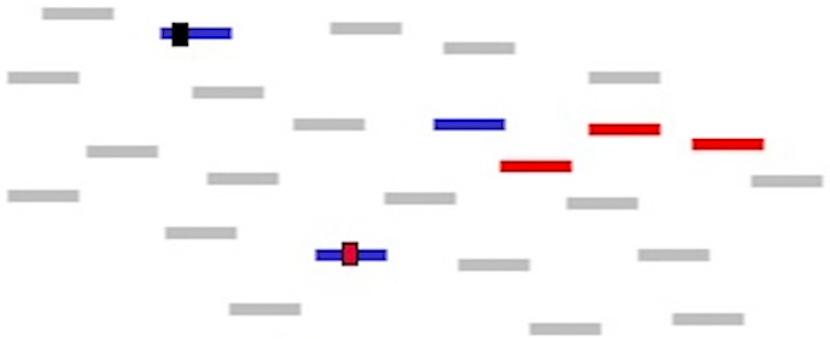
Modern vs Ancient DNA: long strands vs short fragments (50-60 bps); type of damage becomes method of id



Fresh DNA

~1µg DNA
per gram tissue

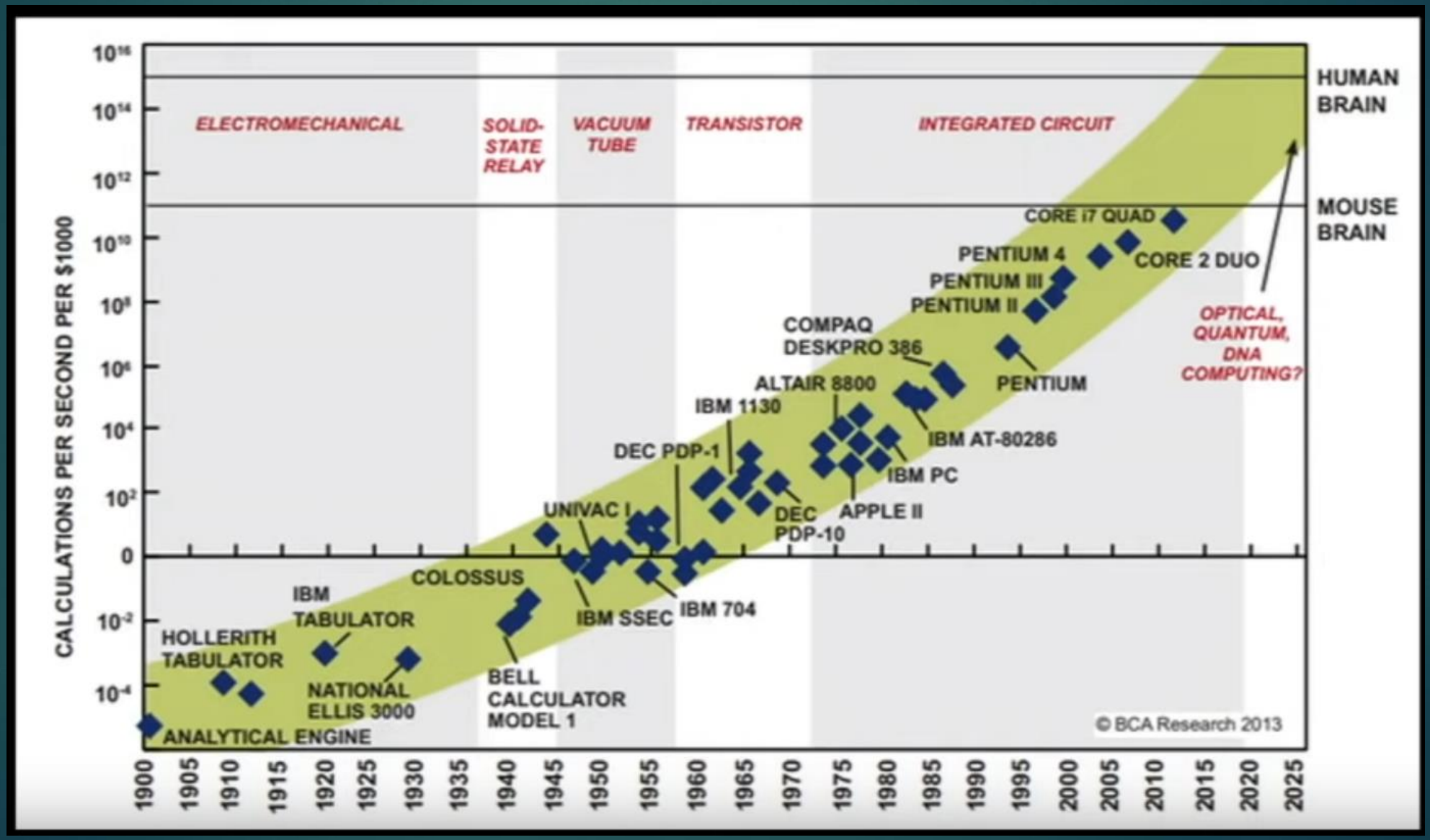
Ancient DNA



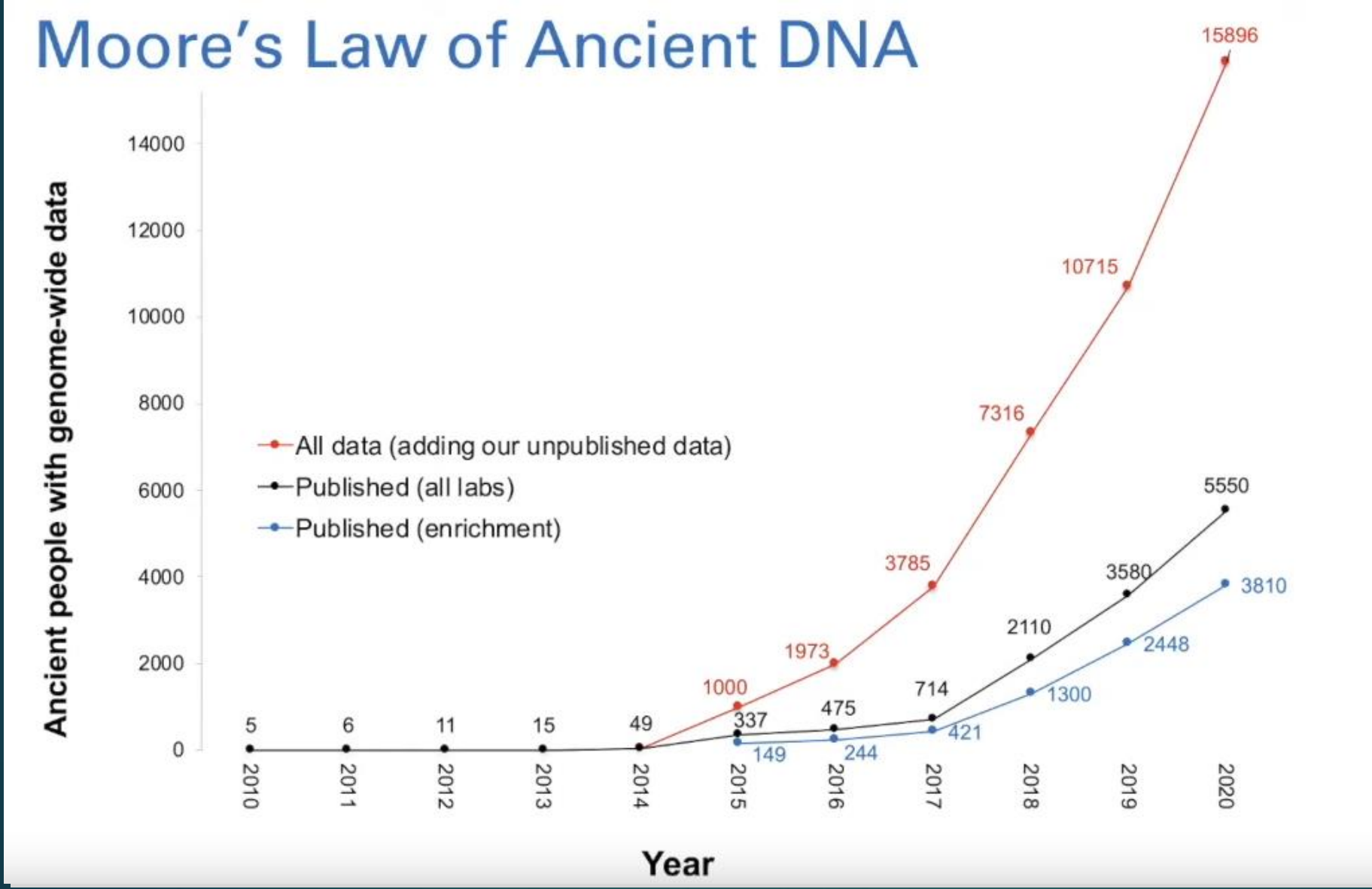
~0.0000001-0.0001µg DNA
per gram tissue

Microbial DNA

Moore's law of Computer Processing:
double power every 2 years



Published ancient full genomes:



- 0 in 2009
- 50 in 2014
- 5500 in 2020
- unpublished estimate = 15,896

Limiting factors on aDNA search

- Polymerase chain reaction = DNA duplication technique

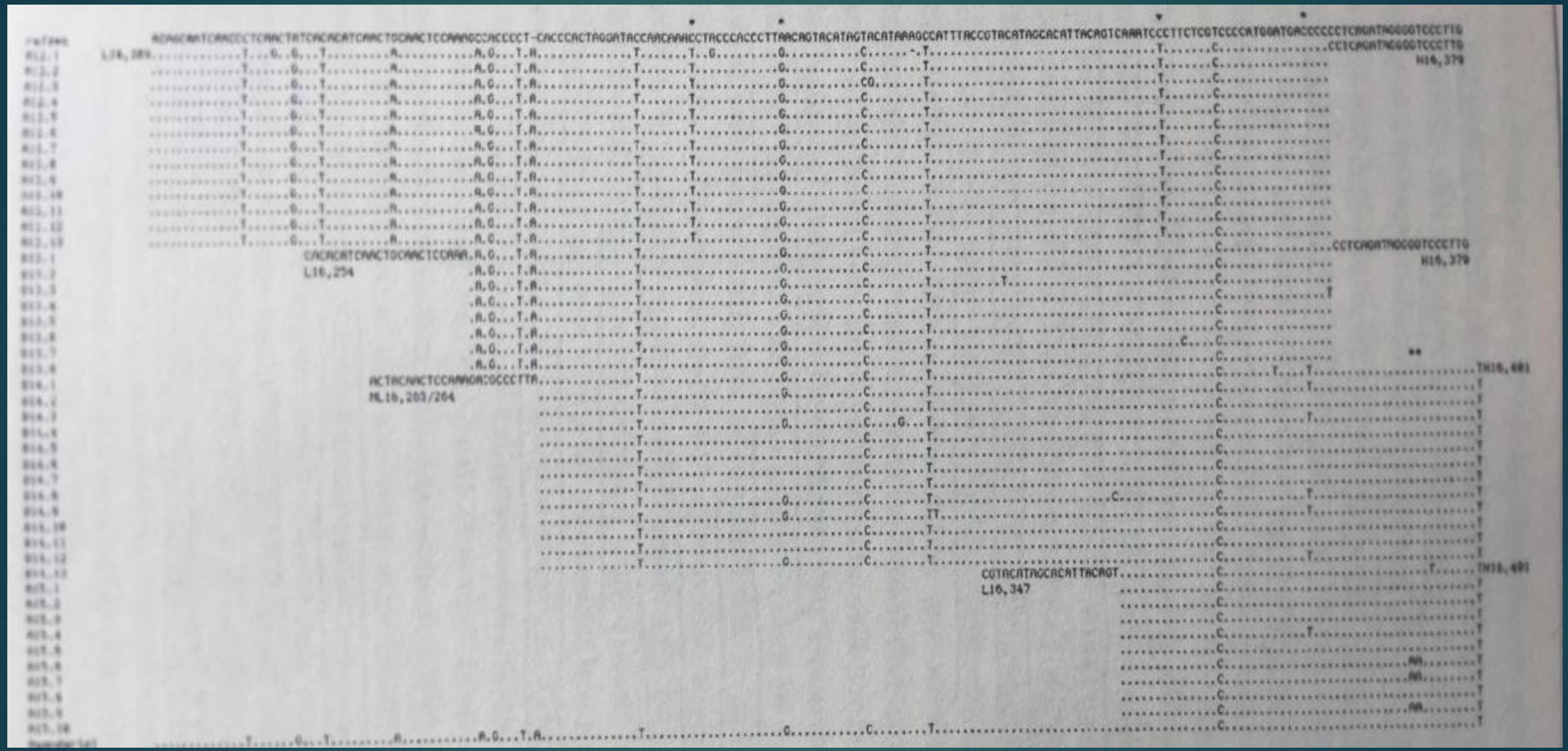
- Limiting factors on analysis:
  - 1) In Human specimens: PCR amplifies human contamination instead of degraded DNA

  - 2) In Bacterial specimens: Current knowledge of only 2% of bacterial diversity; how do you know bacterial DNA is ancient

Effects of DNA damage = ways to identify aDNA

- DNA Backbone breakage– increased fragmentation into very small fragments
- Deamination of C causes wrong base to be added during PCR -- false mutations
- C and T residues converted to hydantoins, blocking DNA polymerases (PCR)
- Ancient DNA is more fragmented and prone to have specific types of DNA damage.
- For example, “A” and “G” nucleotides are enriched at the ends of ancient sequences because the genome seems to break at these locations over time.

Is it a Neandertal basepair?



Top line: human reference sequence; Bottom line: Neanderthal sequence. Repeat clones in between.

Reconstruction of a piece of mtDNA from the Neanderthal from Neander Valley

- First line above, a modern reference sequence is shown.
- Bottom line is the reconstructed Neanderthal nucleotide sequence.
- Each line below represents one cloned molecule amplified from the Neanderthal type specimen.
- Where these sequences are identical to the reference sequence, a dot is placed; where they differ from the nucleotide, I have written them out.
- At each position, you require that a change from the reference sequence is seen in a majority of clones and in at least two independent PCR experiments (either the ones shown or others).
- Matthias Krings et al.. "Neandertal DNA sequences and the origin of modern humans," *Cell* 90,19-30 (1997).

aDNA in “fossils”

- Please note that if you have complete fossilization, it means that all the organic components have turned to minerals
- aDNA can only be found in “fossils” that have not been completely fossilized
- There is still organic compounds in the fossilizing animal body
- aDNA is organic, not fossilized
- Ancient proteins can be discovered in completely fossilized specimens

The half-life of DNA in bone: 50% less every 521 years

- 2012 study: Claims of extreme survival of DNA have emphasized the need for reliable models of DNA degradation through time.

- Study: analysis of mitochondrial DNA (mtDNA) from 158 radiocarbon-dated bones of the extinct New Zealand moa

- The average DNA half-life within this geographically constrained fossil assemblage was estimated to be 521 years for a 242 bp mtDNA sequence, corresponding to a per nucleotide fragmentation rate (k) of 5.50×10^{-6} per year.

Morten E. Allentoft, et al., 2012

DNA half-life

- Short fragments of DNA could be present for a very long time.
- At –58 C, nuclear DNA has degraded at least twice as fast as mtDNA
- This model predicts a half-life of 158,000 years for a 30 bp mtDNA fragment in bone
- Mitochondrial DNA degrades to an average length of 1 base pair after 6,830,000 years at −5 °C

Long-term persistence of bacterial DNA - Eske Willerslev, et al., 2003

- First study of DNA durability and degradation of a broad variety of bacteria preserved under optimal frozen conditions
- Twelve permafrost samples (0 to 8.1 Ma years old)
- Actinobacteria (break down matter in soil) are by far the most durable,
- Sequences of 120 bp and 600 bp could be reproducibly amplified from samples up to 400–600 Ka, and show an inverse relationship between PCR amplification efficiency and fragment length that is typical of ancient DNA
- Indicates a limit for PCR amplifiable DNA between 400 Ka and 1.5 Ma, beyond which DNA is either severely crosslinked or non-detectable.

DNA preservation

- Moreover, DNA preservation is also affected by other factors such as the treatment of the unearthed fossil like (e. g. washing, brushing and sun drying), pH, irradiation, the chemical composition of bone and soil, and hydrology.
- There are three perseveration diagenetic phases.
  - The first phase is bacterial putrefaction, which is estimated to cause a 15-fold degradation of DNA.
  - Phase 2 is when bone chemically degrades, mostly by depurination (loss of basepair site).
  - Phase 3 occurs after the fossil is excavated and stored, in which bone DNA degradation occurs most rapidly.

Fossil DNA Preservation

- To avoid contaminating the ancient DNA, specimens are handled with gloves and stored in -20 °C immediately after being unearthed.
- Bones are milled to a powder and treated with a solution before the polymerase chain reaction (PCR) process.
- The best time to extract DNA from a fossil is when it is freshly out of the ground as it contains six times the DNA when compared to stored bones.
- The temperature of the extraction site also affects the amount of obtainable DNA, evident by a decrease in success rate for DNA amplification if the fossil is found in warmer regions.

Use of aDNA damage as evidence of authenticity of aDNA

- If you retain only those DNA fragments that show a pattern characteristic of ancient DNA sequences (the replacement of Cyt bases by Thy bases at the ends of sequence strands), this will reduce the data by over 90% for each sample.

- This step is necessary, although extreme, as it will undoubtedly have resulted in the loss of many ancient, but undamaged, strands of DNA

Evo-Devo (1980-present): Molecular phylogenetics

- Evolutionary developmental biology
- All animals are built from essentially the same genes:
  Field of biology that compares the developmental processes of different **organisms** **to determine the ancestral relationship** between them, and to discover how developmental processes evolved
- Not until the 1980s and 1990s, however, when more comparative molecular sequence data between different kinds of organisms was amassed
- Change from genes to protein-centric perspective; move to gene switching/regulator perspective

DNA analysis for phylogeny after you get genome data

- External morphology (skeleton) is often an unreliable indicator of genetic relatedness.
- DNA is definitive method for phylogenetic relationships.
- DNA can produce timing of divergence of species from common ancestor by analysis of number of allele (gene variant) differences that have accumulated in DNA sequence, since these differences occur roughly as a function of time. This is basis of the "molecular clock".
- Zoological museums (Smithsonian; London NHM) created molecular labs to analyze their collections.
- Applications: paternity tests, wrongfully convicted prisoners, etc.

Molecular clock

- The molecular clock is the term for a technique that uses the mutation rate of biomolecules to deduce the time in prehistory when two or more life forms diverged.
- Data = nucleotide sequences for DNA, RNA, or amino acid sequences for proteins
- Neutral mutations (random changes) occur at a constant rate in a species
- Mutation rate = clock-like rate of molecular change; further back, more mutations
- Can compare number of mutations in 2 species to arrive at time of divergence
- Most phylogenies require that the molecular clock be calibrated against independent evidence about dates, such as the fossil record

Characteristics of Ancient Biomolecules

- **Ancient DNA:** Ancient DNA is normally heavily fragmented and chemically modified.

- After the death of an organism, DNA is initially degraded by normal endogenous nucleases (enzyme capable of cleaving DNA).

- This is soon followed by exogenous degradation processes, such as oxidation, hydrolysis (water damage), and background radiation, which alter the nitrogenous bases and cleave the sugar-phosphate backbone of the DNA molecules, leading to their destabilization and fragmentation.

Characteristics of Ancient Biomolecules

- There are four dominant types of aDNA damage:
  - (*a*) fragmentation,
  - (*b*) abasic sites (missing DNA bases),
  - (*c*) cross-linking (condensation reactions between DNA and proteins or sugars),
  - (*d*) miscoding lesions (base pair modifications leading to the incorporation of incorrect bases during DNA amplification).
- Fragmentation, abasic sites, and crosslinking all inhibit the amplification of aDNA, whereas miscoding lesions produce erroneous sequences that can significantly impact downstream analyses.

Characteristics of Ancient Biomolecules

- The number of aDNA fragments exponentially decreases with the increase of their length, as the random breakage of long molecules results in an accumulation of shorter ones.

- The rate of the fragmentation depends on different environmental factors:
  - temperature,
  - pH/high acidity,
  - water availability,

- It appears to be initially rapid, most likely due to high enzymatic activity, and followed by reduced rates over the long term.

Ancient DNA

- Cytosine conversion to uracil, a thymine analog, is the most common base modification (90% of damage)

- Postmortem aDNA damage patterns became a key criterion for distinguishing endogenous sequences from contaminant DNA

Ancient DNA preservation

- Particularly favorable environmental conditions for aDNA preservation include
  - low temperatures,
  - rapid desiccation (drying),
  - high salt concentration.

- These factors facilitate destruction and/or inactivation of DNA cleaving enzymes and reduce bacterial metabolic activity and water damage.

- Consequently, the oldest genomes sequenced to date come from specimens preserved under cold conditions.

Ancient DNA & temperature: heat matters

- Oldest aDNA has been sequenced from ice and permafrost ranging between 400 and 1.6 Ma in age.

- In contrast, the age of the oldest aDNA reads from the tropics is ~2 orders of magnitude lower

Ancient lipids

- Lipids are the organic solvent–soluble components of living organisms; fatty acids or their derivatives and are insoluble in water but soluble in organic solvents, i.e., oils, fats, waxes, and resins.

- They are inherently resistant to biodegradation and abiological decay compared to DNA and proteins, especially when they are protected within mineral or organic matrices.

Ancient lipids

- Entrapment in either organic or mineral matrices, e.g., sediment aggregates, pottery, bone, etc., reduces the loss of biomolecules by diffusion and limits microbial activity by impeding access to lipid substrates.

- The resistance of lipids to decay, combined with their persistence at the original site of deposition, makes them excellent candidates for use as biomarkers in molecular stratigraphic (i.e., chronological) investigations.

aDNA: Recovery & Sampling

Over the last 30 years, a tremendous amount of effort has been invested into maximizing

- aDNA recovery. The introduction of high-throughput sequencing is arguably the single most important contributor to the expansion of the field. The associated explosion in data generation prompted a transformation of the procedures for their analysis and interpretation.

- *Sampling.* The two key aims of aDNA retrieval are maximizing endogenous DNA content and minimizing contamination. Optimal sampling of ancient specimens is an important first step to achieve this.

Perseveration of aDNA in Bones

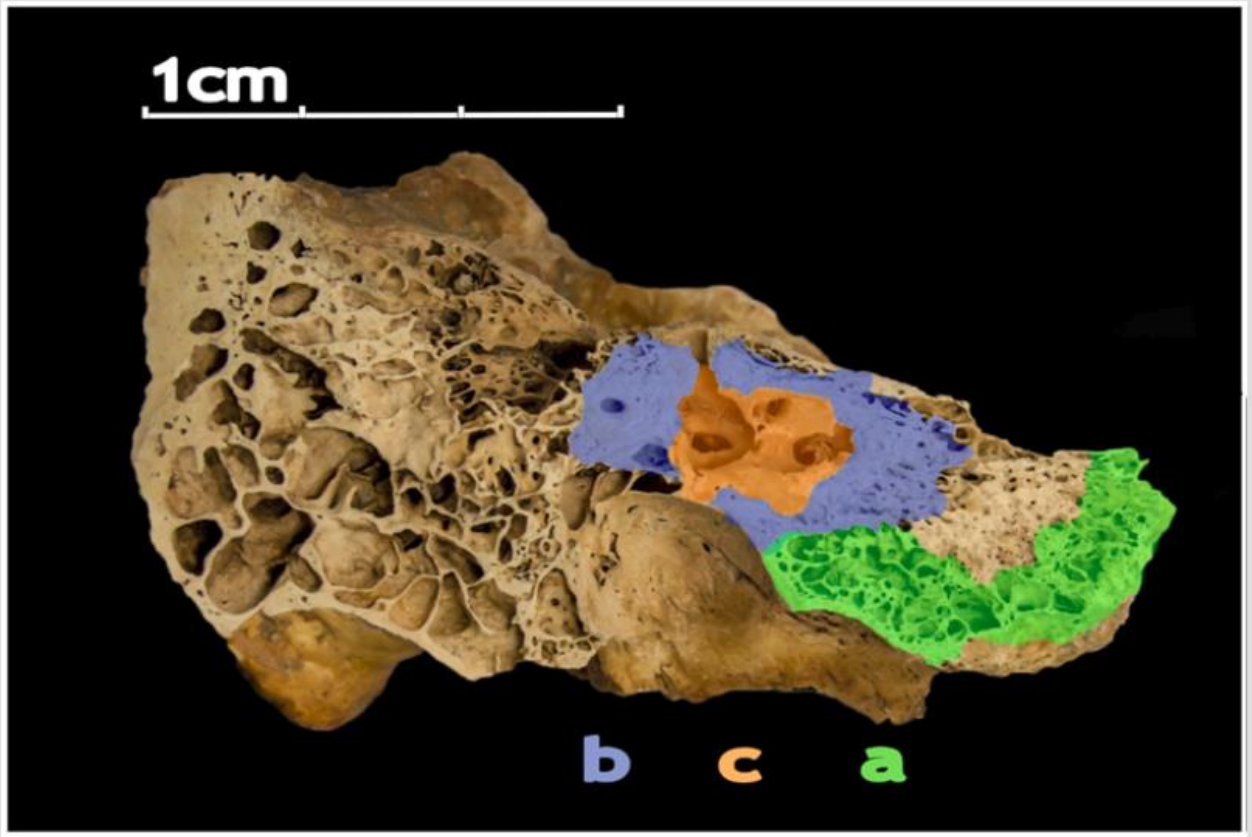
- DNA is stable in bones: the mineral hydroxyapatite binds to it and remains stable for long time.

- Hydroxyapatite-bound DNA is more resistant to decay and less susceptible to degradation by serum and nucleases, which may account for the long-term persistence of DNA in bone and tooth.

- Soft tissue does not preserve DNA for long.

A human petrous bone being analyzed at the Max Planck Institute for the Science of Human History in Jena, Germany.

Skull's Petrous Bone; best source of paleo DNA



The human petrous bone in the skull protects the inner ear structures. Though it is one of the hardest, densest bones in the body, some portions (such as the area in orange, protecting the cochlea) are denser than others. Possibly because the petrous bone is so dense, DNA within the petrous bone is better preserved than in other bones. In some cases, scientists have extracted more than 100 times more DNA from the petrous bone than other bones, including teeth. Credit: Pinhasi et al., 2015, PLOS ONE.

- An inch-long section of the human skull. Found near our ears, this pyramid-shaped portion of the temporal bone is nicknamed the petrous bone.
- The bone is very hard, possibly because it needs to protect fragile structures such as the cochlea, which translates sound into brain signals, and the semicircular canals

Ron Pinhasi, et al., 2015

Petrous bone

- The inner portion of the petrous part of the temporal bone in the skull, *pars petrosa*, and the cementum layer in tooth roots are considered to be the most favorable skeletal substrates for aDNA analysis owing to their high endogenous DNA contents
- It is the hardest and densest bone part in the mammalian body.
- Density of a bone is positively correlated with DNA preservation
- Perhaps because the petrous bone is so dense, it also is the bone in the body that best preserves DNA after a person dies.

Petrous bone

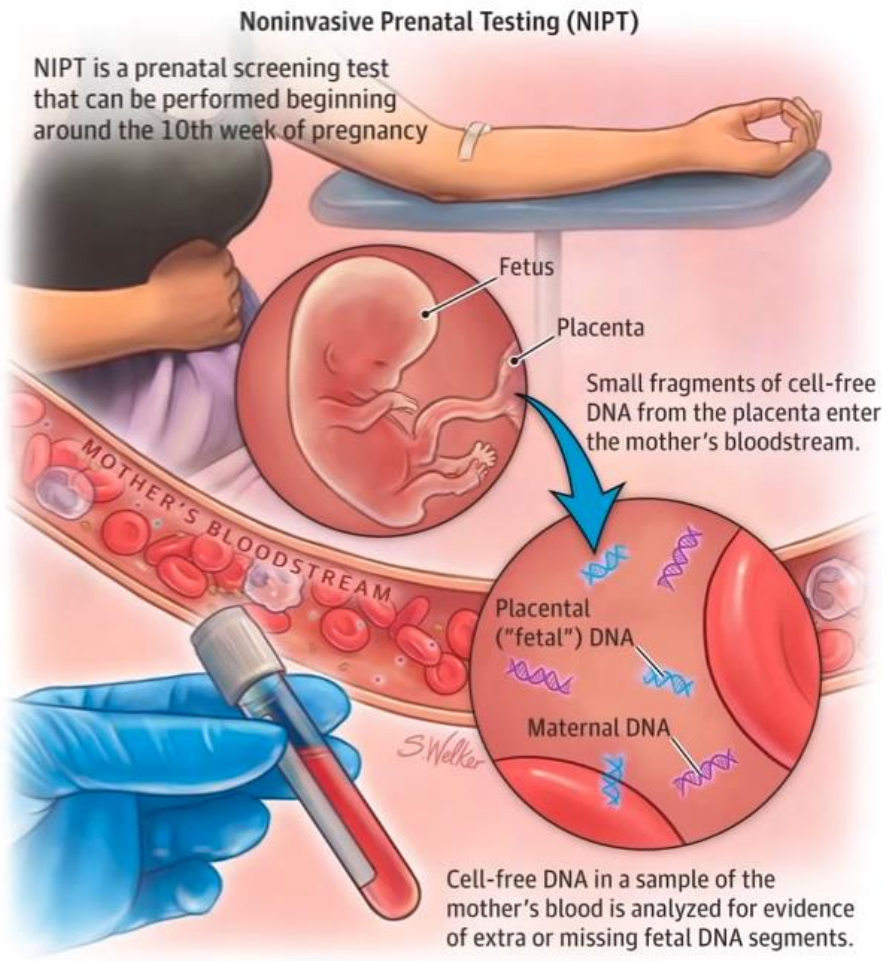
- Most fossil specimens contain only low (~ 1% or less) percentages of endogenous (original to organism) DNA.
- Petrous part of the temporal bone = contains most endogenous DNA
- Bone powder taken from the petrous bone yields on average up to 100 times more DNA than powder from other, softer bones.
- Teeth are second best; petrous bone gives 16x more DNA

Petrous bone & temperature

- The thermal history of a sample is the key factor influencing DNA survival.
- This is evident as many of the most successful aDNA studies utilized samples from permafrost regions
  - In non-petrous bones and teeth, endogenous DNA contents ranged from 0.3 to 21%,
  - while the levels for petrous bones ranged from 37 to 85%.
- So, it's been a real game changer for the field of ancient DNA.

PCR sequencing medical uses

- Prenatal testing
- Cancer detection
- Transplant rejection
- Detect pathogens
- Determine cancer treatment



JAMA. 2015;314(2):198

Methods: PCR and ancient DNA

- Since the late 1980s, researchers have been recovering and characterizing DNA from archaeological specimens using the PCR assay.
- PCR has many useful attributes for this type of work, but it is difficult to scale up to high-throughput, genome-wide analyses.
- It is challenging to apply PCR to the very short DNA fragments that are characteristic of ancient samples, especially when PCR is preferentially prone to recover chemically unaltered, modern contaminant DNA from excavators and laboratory workers.
- In the past decade, the development of next-generation sequencing technologies has driven a renewed rush of interest into the application of ancient DNA

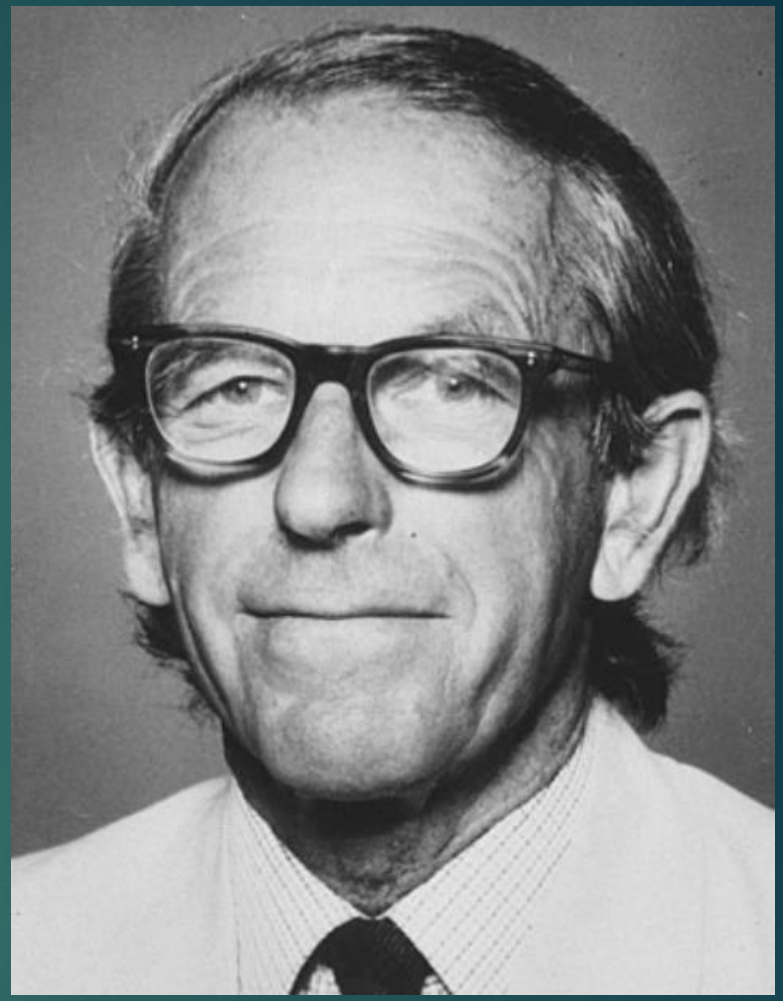
** Sequencing Methodologies

Genome sequencing

- DNA sequencing involves determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a DNA molecule.

- Sequencing an entire genome (all of an organism’s DNA) remains a complex task. It requires breaking the DNA of the genome into many smaller pieces, sequencing the pieces, and assembling the sequences into a single long "consensus."

Frederick Sanger (1918-2013)



- A pioneer of sequencing – 1st to sequence all amino acids in insulin
- A new method for "DNA sequencing with chain-terminating inhibitors" in 1977.
- Sanger is one of the few scientists who was awarded two Nobel prizes in chemistry, one for the sequencing of proteins in 1958, and the other for the sequencing of DNA in 1980

Genome sequencing: cloning

- In **Sanger sequencing**, the target DNA is copied many times, making fragments of different lengths.
- Fluorescent “chain terminator” nucleotides mark the ends of the fragments and allow the sequence to be determined.
- Sanger sequencing is still in wide use for the sequencing of individual pieces of DNA, such as fragments used in DNA cloning or generated through polymerase chain reaction (PCR).

Sanger

- The chain-termination method developed by Frederick Sanger and coworkers in 1977 soon became the method of choice, owing to its relative ease and reliability. When invented, the chain-terminator method used fewer toxic chemicals and lower amounts of radioactivity than the Maxam and Gilbert method.
- Because of its comparative ease, the Sanger method was soon automated and was the method used in the first generation of DNA sequencers.
- Primary method from the 1980s until the mid-2000s. Over that period, great advances were made in the technique, such as fluorescent labelling, capillary electrophoresis, and general automation.

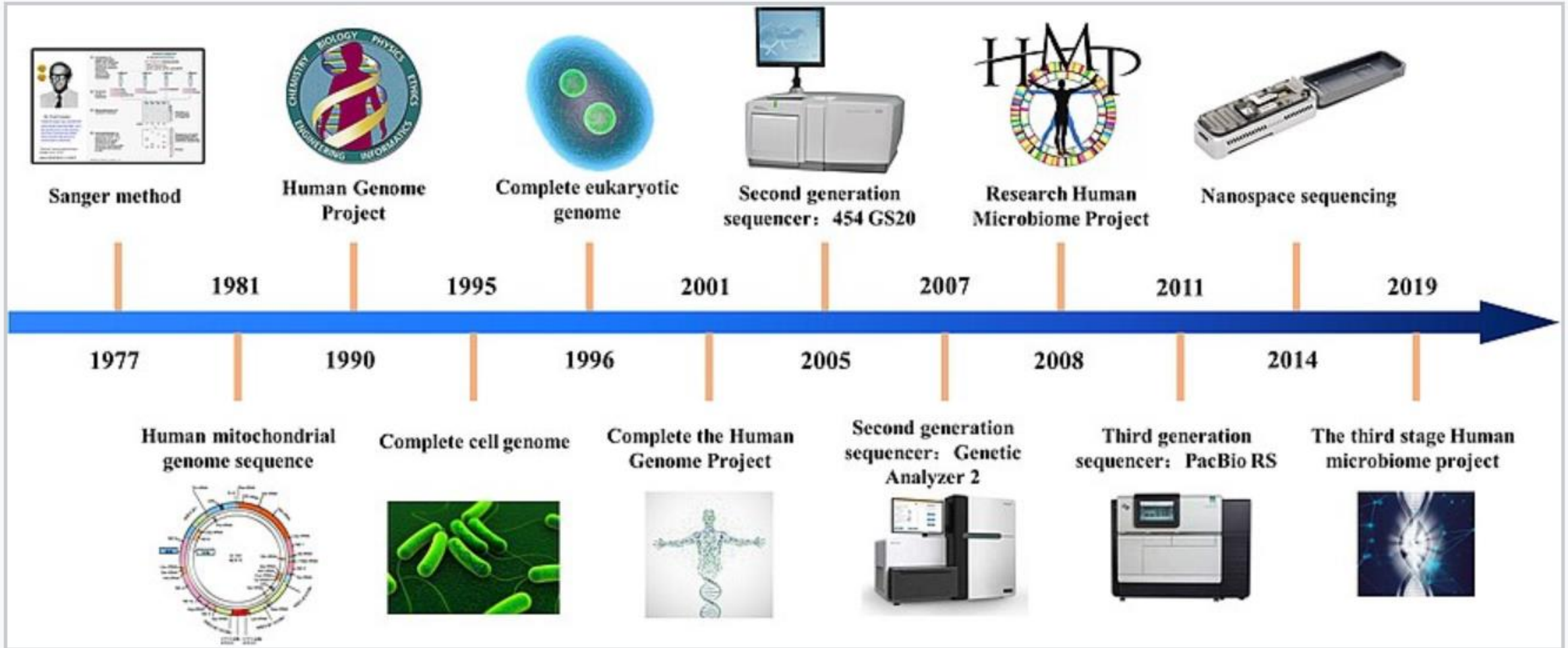
2nd generation sequencing (NGS)

- **Next-generation sequencing** techniques are new, large-scale approaches that increase the speed and reduce the cost of DNA sequencing.
- The "next-generation" or "second-generation" sequencing (NGS) methods, are named thus in order to distinguish them from the earlier methods, including Sanger sequencing.
- In contrast to the first generation of sequencing, NGS technology is typically characterized by being highly scalable, allowing the entire genome to be sequenced at once.
- Usually, this is accomplished by fragmenting the genome into small pieces, randomly sampling for a fragment, and sequencing it using one of a variety of technologies.

Next-generation sequencing

- There are a variety of next-generation sequencing techniques that use different technologies. However, most share a common set of features that distinguish them from Sanger sequencing:
  - **Massively parallel:** many sequencing reactions take place at the same time; entire genome is possible
  - **Micro scale:** reactions are tiny and many can be done at once on a chip
  - **Fast:** because reactions are done in parallel, results are ready much faster
  - **Low-cost:** sequencing a genome is cheaper than with Sanger sequencing
  - **Shorter length reads:** reads typically range from 50-700 nucleotides in length

High-throughput sequencing (HTS) methods



History of sequencing technology[51]

Next-generation sequencing

- Next-generation sequencing (NGS) has now replaced Sanger sequencing in the aDNA field.
- By rendering the previously required transfer of plasmid libraries and bacterial cloning unnecessary, NGS tremendously increases the amount of retrievable data.
- The NGS workflow can be summarized as follows:
  - first, building DNA sequencing libraries using DNA ligation technologies;
  - second, amplifying libraries using polymerase chain reaction (PCR);
  - third, performing massively parallel sequencing;
  - fourth, conduct downstream bioinformatics analyses.
- Leading company: The Illumina platform has largely outcompeted the various other commercial options, primarily owing to its massive output of short DNA reads.

Enrico Cappellini, et al., 2018

Illumina machines

	MiSeq	HiSeq	NovaSeq	Sanger
Reads (millions)	30	3,000	13,000	0.0004
Gigabases/day	7	500	4000	0.001

Illumina

- 90% of sequencing market
- Imaging-based method
- Many reads - millions to billions per run
- 300 to 600 bases per read
- High fidelity: >99.9% accuracy
- $1,000 human genome in 48 hours

Next-generation sequencing

- The two most widely used sequencing approaches today are
  - shotgun sequencing
  - target-enriched sequencing.

- With shotgun sequencing, the extracted DNA is directly converted into a digital sequencing library.

- In contrast, target-enrichment by hybridization, a procedure commonly known as capture, selects for DNA library fragments of interest with either DNA or RNA baits. It is particularly helpful when genotyping, detecting rare variants, and exome sequencing.

What is shotgun sequencing?

- In whole genome shotgun sequencing the entire genome is
  - broken up into small fragments of DNA for sequencing.
  - and then reassembling the sequence by looking for regions of overlap.

- These fragments are often of varying sizes, ranging from 2-20 kilobases (2,000-20,000 base pairs) to 200-300 kilobases (200,000-300,000 base pairs).

What is shotgun sequencing?

- These fragments are sequenced to determine the order of the DNA bases, A, C, G and T.

- The sequenced fragments are then assembled together by computer programs that find where fragments overlap.

- You can imagine shotgun sequencing as being a bit like shredding multiple copies of a book (which in this case is a genome), mixing up all the fragments and then reassembling the original text (genome) by finding fragments with text that overlap and piecing the book back together again.

What are the advantages of shotgun sequencing?

- Whole genome shotgun sequencing uses a fraction of the DNA that clone-by-clone sequencing needs.

- Whole genome shotgun sequencing is particularly efficient if there is an existing reference sequence. It is much easier to assemble the genome sequence by aligning it to an existing reference genome.

- Shotgun sequencing is much faster and less expensive than methods requiring a genetic map.

A Sequencing Timeline

	Samples/person/week	Average read length	Total bps per week
1977 Sanger and Maxam-Gilbert sequencing techniques developed	4	X 50 bp =	200 bp
1980 M13 vector developed for cloning, significantly improved computer technology	20	X 100 bp =	2,000 bp
1990 Improved sequencing enzymes, fluorescent dyes developed, robotics used for high throughout	60	X 300 bp =	18,000 bp
1997 *Sacromycetes Cerevisiae* yeast fungus genome sequenced	180	X 500 bp =	90,000 bp
1999 *C. elegans* nematode, Human chromosome 22, &~20 bacterial genomes	500	X 650 bp =	325,000 bp
2000 *Drosophila melanogaster*, *Homo sapiens*, *Arabidopsis thaliana* flower	5000	X 600 bp =	3,000,000 bp

Comparison of high-throughput sequencing methods[78][79]

Method	Read length	Accuracy (single read not consensus)	Reads per run	Time per run	Cost per 1 billion bases (in US$)	Advantages	Disadvantages
Single-molecule real-time sequencing (Pacific Biosciences)	30,000 bp (N50); maximum read length >100,000 bases[80][81][82]	87% raw-read accuracy[83]	4,000,000 per Sequel 2 SMRT cell, 100–200 gigabases[80][84][85]	30 minutes to 20 hours[80][86]	$7.2-$43.3	Fast. Detects 4mC, 5mC, 6mA.[87]	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (Ion Torrent sequencing)	up to 600 bp[88]	99.6%[89]	up to 80 million	2 hours	$66.8-$950	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (454)	700 bp	99.9%	1 million	24 hours	$10,000	Long read size. Fast.	Runs are expensive. Homopolymer errors.

Nanopore Sequencing	Dependent on library preparation, not the device, so user chooses read length (up to 2,272,580 bp reported[93]).	~92–97% single read	dependent on read length selected by user	data streamed in real time. Choose 1 min to 48 hrs	$7–100	Longest individual reads. Accessible user community. Portable (Palm sized).	Lower throughput than other machines, Single read accuracy in 90s.
GenapSys Sequencing	Around 150 bp single-end	99.9% (Phred30)	1 to 16 million	Around 24 hours	$667	Low-cost of instrument ($10,000)	
Chain termination (Sanger sequencing)	400 to 900 bp	99.9%	N/A	20 minutes to 3 hours	$2,400,000	Useful for many applications.	More expensive and impractical for larger sequencing projects. This method also requires the time-consuming step of plasmid cloning or PCR.

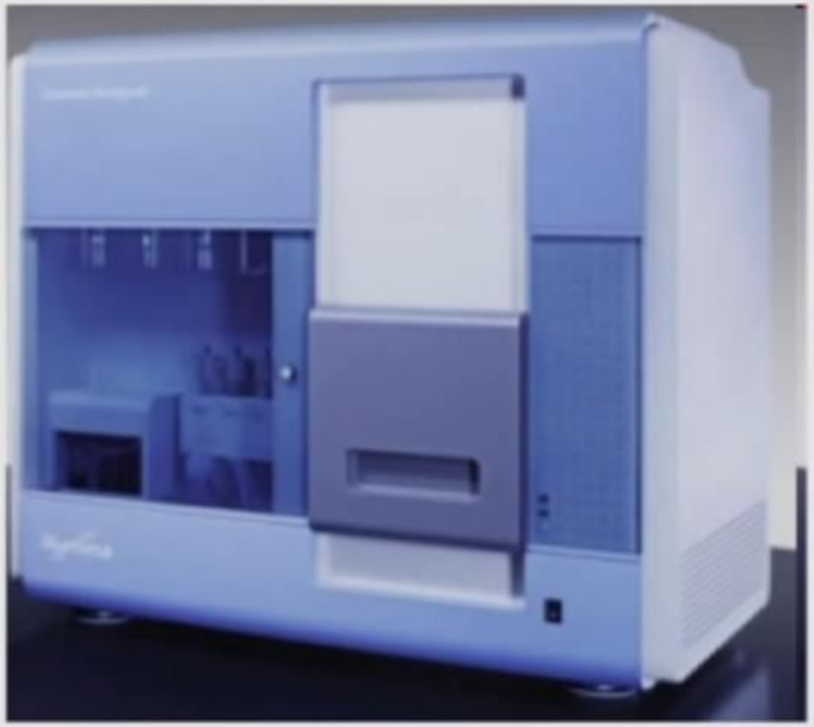
2006 DNA Next generation sequencing lab

Lawrence Berkeley Natl Lab./SPL

By 2006, DNA sequencing required much less manpower.

Next generation sequencing

-500 million bp; -500 bp reads



-95 billion bp; -150 bp reads

Modern DNA sequencing is 100s of times faster than machines 10 years ago;

454 Machine

PCR Platforms

2004

ABI 3730
~10^2 fragments per run

2005

454 GS20
~10^5 fragments per run

2009

Illumina GA2
~10^8 fragments per run

2016

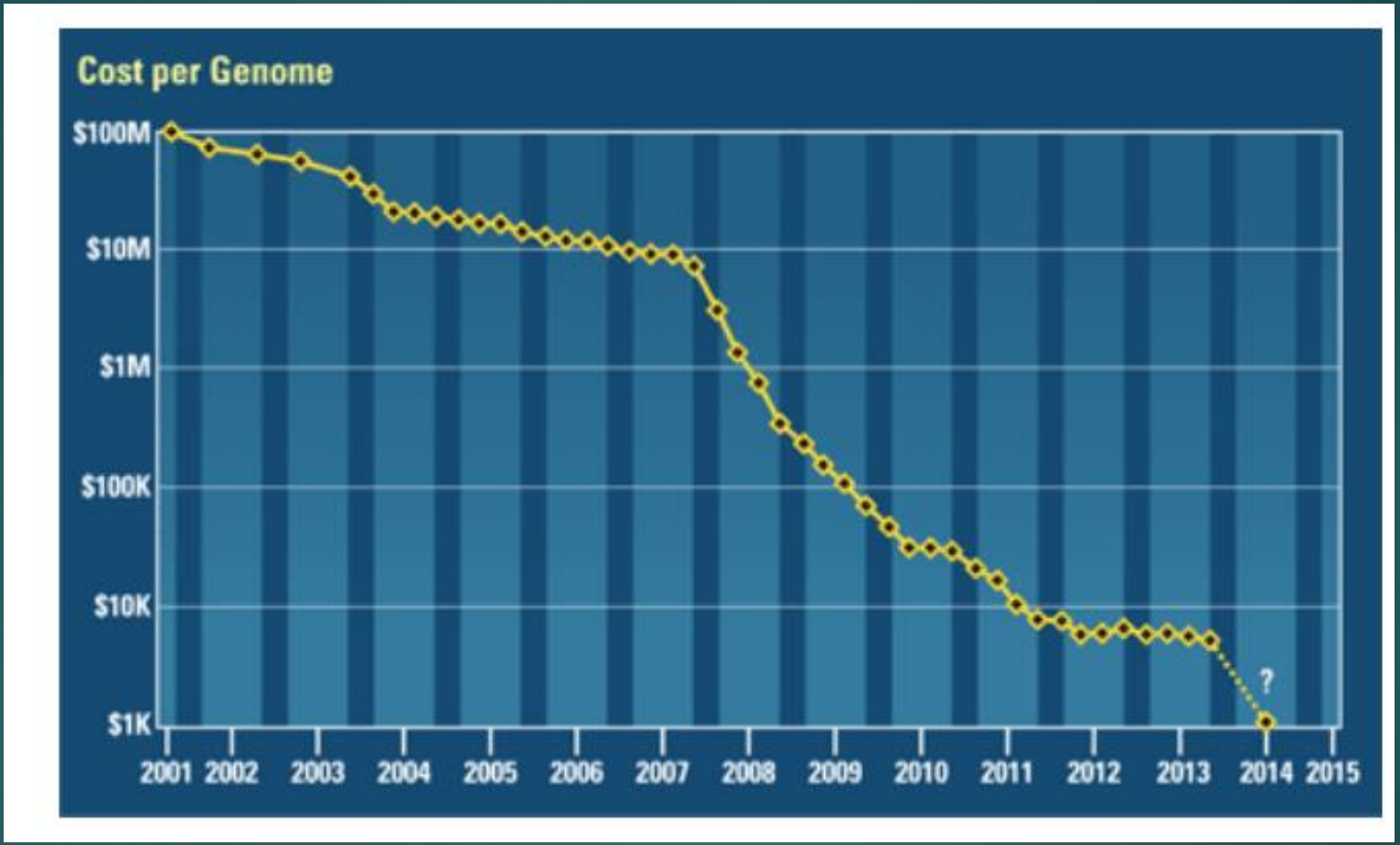
Illumina hiseq X
6 x10^9 fragments per run

Does Illumina Have the First 1,000 Genome?

Cost of machine = $1 M

The Illumina platform has largely outcompeted the various other commercial options, primarily owing to its massive output of short DNA reads

Price for whole genome sequencing has dropped in the past decade from about 22 million to as little as $1,000 per human genome



Long fall: After many years of decline, the cost of sequencing a genome had leveled off, but may dive again (dashed line) if Illumina's promise of a $1000 genome holds up.

Next-generation sequencing

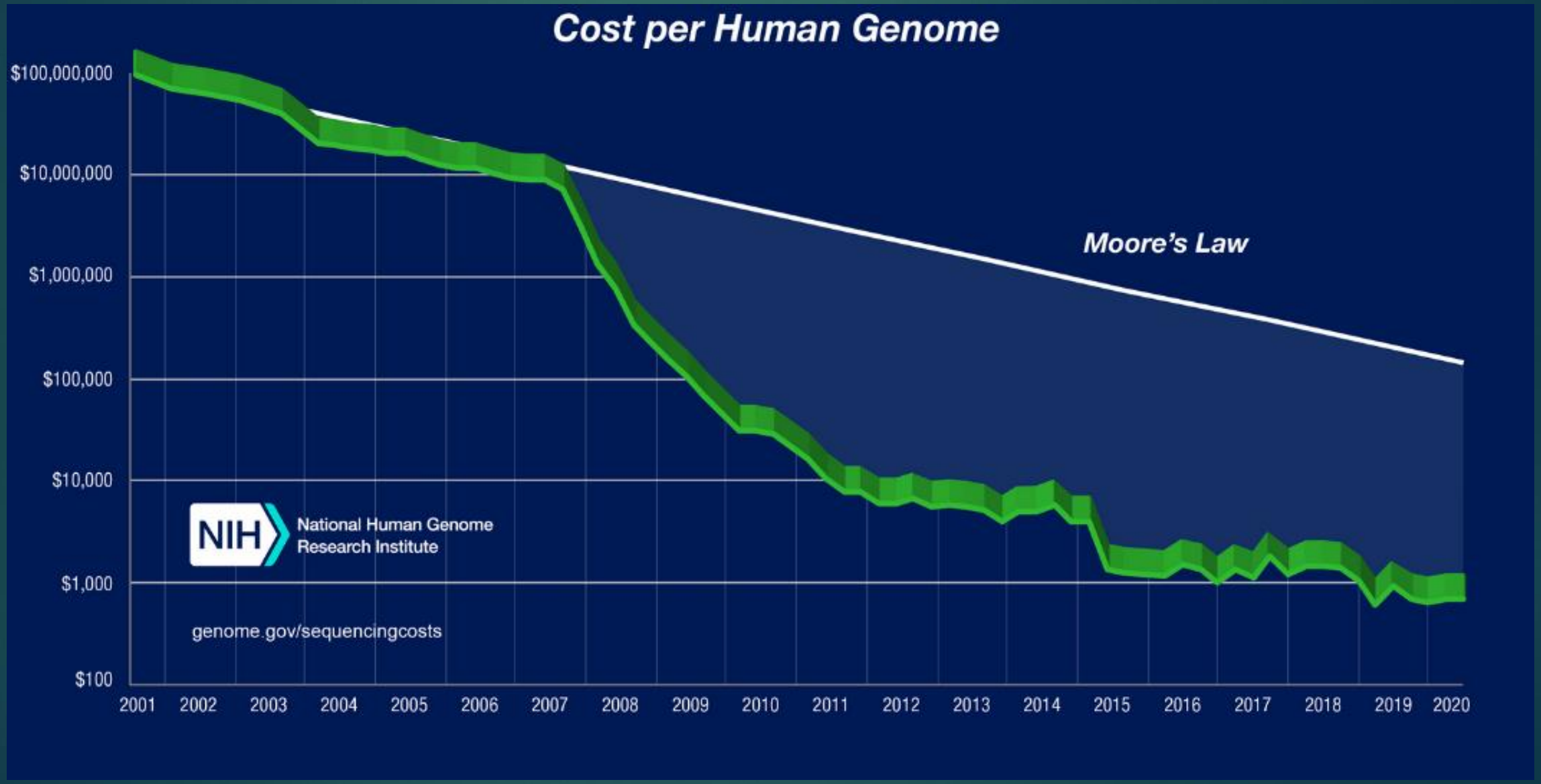
- The ancient genome(s) can finally be used for downstream analyses to infer evolutionary histories.

- These analyses often include exploratory statistical approaches, such as
  - principal-component analysis
  - latent class modeling,
  - demographic estimation methods

X-fold: Meaning of x-coverage = number of times bp has been analyzed

- X coverage (or x-fold coverage) is used to describe the sequencing depth. Refers to number of times a single bp is scanned.
- For example, if your genome has a size of 10 Mbp and you have 100 Mbp of sequence in data that is assembled to said 10 Mbp genome, you have 10x coverage.
- The word "coverage" has several different meanings. First, it can represent the same concept as the words "folded", "depth", and "redundancy", which is the number of times the *average* base position is scanned by data (a read, a clone, etc.) Similar to example above, 40Mb of sequence data for a 4Mb bacterial genome would be 10X or 10-fold "coverage".
- Clinical grade (30X coverage) = the gold standard of sequencing quality.

Now under $ 600 occasionally; Veritas Genetics: $1000

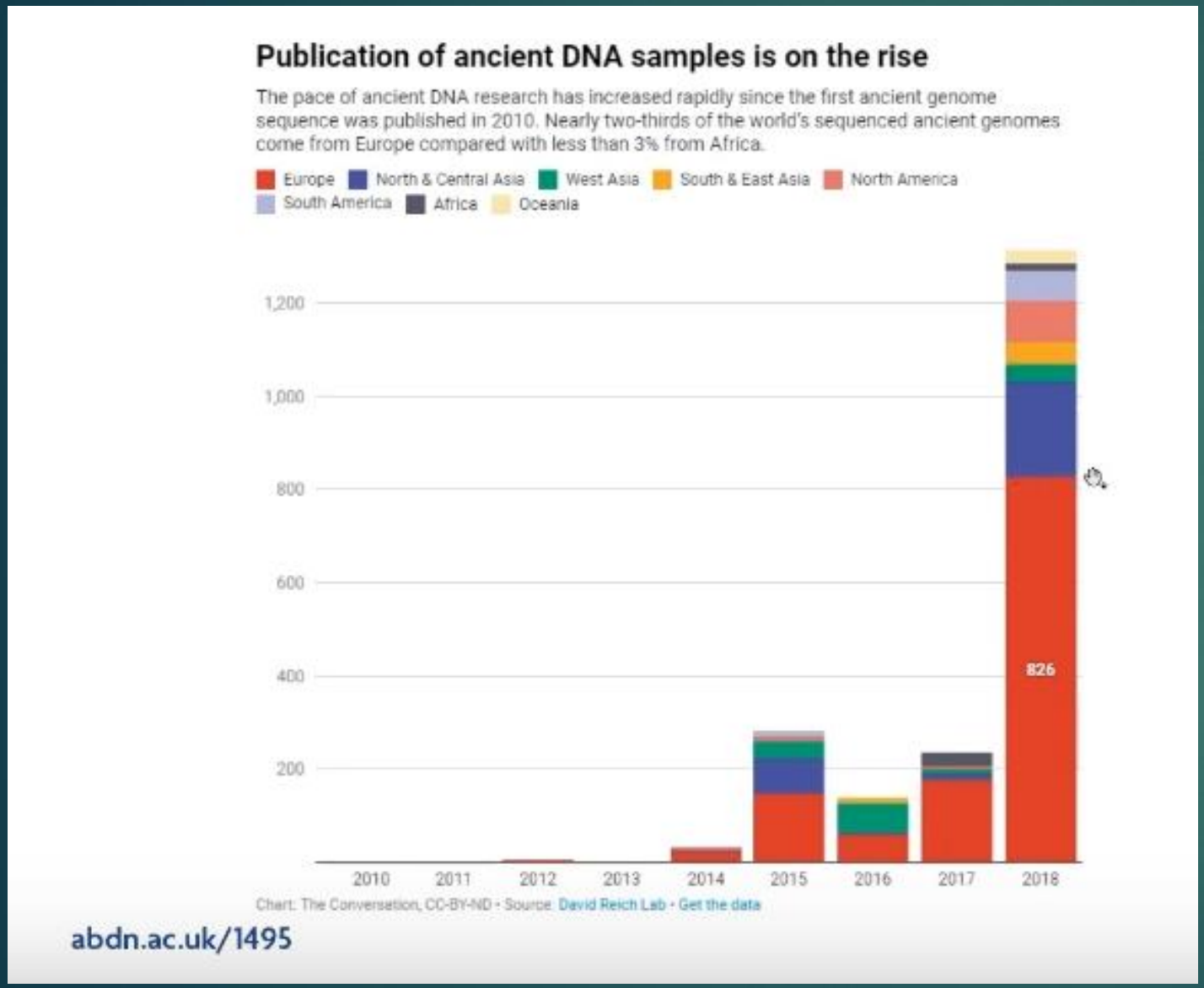


Cost has declined

- The first human genome sequence was estimated to cost 2.7 billion dollars in 2001.
- 2015, $2000 per full genome; Veritas Genetics, $1000 in 2016; $699 in 2019
- 23andMe and Ancestry: use genotyping, which looks at 1% of genome; tests only 3 BRCA genes for breast CA (ones commonly found in individuals of Ashkenazi Jewish descent); Veritas tests for 1000 BRCA mutations
- One recent analysis found that from 2016 through 2017, more than 7,500 mutations were reclassified, most of them from “pathogenic” or “likely pathogenic” to “unknown” or “conflicting significance.”
- Whole genome = 150 gigabytes of data

Ethics

- Ethics: One key issue is the ownership of an individual's DNA and the data produced when that DNA is sequenced
- *Moore v. Regents of the University of California* (1990) ruled that individuals have no property rights to discarded cells or any profits made using these cells (for instance, as a patented cell line). However, individuals have a right to informed consent regarding removal and use of cells. Regarding the data produced through DNA sequencing, *Moore* gives the individual no rights to the information derived from their DNA
- In May 2008, the Genetic Information Nondiscrimination Act (GINA) was signed in the United States, prohibiting discrimination on the basis of genetic information with respect to health insurance and employment.



- Increased publication rate since 2010.
- 75% of publications are of European genomes, only 3% from Africa
- By 2015, Reich lab was publishing more than 50% of all world's aDNA genomes
- Now have 10,000 genomes from 1000 current human populations: mainstay of modern study of human variation

PCR: Waking the Dead

- The advent of PCR in the 1980s made ancient DNA (aDNA) sequencing a reality,
- But early attempts to sequence human aDNA were frustrated by sample contamination and degradation.
- In 2010, the first nuclear draft sequence of a Neanderthal genome heralded a revolution in paleogenomics, advancing our understanding of the relationships between extinct and extant hominin lineages and how modern humans spread throughout the world.
- Ancient DNA research has been limited only by the technology, and never by a lack of interesting questions to be asked.
- The first aDNA studies using soft tissues from museum specimens were hampered by depurination and fragmentation in the sequenced DNA.

Waking the Dead

- By the late 1980s, it was possible to extract DNA from ancient bone, but the limited throughput of the Sanger sequencing technology, and the absence of human reference genomes for comparisons or filtering, made the detection of genuine nuclear aDNA sequences challenging.

- While the field moved forwards with studies of plants and non-human animals, hominins were somewhat neglected, until next-generation sequencing revolutionized the genomics field as a whole.

- February 2010, Rasmussen, et al.: first ancient human genome sequence for an extinct Paleo-Eskimo.

Waking the Dead

- In their 2010 study, Green et al. generated libraries from three Neanderthal bones from Croatia, dating to >38,000 years ago, fine-tuned these to screen out contamination from microorganisms and modern humans, and sequenced them with a combination of 454 and Illumina technologies, combining the three individuals into a 1.3× coverage genome.

- Comparisons with the human and chimpanzee genomes allowed the identification of Neanderthal sequences, leading to extensive new inferences about hominin molecular evolution, adaptation and — perhaps most controversially — gene flow between hominin groups.

Waking the Dead

- The Neanderthal genome shared more genetic variants with present-day Europeans and Asians than with Africans, suggesting some gene flow after the divergence of these lineages of modern humans.

- Genomic segments with high similarity to Neanderthal DNA were detected in present-day non-African genomes, providing direct evidence for this introgression (and allowing estimation of the time when it occurred).

Waking the dead

- 2012: To tackle the problem of limited endogenous aDNA quantities, Meyer et al. developed a single-stranded DNA library preparation method for a Denisovan sample.
- Their approach substantially increased the number of ancient molecules that could be incorporated into the DNA sequencing libraries, thereby yielding enough DNA sequence to obtain the first high-quality ancient genome, with 30× coverage of a single individual.
- This study provided further evidence for hominin admixture.
- DNA capture technologies have revolutionized our understanding of human disease and their introduction into the paleogenomics field enabled the study of polymorphisms present in tens or hundreds of ancient genomes.

Waking the dead

- In the first large-scale study by Haak et al. in 2015, capture technology was used for the analysis of 394,577 polymorphisms in 69 European individuals dating from 8,000–3,000 years ago, allowing the authors to make conclusions about
  - population movements and turnover during the Neolithic period and
  - the spread of Indo-European languages into Europe.

- With rapid technological advances and many questions already tackled, the limiting factor may now become the availability of suitable samples, in itself a potentially controversial topic for many reasons.

Modern Human Ancestry and Demography

- Sequencing the genome of the Denisovans has shown that interbreeding occurred between them and the ancestors of present-day Oceanian peoples.

- This and other archaic introgression events may have helped modern humans to adapt to local environmental conditions, such as high altitudes in Tibet, and contributed to a wide range of modern human phenotypic traits.

Modern Human Ancestry and Demography

- Much of the aDNA work to date has focused on the evolution and global dispersal of anatomically modern humans.

- One particularly fruitful research area has been focusing on the admixture history and on the migration routes that have given rise to the present-day patterns of human genetic diversity.

Modern Human Ancestry and Demography

- Genome of the 24,000-year-old Mal'ta child in south-central Siberia showed strong genetic affinities with both western Eurasians and Native Americans, suggesting a dual ancestry of the First Americans.

- The genome of a 12,600-year-old individual from the Anzick culture revealed closer genetic ties to Native Americans than to Europeans, ruling out a cross-Atlantic European origin for the Paleo-Indian Clovis culture in North America.

Modern Human Ancestry and Demography

- Recently, the genome of an 11,500-year-old individual from interior Alaska revealed that the First Americans derive from a single source population and that they were most likely established in Beringia as early as 20,000 years ago.

- These and similar studies have collectively demonstrated that much of the human genetic diversity we see today was created by migration and admixture events during human (pre)history.

Species Extinctions

- While the Quaternary Megafaunal Extinction is very well recorded in the paleontological record, its causes remain hotly debated, with climate change and human overkill as the two principal candidates.

- Ancient DNA has shed new light on this debate. Using ancient mitochondrial DNA from megafaunal fossil remains spanning the past 50,000 years, Lorenzen et al. showed that
  - while climate may have been the main extinction driver for some megafaunal species, e.g., the Eurasian musk ox and woolly rhinoceros,
  - it was the combined effects of climate and anthropogenic activities that led to the demise of others, e.g., the Eurasian steppe bison and wild horse.

- A follow-up study by Cooper et al. focusing on nuclear DNA regions, reported similar findings: abrupt warming events brought about by interglacial periods have caused repeated population-level turnovers and created faunal metapopulations highly vulnerable to the subsequent human impact.

Animal and Plant Domestication and Exploitation

- The domestication of animals and plants over the past 11,500 years has offered invaluable insights into positive selection and the associated environmental adaptations.

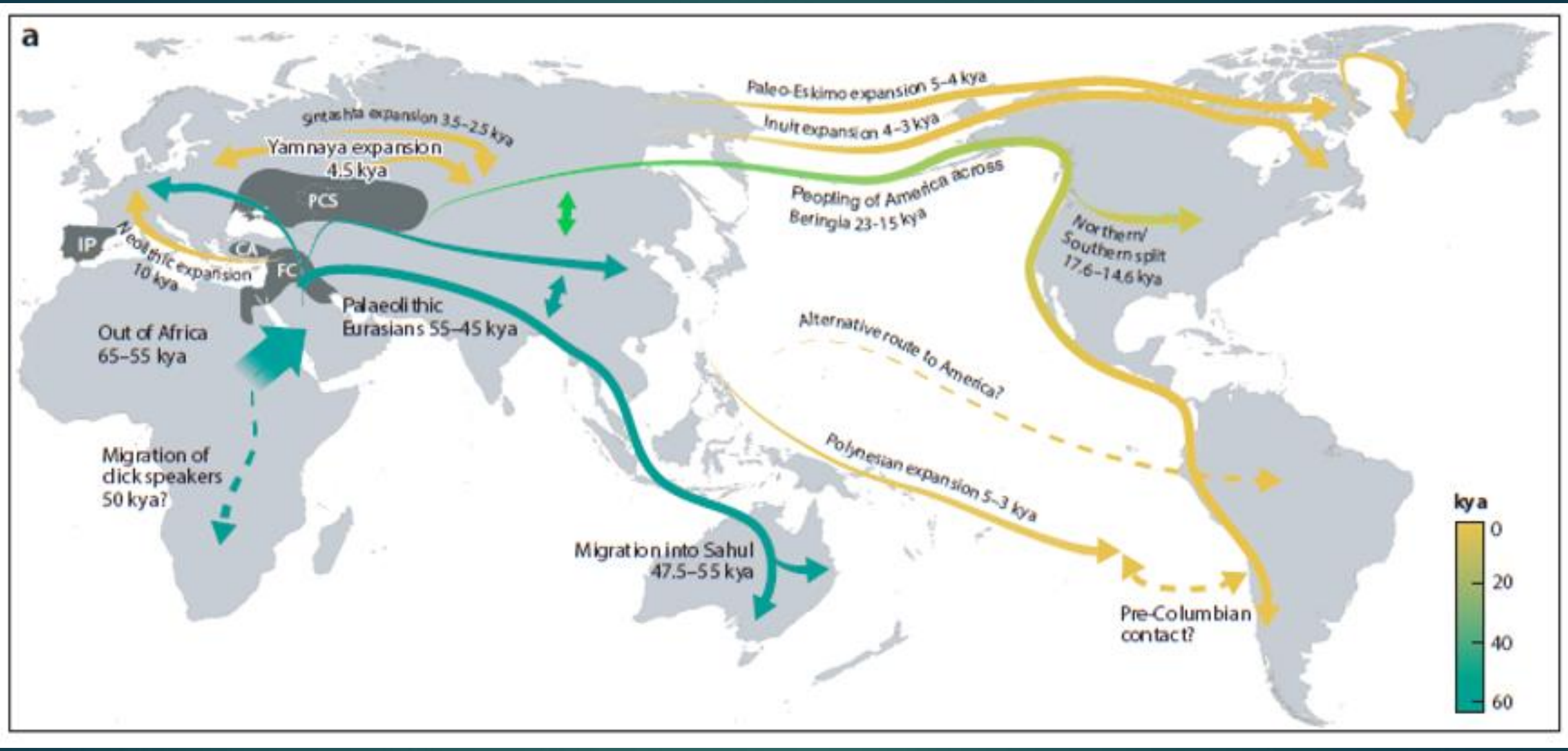
- Studies of aDNA in domestication in various taxa such as pigs, dogs, and chickens; arguably it is the horse that has received most attention

- Horse domestication: this process imposed positive selection on a range of genes involved with cognition, physiology, and locomotion; led to major human migrations and spread of pathogens.

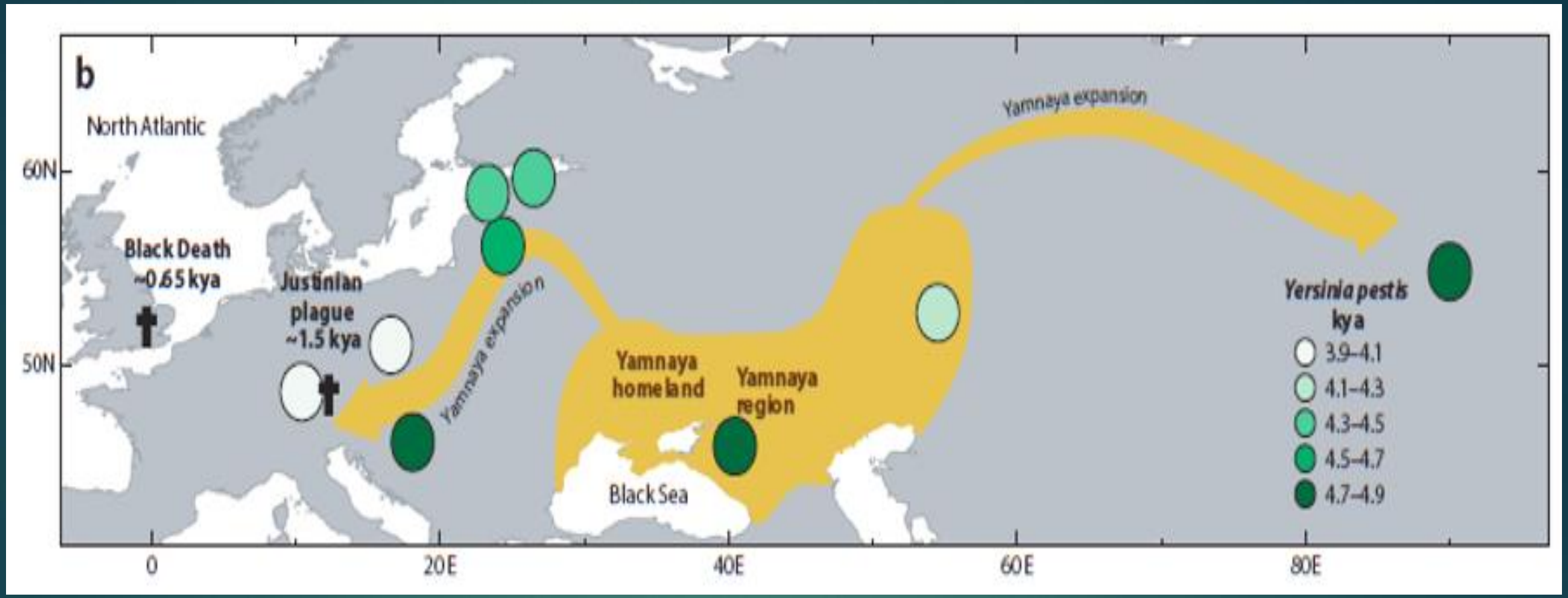
Human Migrations and Plagues

- (*a*) Migrations of modern humans as revealed by genomic data.
  - Many of the depicted migration events were only possible to infer using ancient DNA data sets (e.g., Neolithic, Yamnaya, and Sintashta expansions in Western Eurasia)

- (*b*) Spread of the Bronze Age plague. Genomic evidence of the plague-causing bacterium *Yersinia pestis* was isolated from ancient human remains.
  - Chronology and geography correlate with the expansion of peoples of the Yamnaya culture, as inferred from ancient genomes, suggesting diffusion of the disease through these prehistoric migrations.
- Higher numbers of deleterious mutations; and resulted in a net loss of genetic diversity over the last two millennia



Major human migrations, *(a)* Migrations of modern humans as revealed by genomic data. (e.g., Neolithic, Yamnaya, and Sintashta expansions in Western Eurasia): Abbreviation: CA, Central Anatolia; FC, Fertile Crescent; kya, thousands of years ago; IP, Iberian Peninsula; PCS, Pontic–Caspian



(b) Spread of the Bronze Age plague. Circles indicate the geographic locations and the age of sites where genomic evidence of the plague-causing bacterium *Yersinia pestis* was isolated from ancient human remains.
Chronology and geography correlate with the expansion (*arrows*) of peoples of the Yamnaya culture, as inferred from ancient genomes, suggesting diffusion of the disease through these prehistoric migrations.

Ancient Pathogens and Microbiomes

- Shotgun sequencing of ancient skeletal remains can also reveal genetic information about the microorganisms originally associated with their host, from specific pathogens to entire microbiomes.

- Genetic analysis of the origin and evolution of some of the deadliest pathogens in human history: the Spanish flu and the bubonic plague, the H1N1 influenza virus and *Yersinia pestis*, respectively.
- Rasmussen et al. showed that
  - bubonic plague had been widespread~3,000 years before any known written record

- The presence of cell wall biomarkers of *Mycobacterium tuberculosis*, namely mycolic acids, and the detection of *M. tuberculosis* aDNA in ancient individuals are seen as complementary evidence for ancient tuberculosis.

Microbiomes

- Deep sequencing of dental calculus and coprolites:
  - identified bacteria typical of both oral and gut microbiomes in archaic, Neolithic, and medieval humans.
- These studies suggest that major dietary shifts in human history:
  - such as neolithization and the Industrial Revolution, have
  - caused a marked decrease in microbiome diversity
  - the rise of microbial taxa linked to chronic diseases.
  - Genetic study evolution of human health and disease.

Future Directions: Large-Scale, High-Coverage Genome Panels

- The field of ancient biomolecules is likely to take numerous new directions over the coming years.

- Large-scale ancient genomic projects have so far relied either on low-coverage genome sequences (~1X average coverage) or targeted capture of common genomic variants, owing to the prohibitive costs of generating high-coverage ancient genomic data.

- An important next step in aDNA research will be to routinely sequence large numbers of high-coverage genomes spanning larger spatial and temporal scales.

Future: Deep-Time Phylogenetics

- Ancient proteins can survive considerably longer than aDNA.

- Consequently, paleoproteomics has the potential to provide access to genetic evidence from epochs and geographic areas incompatible with aDNA preservation and enable investigation into deep-time evolution, which has so far been intractable for molecular phylogenetics.

70% of aDNA research uses M. Meyer's 1.24 million SNPs microarray

In-Solution Enrichment Revolution in Ancient DNA

Problem: Ancient DNA is expensive because so little is human (often <1%)
Solution: Enrich for positions that are informative about human variation

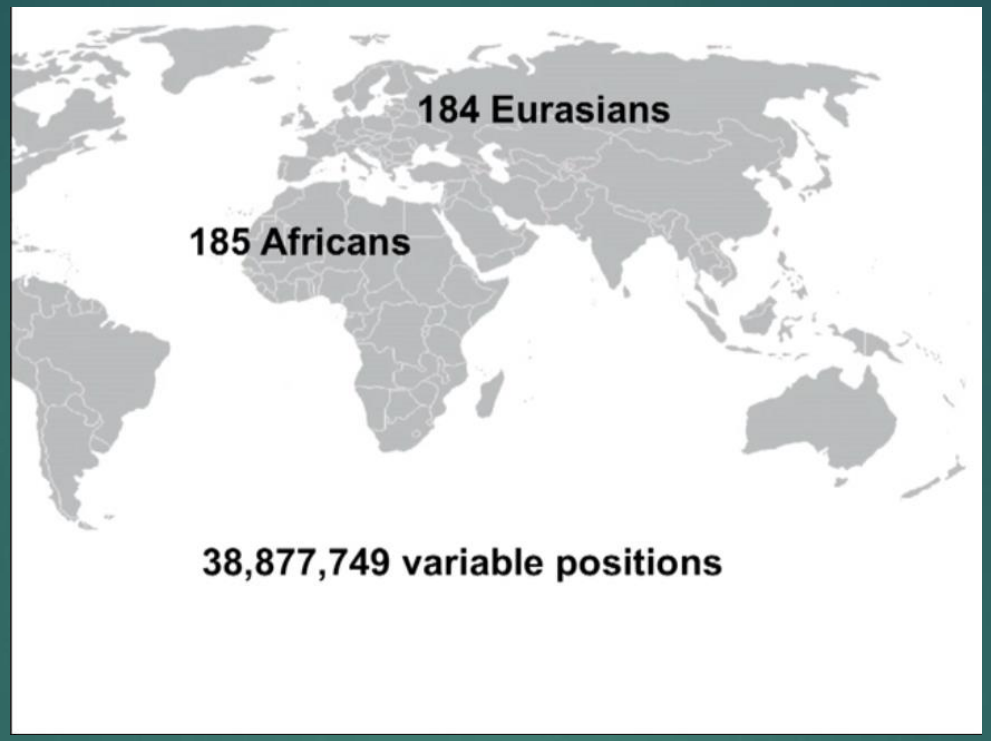
We worked with Matthias Meyer to order custom microarrays targeting ~1.24 million SNPs ('1240K')

Our design if generically available could compete with the microarray company's in-solution enrichment products as it allows probes to be replicated and used in many experiments instead of just one. The company agreed to provide arrays with this design under the condition we use them for aDNA only and not secondarily distribute the reagent.

We fully published the probe design with our first paper so others could replicate them (Fu et al. *Nature* 2015). For example, Johannes Krause's group separately ordered the 1240K targeting set.

The 1240K design has been responsible for >70% of whole genome human data in the aDNA literature

2008: 1000 Genome Project



- **The 1000 Genomes Project**, launched in January 2008, is an international research effort to establish by far the most detailed catalogue of human genetic variation; 38 M variations in 2 groups

2015: 1000 Genome Project

- Sequenced the full genomes of 2,504 people, from 26 different populations across the Americas, Eurasia, and Africa.
- This took seven years.
- In total, the project identified 88 million variants in the human genome. The typical person's DNA was dotted with 4 to 5 million of them.
- Mostly single-nucleotide polymorphisms (SNPs) variants, but your average Joe also carries a couple of thousand structural variants, such as deletions and insertions.
- Data set accounts for more than 99 percent of SNPs and 85 percent of larger variants

1000 Genome Project

- People in Africa exhibited the most variable genomes; Africa houses the oldest human populations, the ones who have had the most time for genetic drift (allele change by chance) to create variety.

- Other populations underwent bottlenecks in genetic diversity as small founder groups emigrated from Africa carrying only a smidgen of the total population diversity with them.

- https://catalog.coriell.org/1/NHGRI/Collections/1000-Genomes-Collections/1000-Genomes-Project

Paleogenetics, Part 3
End