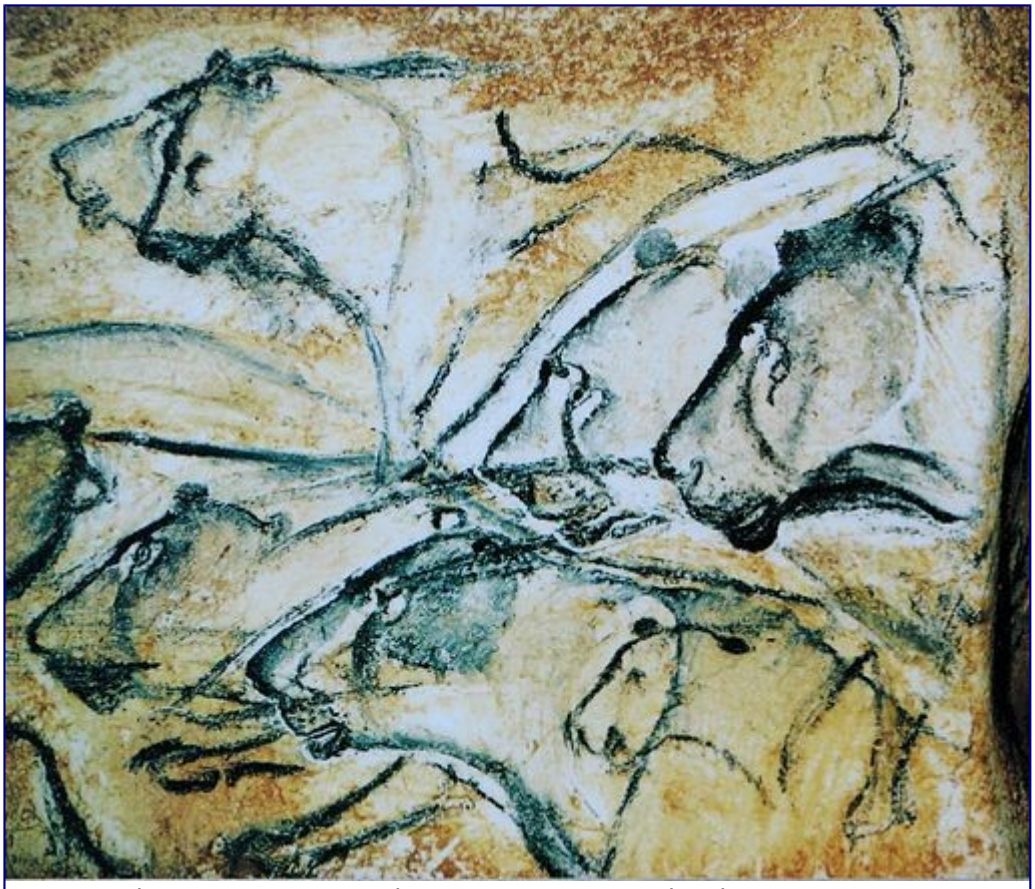
Being Human



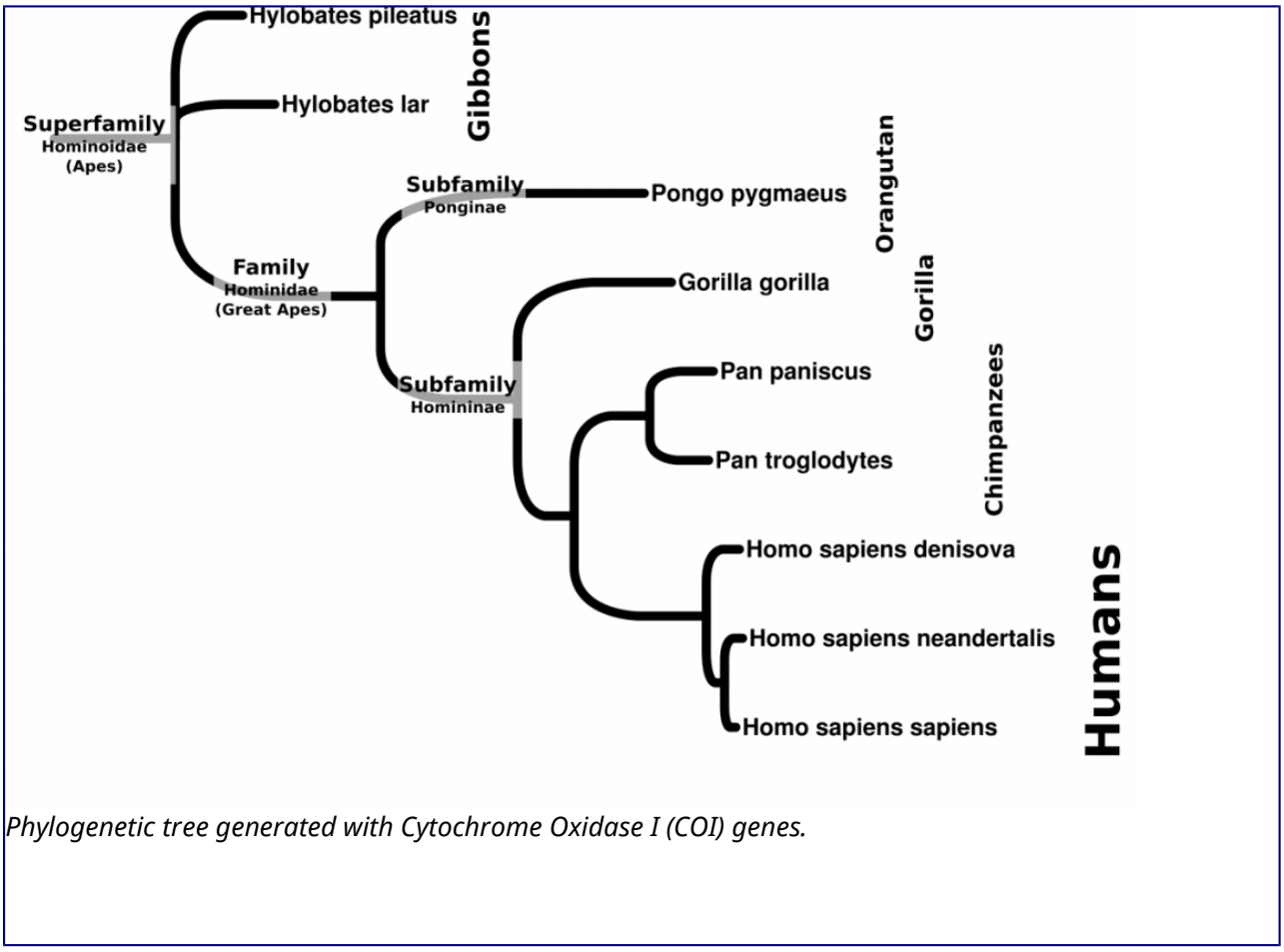
Drawings dating to approximately 30,000 years ago in the Chauvet Cave

What constitutes being human? Many will point at a cultural identity and leaving long-standing remnants of that culture. Such prehistorical artifacts like cave drawings and tools provide an anthropological framework for identifying what it is to be human, but the biological identity remains locked in the history of our DNA.

Spear points of the Clovis Culture in the Americas dating to approximately 13,000 years ago.

Jeremy Seto jseto@citytech.cuny.edu

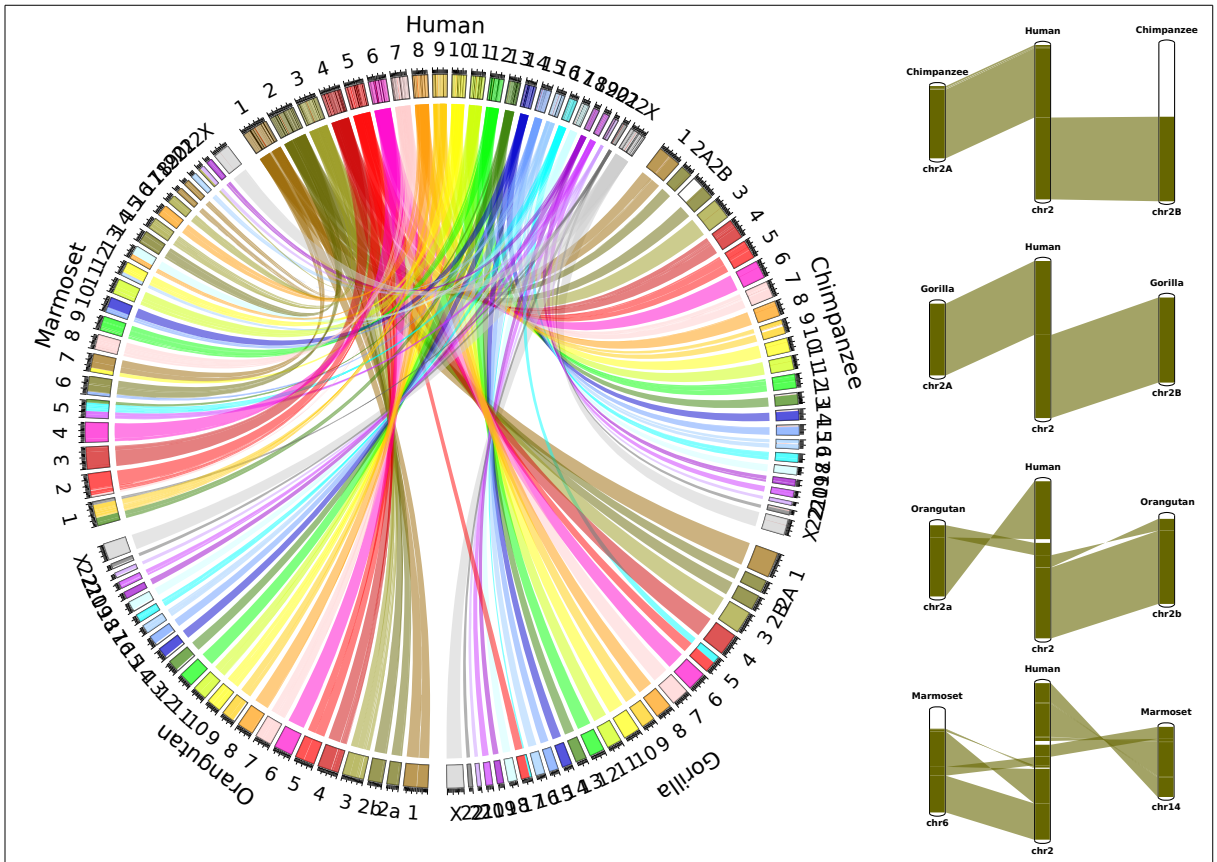
The Great Apes



Phylogenetic tree generated with Cytochrome Oxidase I (COI) genes.

Homo sapiens represent a branch of primates in the line of Great Apes. The family of Great Apes consists of four extant genera: *Homo, Pan, Gorilla, Pongo*. Karyotype analysis (Yunis *et al.*, 1982) reveals a shared genomic structure between the Great Apes. While humans have 46 chromosomes, the other Great Apes have 48. Molecular evidence at the DNA level indicates that Human Chromosome 2 is a fusion of 2 individual chromosomes. In the other Great Apes, these 2 Chromosomes are referred to as 2p and 2q to illustrate their synteny to the human counterpart.

Jeremy Seto jseto@citytech.cuny.edu



Synteny map of Human, Chimpanzee, Gorilla, Orangutan and Marmoset (non-ape primate). Mapping of chromosome 2a and 2b in the apes compared to 6 and 14 in the marmoset illustrates the relatedness of the chromosomal structure of the apes. Minor inversions are apparent in the orangutan chromosome.. Credit: Jeremy Seto [CC-BY-NC-SA]

Chimpanzees (*Pan*) are the closest living relatives to modern humans. It is commonly cited that less than 2% differences in their nucleotide sequences exist with humans (Chimpanzee Sequencing and Analysis Consortium, 2005). More recent findings in comparing the complement of genes (including duplication and gene loss events) now describes the difference in genomes at about 6% (Demuth JP, et al., 2006).

The Pan-Homo divergence. A display at the Cradle of Humankind illuminates the skulls of two extant Hominini with a series of model fossils from the Hominina subtribe of Australopithecina and Homo. Credit: Jeremy Seto (CC-BY-NC-SA) https://flic.kr/p/SmhHTd

Jeremy Seto jseto@citytech.cuny.edu

The Genus *Homo*

The rise of the human lineage is thought to arise in Africa. Fossils of *Austroloptihs* (southern apes) found in death traps, like those at the Cradle of Humankind, reveal a historical record of organisms inhabiting the landscape. The breaks in the ceiling of the caves provide opportunities for animals to fall inside these caves to their death. The limestone deposits of the caves serve as an environment for fossilization and mineralization of their remains. An abundance of fossilized hominids in these caves including *Australopithecus africanus Australopithecus prometheus, Paranthropus boisei,* and the newly discovered *Homo naledi* continue to reveal the natural history of the genus *Homo* from 2.6 million to 200,000 years ago.

An underground lake at inside the Sterkfontein Cave system at the Cradle of Humankind (South Africa)

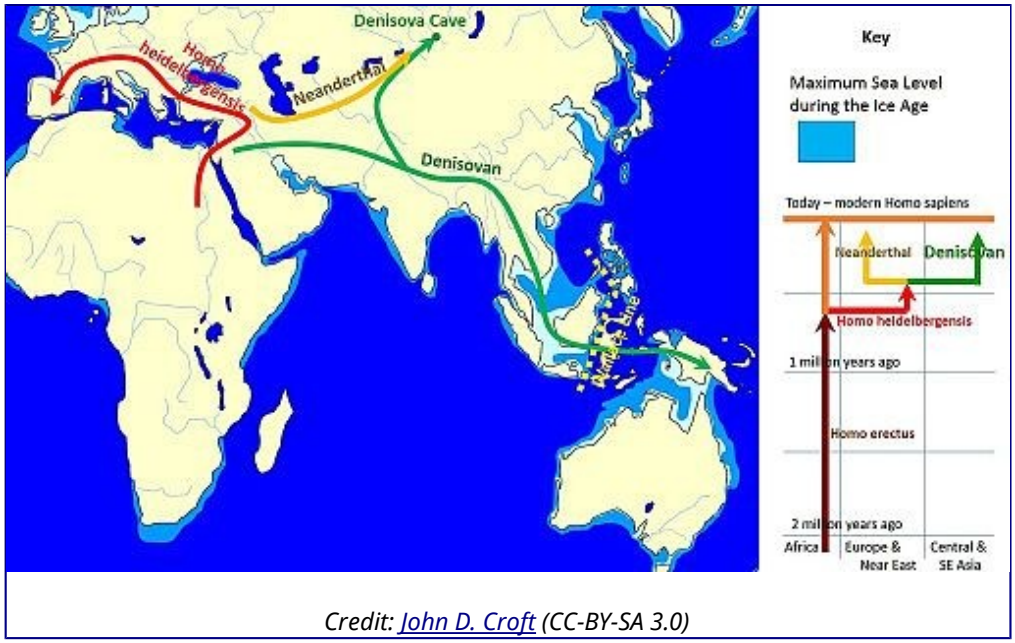
Credit: Jeremy Seto (CC-BY-NC-SA) https://flic.kr/p/RczrEg

The entrance to the archaeological site at Sterkfontein, Cradle of Humankind (South Africa).

Credit: Jeremy Seto (CC-BY-NC-SA) https://flic.kr/p/ULs2Sv

Jeremy Seto jseto@citytech.cuny.edu

Ancient DNA of Humans



Credit: John D. Croft (CC-BY-SA 3.0)

In 2008, a piece of a finger bone and a molar from a Siberian Cave were found that differed slightly from that of modern humans. The cave, called Denisova Cave, maintains an average temperature of 0ºC year round and was suspected to contain viable soft tissue. Bones in this cave were discovered that had similarities to modern humans and Neandertals. An initial mitochondrial DNA analysis revealed that these beings represented a distinct line of humans that overlapped with them in time (Krause *et al.*, 2010). Analysis of the full nuclear genome followed and indicated that interbreeding existed between these Denisovans, Neandertals and modern humans (Reich *et al.*, 2010). Furthermore, analysis of DNA from a 400,000 year old femur in Spain revealed that these three lines diverged from the species *Homo heidelbergensis* and that Denisovans were closest in sequence (Meyer *et al.*, 2016).

Between modern humans, markers found in the mtDNA can be used to trace the migrations and origins along the maternal line. Similarly, VNTRs found on the Y chromosome have revealed migration patterns along paternal lines within men. Other markers, like the insertion points of transposable elements can be used to further describe the genetics and inheritance of modern humans while providing a snapshot into evolutionary history.

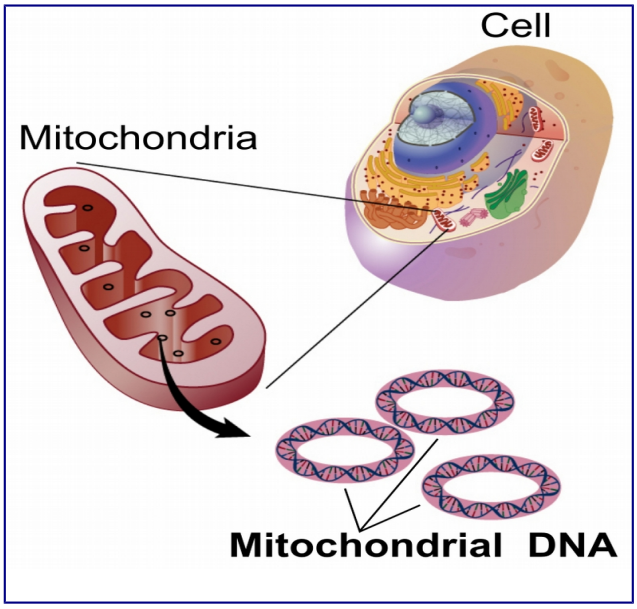
Jeremy Seto jseto@citytech.cuny.edu

Other Resources

- **Great Ape Mitochondrial Sequences**
- http://media.hhmi.org/biointeractive/click/Origins/01.html
- Yunis JJ, Prakash O. The origin of man: a chromosomal pictorial legacy. Science. 1982 Mar 19;215(4539):1525-30.
- Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005 Sep 1;437(7055):69-87.
- Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. **The evolution of mammalian gene families**. PLoS One. 2006 Dec 20;1:e85.
- Krause, Johannes; Fu, Qiaomei; Good, Jeffrey M.; Viola, Bence; Shunkov, Michael V.; Derevianko, Anatoli P. & Pääbo, Svante (2010), "The complete mitochondrial DNA genome of an unknown hominin from southern Siberia", Nature, **464** (7290): 894–897, doi:10.1038/nature08976, PMID 20336068
- Reich, David; Green, Richard E.; Kircher, Martin; Krause, Johannes; Patterson, Nick; Durand, Eric Y.; Viola, Bence; Briggs, Adrian W. & Stenzel, Udo (2010), "Genetic history of an archaic hominin group from Denisova Cave in Siberia", *Nature*, **468** (7327): 1053–1060, doi:10.1038/nature09710, PMID 21179161
- Meyer M, Arsuaga JL, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, Martínez I, Gracia A, Bermúdez de Castro JM, Carbonell E, Viola B, Kelso J, Prüfer K, Pääbo S. **Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins.** Nature. 2016 Mar 24;531(7595):504-7. doi: 10.1038/nature17405. PMID:26976447

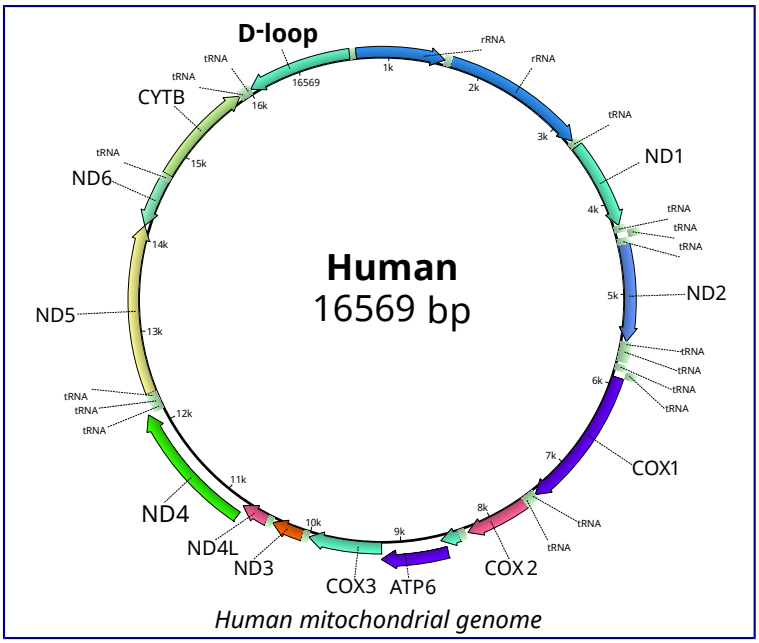
Jeremy Seto jseto@citytech.cuny.edu

Mitochondrial and Maternal Inheritance



In addition to the 23 chromosomes inherited from mother and 23 chromosomes inherited from father, humans have an additional genome that is only inherited from the mother. This genome comes from the **endosymbiotic** organelle, the mitochondrion.

Mitochondria are thought to have arisen in the eukaryotic line when bacteria capable of detoxifying the deadly effects of atmospheric oxygen were engulfed by a eukaryote that did not proceed to consume it. Over the course of time, these formerly free-living bacteria became dependent on the eukaryotic cell environment while providing the benefit to the host cell of aerobic respiration. Hallmarks of this endosymbiotic event include: the inner prokaryotic membrane surrounded by the outer eukaryotic membrane, the presence of prokaryotic ribosomes and most significantly, the circular prokaryotic chromosome. Mitochondria still replicate independently of the host cell but can not survive outside of this cellular environment. Animal mitochondria have the simplest genomes of all mitochondrial genomes, ranging from 11-28kb. The human mitochondrial genome consists of 37 genes which are almost all devoted to processing ATP through oxidative phosphorylation.

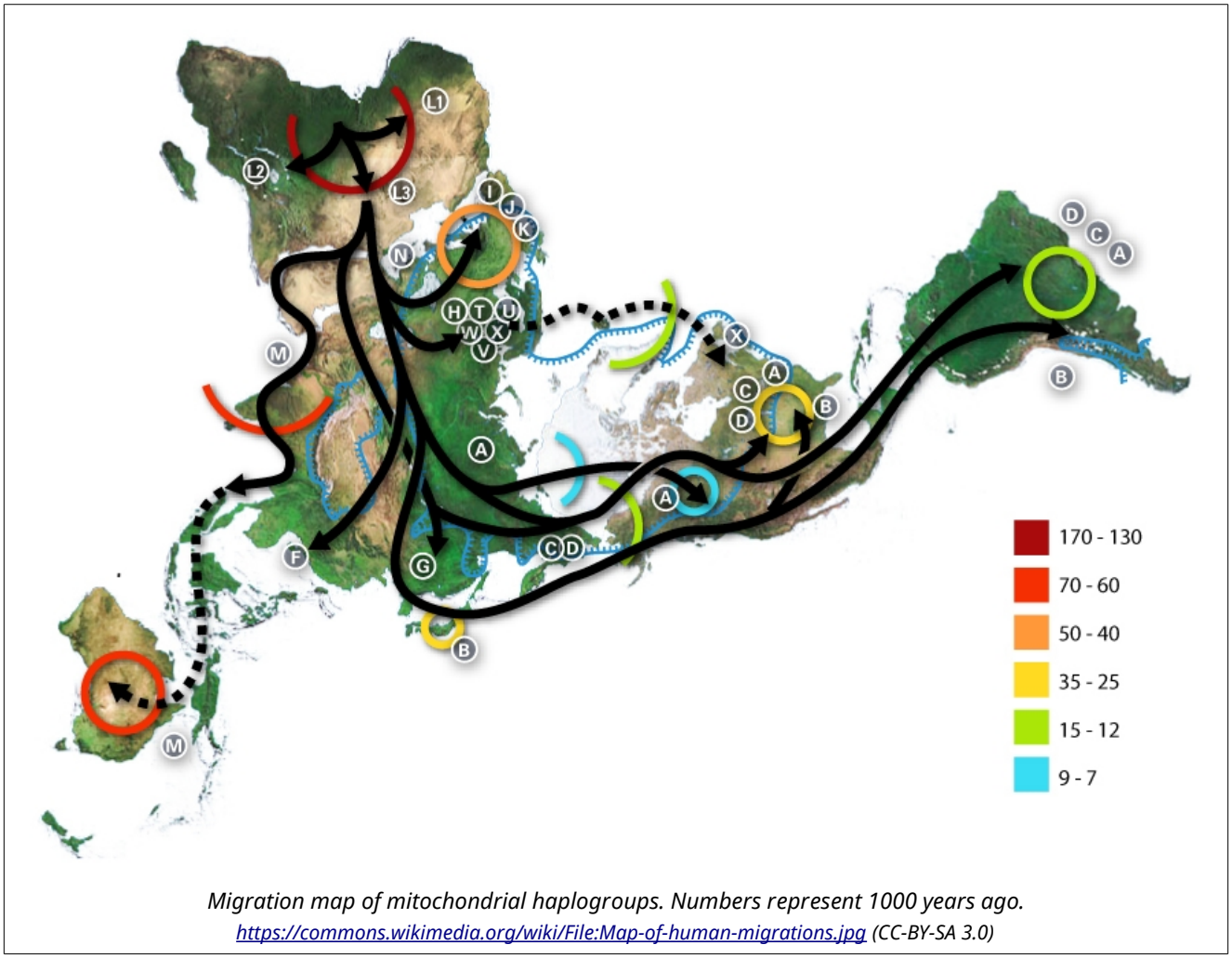


Human mitochondrial genome

The human mitochondrial genome (genbank file) consists of 16,569 nucleotides (16.6kb). While most of this 16.6kb genome consists of protein encoding genes, approximately 1.2kb non-coding DNA takes part in signals that control the expression of these genes and replication processes. It is the area of DNA where the double-strandedness is displaced and having the name **D-loop** (displacement loop). Mutations in this area generally have very little effect on the functioning of the mitochondria. Because of this reduced selection pressure on this area, this **control region** is also referred to as the **hypervariable region**. This hypervariable region actually has 10 times more SNPs than the nuclear genome. Due to this abundance of mutations, it is possible to track down the maternal line of an individual. Why just maternal? The human oocyte contains many mitochondria while sperm cells only contain mitochondria that power the flagellar motion. Upon fertilization, the flagellum and the associated mitochondria are lost, leaving the zygote with only maternal mitochondria.

Jeremy Seto jseto@citytech.cuny.edu

The cluster of SNPs found in the mitochondrial control region are linked and are always inherited together. Because of the lack of paternal contribution, this linkage is referred to as a **haplotype**, or "half-type". Tracking these polymorphic haplotypes, a family tree of humans was developed in the 1980s which concluded that humans arose from a metaphorical "Mitochondrial Eve" 200,000 years ago. As a metaphor to the Biblical Eve, this alludes to an origin but unlike the Biblical event, this does not mean that it was a single woman that gave rise to all of modern humanity. On the contrary, the metaphor merely indicates that a series of females; sisters and cousins, of this line gave rise to modern humans.



Migration map of mitochondrial haplogroups. Numbers represent 1000 years ago.
https://commons.wikimedia.org/wiki/File:Map-of-human-migrations.jpg (CC-BY-SA 3.0)

The use of mitochondria for this analysis provides great flexibility, especially from ancient sources. Unlike the nuclear genome which only has 2 copies of DNA per cell, the mitochondria are abundant in number and provide many copies of genome per cell. Ancient sources of DNA in fossils will most often have degradation of the DNA. The mitochondrial genome is just as likely to undergo degradation over time, however the high copy number allows for gaps to be filled in easily. SNPs do not alter the overall size of the hypervariable region, therefore amplification by PCR can not resolve these differences based on agarose gel migration. However, **amplicons** (amplified copies) can be sent for sequencing whereby each nucleotide can be called out in succession and reveal the specific SNPs.

The PCR amplification of the mitochondrial control region

There are 2 hypervariable regions within the control region of the mitochondria. This exercise amplifies just one of these. For more definitive results, both should be amplified and sequenced. This exercise will permit us to have a rough idea of the origins of our maternal line and we will be able to attribute ourselves to various tribes throughout the world. The human mitochondrial genome (genbank file).

Forward Primer 5'-TTAACTCCACCATTAGCACC-3'
Reverse Primer 5'-GAGGATGGTGGTCAAGGGAC-3'

1. PCR the previously extract DNA samples
    - Pour 2% agarose into casting apparatus in refrigerator
    - 2 gels per class need to be made → 100ml of TBE with 2g agarose
    - add 5µl SYBR safe solution into the molten agarose before casting
    - place 2 sets of combs into the gel → at one end and in the middle
2. load gel with DNA ladder and PCR
3. Run gel at 120V for 20 minutes
4. Visualize on UV transilluminator
5. Document with camera to verify amplification
6. The instructor will submit the viable reactions for sequencing
7. **Analyze data during Bioinformatics Lab session**
    1. Using NYCCT email address, register for account at https://dnasubway.cyverse.org/
    2. retrieve reference mitochondrial sequences
    3. perform multiple sequence alignment using MUSCLE
    4. draw phylogenetic trees using PHYLIP and visualize using FigTree

Jeremy Seto jseto@citytech.cuny.edu

Transposable Elements

Mobile genetic elements called **transposable elements** or **transposons** are located throughout the genome. These elements were first described in maize by Barbara McClintock at the Cold Spring Harbor Laboratory where she observed a disruption of coloring in corn kernels that did not follow simple Mendelian inheritance.

McClintock's Corn Kernels

Each kernel represents a distinct new individual organism. Kernel color is described through simple Mendelian inheritance where purple is dominant over yellow. Dr. McClintock noticed that some kernels contained spots. She noticed that the coloration disruption could later reverse in subsequent generations. She described the phenomenon of this break in the Mendelian characteristics as a "genetic instability". Over time, she would come to realize that the spots in these kernels arose from the insertion of DNA into the area of genes that were involved in controlling kernel coloration.

Dr. McClintock's description of this phenomenon and the underlying mechanisms was extremely unpopular as it violated what was already known about the *fixity* of genetics. Though initially skeptical, biology has found that these “jumping genes” are found in every taxa including prokaryotes (where they are often associated with genes conferring antibiotic resistance) and she was later awarded the Nobel Prize. Approximately half of the human genome consists of transposons, making up the bulk of what was previously referred to as “junk DNA”.

(See Animation: http://www.dnaftb.org/32/animation.html)

Advanced Video of gene disruption by jumping genes

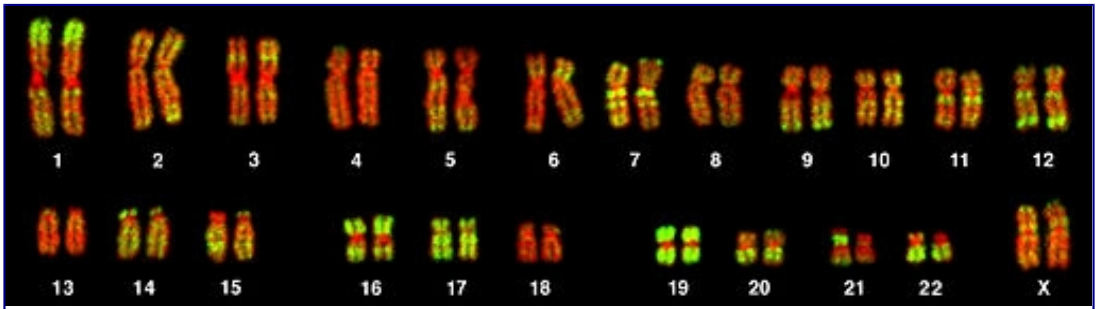
https://youtu.be/_cJfsWYR42M

Classes of Transposons

Transposons can be autonomous or non-autonomous. **Autonomous transposons** encode their own **transposase** enzyme that facilitates the jumping of the gene while **non-autonomous transposons** require the transposase activity of another transposable element. Functional DNA transposons are autonomous and work through a “*cut and paste*” mechanism. DNA transposons are delineated by flanking terminal repeats that mark the location that

the transposase excises the DNA. These DNA elements then re-integrate at a different location within the genome. The excision from DNA leaves marks of these flanking repeats that can be used to study the rate and level of DNA transposition events within a genome. The insertion of these transposons can effect the expression of nearby genes and can completely disrupt genes they land into as evidenced in the speckled corn kernels that McClintock described.

RNA transposons are called **retrotransposons** because they are transcribed into an mRNA and require a reverse transcription to integrate into the genome. The most common mobile element in the human genome are the Long Interspersed Nuclear Elements (**LINEs**) and the Short Interspersed Nuclear elements (**SINEs**). These retrotransposons are most abundantly represented by the autonomous LINE1 (L1) and non-autonomous *Alu* elements, respectively. *Alu* elements rely on the expression of the L1 in order to be reverse-transcribed and integrated into the genome. These retrotransposons work in a "*copy and paste*" mechanism and are responsible for genomic expansion. As their classifications signify, LINEs are longer than SINEs. This is due to the presence of a second reading frame that encodes the transposase.



The Human Genome is littered with "junk" DNA. Green coloration corresponds to Alu sequences in this karyogram of the human genome.
Credit: Andreas Bolzer, Gregor Kreth, Irina Solovei, Daniela Koehler, Kaan Saracoglu, Christine Fauth, Stefan Müller, Roland Eils, Christoph Cremer, Michael R. Speicher, Thomas Cremer (CC-BY 2.5)

- Human L1 sequence (fasta)
- Human Ya5 Alu sequence (fasta)

Advanced Video of Classification of Transposons and the Evolution of Genomes

https://youtu.be/IOXvZXtc93U

Further Reading

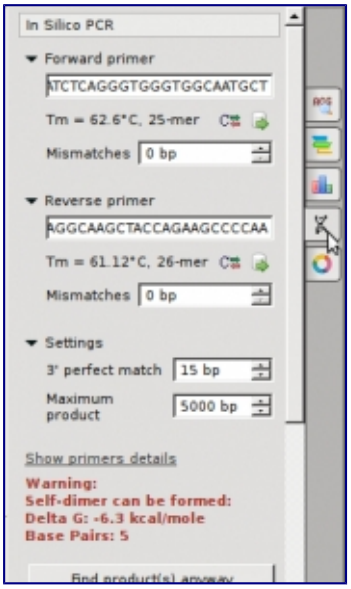
- http://www.nature.com/scitable/topicpage/barbara-mcclintock-and-the-discovery-of-jumping-34083
- http://www.nature.com/scitable/topicpage/functions-and-utility-of-alu-jumping-genes-561

Alu's are unique SINEs that appear in the primate lineage and reveal the lineage and diversification of primates. While retrotransposons can disrupt gene (as in some cases of hemophilia), they often land outside of genes or within introns without effect. One example of a non-disruptive Alu element in humans is found in the location called PV92 on chromosome 16. This element is of the youngest subfamily of Alu, called Ya5.

Since PV92 does not cause any deleterious effects, it can be used as a non-selected marker to illustrate lineage. Some people have an *Alu* element int his location while others do not. The presence or absence of this marker is viewed as an allele. This lab uses primer that flank the location of the *Alu* insertion that span 416 bp. If an *Alu* is present, the amplified DNA will be 300bp larger (the size of an *Alu*) at 731bp.

Exercise: In silico PCR of PV92

Forward primer: 5' GGATCTCAGGGTGGGTGGCAATGCT 3'
Reverse primer: 5' GAAAGGCAAGCTACCAGAAGCCCCAA 3'

1. Perform Virtual PCR **Informatics Exercise/Discussion**
2. Visit BLAST: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch
3. Paste both primers: GGATCTCAGGGTGGGTGGCAATGCT GAAAGGCAAGCTACCAGAAGCCCCAA
4. Choose "Somewhat Similar"
    - Locate the locus of the product and the size
5. Find the PCR fragments in Ugene
    - Download the sample FASTA file: PV92 sample
    - Open the file in Ugene and select option "As Separate Sequences in Viewer"
    - Select the "In Silico PCR" button on the far right (double helix button) and insert the primers
    - 

    - A PCR product should be noted for one of the sequences after pressing "Find Products anyway"
    - Click on the second sequence in the viewer and Press "Find Products anyway"

Exercise: PCR genotype PV92 locus

1. PCR the individual samples
2. Pour 2% agarose into casting apparatus in refrigerator
    - 2 gels per class need to be made → 100ml of TBE with 2g agarose
    - add 5µl SYBR safe solution into the molten agarose before casting
    - place 2 sets of combs into the gel → at one end and in the middle
3. Load DNA ladder and PCR samples
4. Run gel at 120V for 30 minutes
5. Visualize on UV transilluminator
6. Score gels for the presence/absence of the alleles to determine genotype frequency in the class