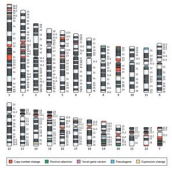
Genetics, Genomics and Human Origins



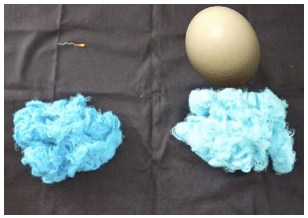
Introduction to Anthropogeny
Pascal Gagneux

Lecture 2
Thursday, October 8, 2020

Question:

- What is the genetic difference between humans and apes in %?

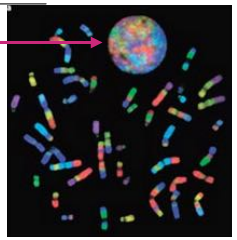
Molecules of inheritance Recap: DNA



1000 x model of the genome in Sperm and Egg

interphase nucleus

spread from metaphase (cell dividing by mitosis)



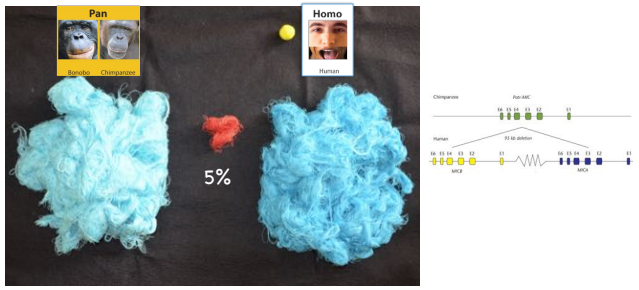
Human chromosomes painted with probes from sorted gibbon chromosomes

Ferguson Smith 1997 ***European Journal of Human Genetics***

Human egg 130 microns (the size of a grain of table salt) magnified times 1000 =130 mm = 13 cm; Sperm 60 micron (invisible to the naked eye) magnified times 1000 = 6 cm long; Somatic cell: 20 micron, magnified times = 2 cm diameter (large marble); Haploid Genome: ~1 meter, magnified times 1000 = 1 km of thread.

The human genome consists of 22 pairs of autosomal chromosomes and one pair of sex chromosomes. Each haploid genome is thus 23 chromosomes (each chromosome is a single strand of DNA tightly wrapped around histone proteins, the combined length of these 23 chromosomes is ~ 1m). Metaphase, when the cell is ready to divide, is when chromosomes can be seen as compact rods under a microscope after staining. When a cell is busy living (not replicating) during interphase , the DNA is decondensed into “puffs” as seen above. Sorted gibbon chromosomes can be stained and used for “painting” of human chromosomes. The gibbon fluorescent chromosomes will bind to places on the human chromosomes that have similar DNA

1000 x model of haploid genome



Including Insertions and deletions (peices of the genomes that have no direct counterpart in the other genome:
Human and chimpanzee genomes differ by ~ 5% of their DNA.
Example of a region where humans have two copies of a gene (MIC) but chimpanzees lost one of the two…

Question:

- How much genetic change is necessary for speciation to occur?

Speciation


Cercopithecus mitis


Cercopithecus mitis x ascanius


Cercopithecus ascanius

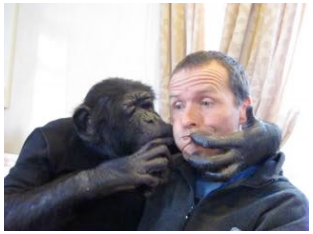

Papio anubis


Papio anubis X hamadryas


Papio hamadryas

Many species in the wild hybridize. There is not linear metric of % DNA difference that can be applied to the possibility of viable hybridization.

Hybridization happens



Female F1 hybrid of accidental cross between male bonobo and female common chimpanzee.

Genomic evidence for lack of pre-zygotic species barrier between chimpanzees and bonobos

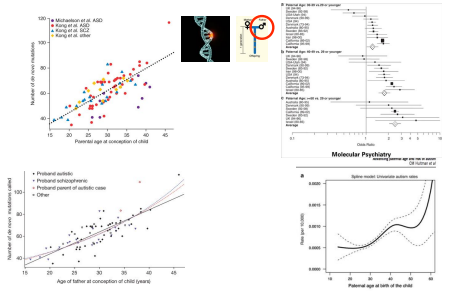
Samuel H. Vohr, Brendan O'Connell, Beatriz Freitas, Florence Ollivet-Courtois, Peter D. Heintzman, Pascal Gagneux, Alysson R. Muotri, and Richard E. Green submitted

Graph b shows that the DNA of the F1 Bonobo X chimpanzee hybrid has DNA that is equidistant from both chimp and bonobo parental chromosomes at all locations along each chromosome.

Question:

- How could culture/behavior affect genomic evolution?

Paternal age and de novo mutations

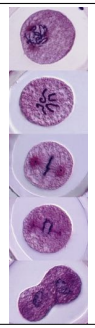


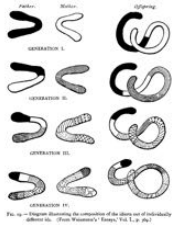
Kong et al. ***Nature*** 2012

Molecular Psychiatry (2011) **16,** 1203-1212;

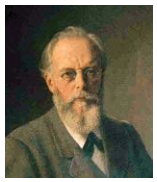
Sequencing parent child trios for their whole genomes now allows to measure the rate of de novo mutation per generation.
About 50 new mutations are inherited by each child, unless the father is older, then many more de novo mutations are inherited.
Parental age is a strong risk factor for autism spectrum disorders.

Soma and Germ Line

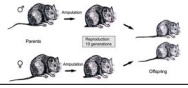




"idants in worm eggs"



August Weismann 1893 Germ Plasm a Theory of Heredity



The stuff of heredity, is particulate and is only passed on through specialized cell: the germ line. The rest of the body (the soma) serves as the vehicle to get these sexual cells to the right place: meeting with sex cells from other individuals.

Practice Question: What is the difference between somatic and germ cells?
Somatic cells are the majority of the cells in the body, germ cells can give rise to gametes.

Inheritance of acquired characteristics?

>200 generations of male circumcision, male babies still born with a prepuce!

4500 years

> 220 generations

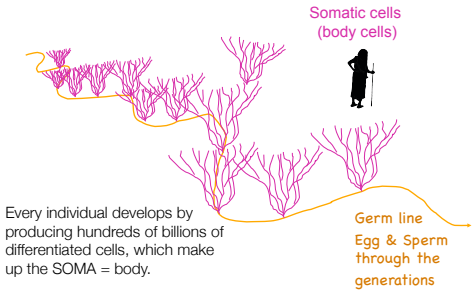
Relief depiction of circumcision from from the necropolis at Saqqara
Dating to Dynasty 6 and specifically to the Service of Nesi (Pharaoh) Teti (2355-2343 BCE)

Just like the mouse tails cut off along successive generations by Weissmann, over 200 generation of male circumcision has not caused baby boys born without a prepuce. **It represents a culturally inherited practice!**

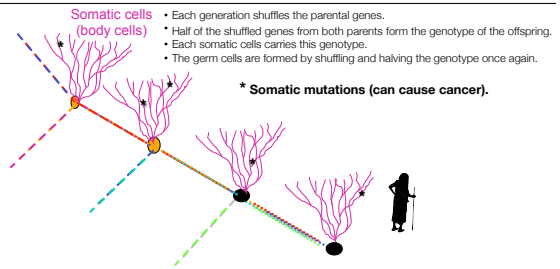
Practice question: What does the ancient practice of circumcision in the Middle East illustrate about the inheritance of acquired characteristics?

Answer: All males in these societies are still born with an intact prepuce, proving that more than 200 generations of cutting has failed to translate into a genetically inherited character.

Germ Line and Soma



The Germ Line is not a simple line:



The germ line is made up of shuffled pieces of DNA that meet and get taken apart again by sexual recombination.—-> **WE ARE DYNAMIC MOSAICS of mixed heritage!**

Each generation shuffles the parental genes. Half of the shuffled genes from both parents form the genotype of the offspring. Each somatic cells carries this genotype.
The germ cells are formed by shuffling and halving the genotype once again. Half of the chromosomes in the gametes are recombinant, the other half are as found in the parents.

Good example of misperception: Blood as vehicle of inheritance? **Completely erroneous notions of "pure blood" and "contamination" are still around.**

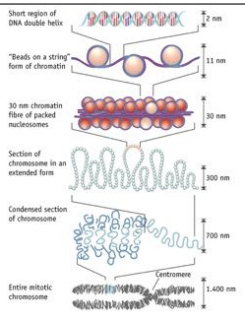
Practice question: Can somatic mutation be passed on to the next generation?

Answer: No, only germ line mutations can.

Practice question: Are most mutations dangerous to the survival of the individual in which they occur?

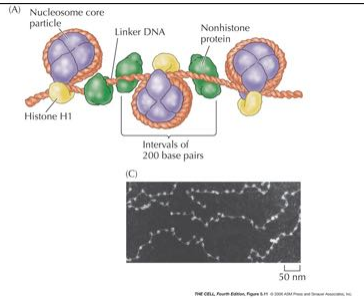
Answer: NO, most mutation appear to be neutral.

Fitting a genome into a cell: genome packaging



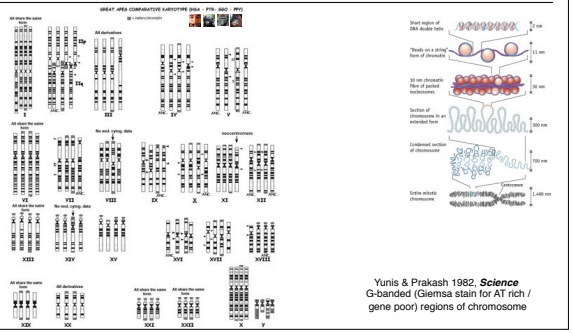
Highly efficient packaging of DNA into chromatin and chromosomes. Each chromosome is based on one segment of genomic DNA.

Genome packaging: chromatin



"Beads on a string" structure of chromatin, visible when chromosomes open up their tight coils in between cell divisions.

Fitting a genome into a cell: genome packaging

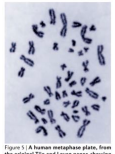


Comparison of human and great ape chromosomes. Humans have 46 chromosomes in each cell (sperm and eggs have 23 each, as they are haploid). The banning patterns obtained after staining with DNA specific dyes are highly similar between apes and humans.

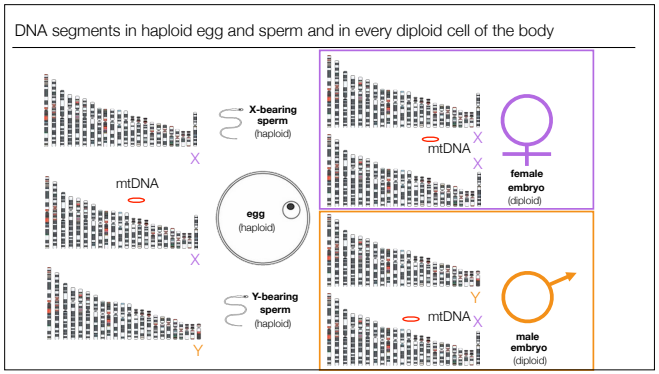
Getting the number right…..

For over 30 years: 48

Since 1956: 46



It took a while to get the correct number of chromosome in human cells….
Diving human cells can easily be found in testes or white blood cells can be triggered to undergo cell division by adding special chemicals: lectin proteins from beans to trigger cell division and colchicine from a flower to arrest cell division in the stage where chromosomes have just been copied.

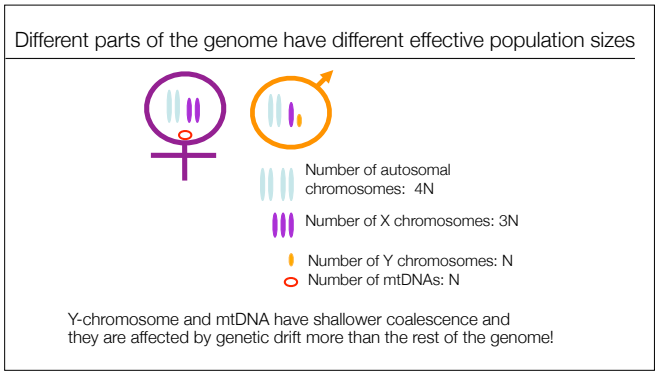


The complement of chromosomes in human gametes (haploid = single copy of each chromosome) and body (somatic) cells (diploid two copies of each chromosomes except for males who have a single X chromosome and a Y chromosome). 22 pairs of **autosomal** chromosomes and a pair of X chromosomes for females, or a single X- and single Y chromosome for males. X and Y chromosomes are known as **sex chromosomes** or **allosomes**. The **karyotype** (full complement of all chromosomes) of a female and male cell is illustrated on there right.

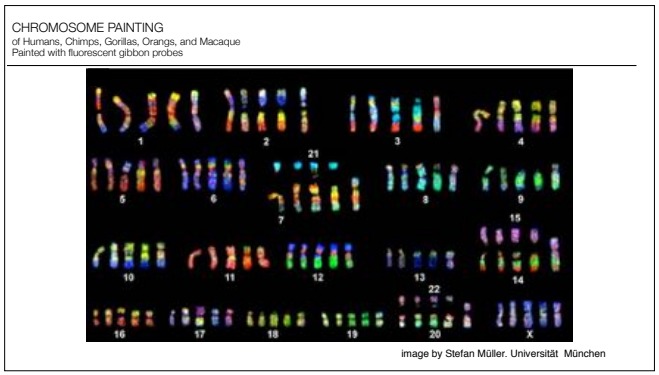
Practice question: What is a karyotype?

Answer: The total set of chromosome in the cells of an individual.

Interesting fact: sperm do not actually carry chromosomes as the 23 segments of DNA of sperm cells are mostly removed from the histone proteins and tightly wrapped around protamine proteins so as to fit snuggly into a minuscule sperm head.

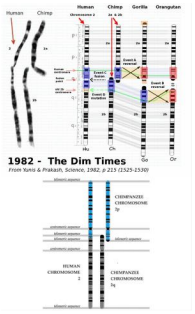


Due to their lower copy number in the “gene pool” Y chromosomes and mitochondrial variants are more easily lost from the population by chance (i.e. they are more strongly affected by genetic drift).



Remarkable similarity between human and great ape chromosomes when painted with sorted and fluorescently labeled gibbon chromosomes.

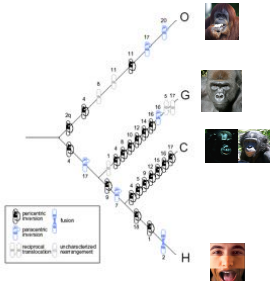
Chromosome fusion e.g. human chromosome 2



Details of the fusion of two ancestral ape chromosomes giving rise to the human chromosome 2.

Taxonomy by CHROMOSOMES

Cytogenetics:
the study of chromosomes
first visible DNA



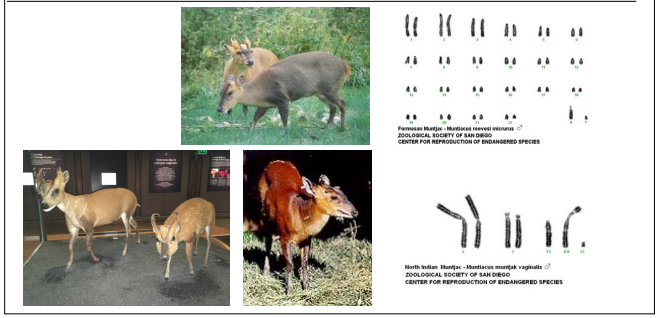
Modified from Mark A. Jobling ,Matthew Hurles ,Chris Tyler-Smith (2003)

Inversion, translocation, fusions and some more complicated changes in chromosome organization can be mapped onto the phylogeny of hominids.

Question:

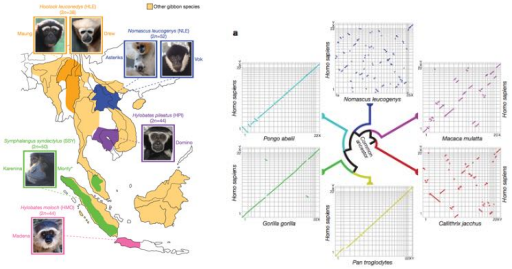
- Can similar species have radically different karyotypes (number and arrangement of chromosomes)?

Chromosomal (karyotypic) evolution
(still poorly understood)



Similar looking mammals (barking deer) have either 46 or 7 chromosomes!

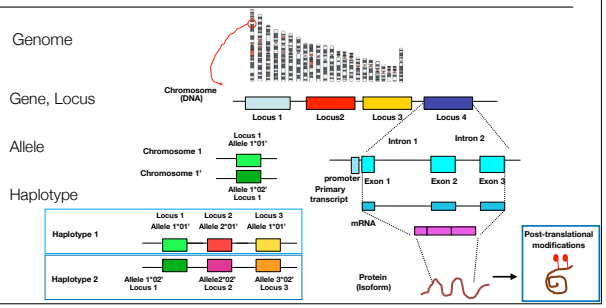
Gibbon Karyotype Evolution



Gibbon genome and the fast karyotype evolution of small apes, ***Nature*** 2014 September

Assembly and analysis of a northern white-cheeked gibbon (Nomascus leucogenys) genome. We describe the propensity for a gibbon-specific retrotransposon (LAVA) to insert into chromosome segregation genes and alter transcription by providing a premature termination site, suggesting a possible molecular mechanism for the genome plasticity of the gibbon lineage. We further show that the gibbon genera (Nomascus, Hylobates, Hoolock and Symphalangus) experienced a near-instantaneous radiation ~5 million years ago, coincident with major geographical changes in southeast Asia that caused cycles of habitat compression and expansion. Finally, we identify signatures of positive selection in genes important for forelimb development (TBX5) and connective tissues (COL1A1) that may have been involved in the adaptation of gibbons to their arboreal habitat.

Genetic vocabulary:
"genome, gene, allele, haplotype"



Some key terms used in genetics: genome, gene, locus ("site" in Latin), allele, haplotype, promoter, exon, intron, mRNA, post-translational modification. The same locus can encode a number of different "versions" of a protein called "isoforms" by splicing and use of different transcription starting site. Haplotypes are long stretches of DNA that carry unique combinations of genetic variants (alleles). Post-translational modifications of proteins include the addition of sugar (glycosylation), or phosphate (phosphorylation) etc… Among other things, these modification regulate the function of proteins.

Wild Type allele: most common variant, not associated with disease.

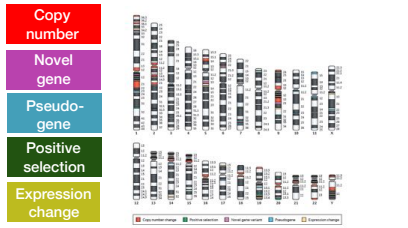
Practice Question: What is a haplotype?

Answer: a unique combination of variants (alleles) along a stretch of DNA.

Practice Question: How many different alleles exist at a particular locus in an individual human?

Answer: two, except for X and Y chromosomes where males have only one allele.

Human-Lineage-Specific Changes are Scattered throughout the Genome

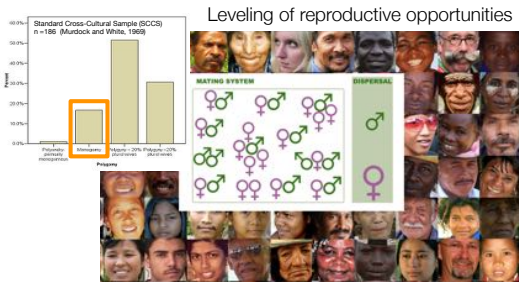


A vast landscape of functional elements still being discovered…

O'Blenness *et al.* ***Nat Rev Genet.*** *2012*

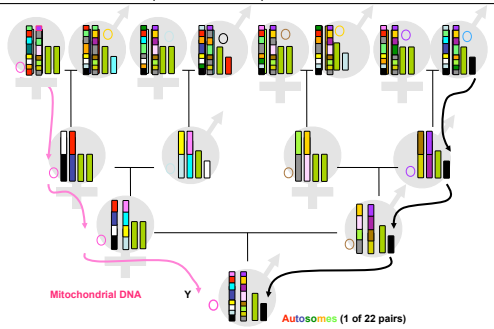
Ape genomes are about 2 meters long 3 billion basepairs on 23 or 24 different chromosomes. Human changes are scattered throughout this vast landscape: 5% of our DNA differs from that of the chimpanzee genome if deletions and duplications are included. The search is on identifying key genetic changes. Important changes range from single nucleotide (DNA "letters") changes to large duplications or deletions, differences in gene copy number and altered regulatory DNA. In addition, identical genes can be expressed at different times and in different tissues, creating large differences in the organism.

Finding a Mate



The majority of documented traditional human societies tolerated some degree of polygamy by men. Individually, the majority of men appear to have been (serially) monogamous. Human mating system are remarkably diverse and do include some with polyandry (one woman with multiple male partners).

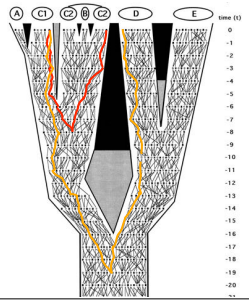
Modes of inheritance: uniparental & biparental



Most DNA is on autosomal chromosomes (all but the X or Y sex chromosomes). These autosomal chromosomes get reshuffled when eggs and sperm are produced. Each one of us is a genetic mosaic that only exists in its present combination once in the history of the universe!

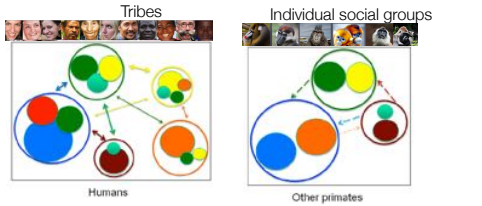
Most of the Y-chromosome and the mitochondria DNA are not subject to recombination.

DNA trickling through generations



Different pieces of our genomes share common ancestors at various time depths (number of past generations back)

Small groups - large networks



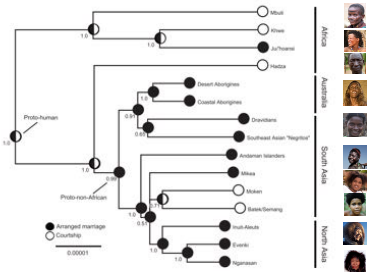
reciprocal exogamy
multiple kin lineages in multiple communities
socio-cultural niches

Walker et al. **PlosOne** 2011

The evolved human social structure (left) of reciprocal exogamy including the exchange of mates, goods, and services (double-headed arrows), involves multiple kin lineages (filled circles) often existing in multiple residential communities (open circles). Extensive cooperation (overlap of filled circles) likely results in economies of scale within and across human communities. In contrast, in other primates (right) one or the other sex emigrates (dotted arrows). The lack of any reciprocal exogamy means that kin lineages are isolated to single communities and thus do not generate a meta-group social structure as found in humans. Kin lineages in humans are directly identified by language., essentially allowing the invention of “tribes”.

Practice question: Explain the term reciprocal exogamy.
Repeated exchanges of mates, goods and services between different social groups.

Evolutionary History of Hunter-Gatherer Marriage Practices

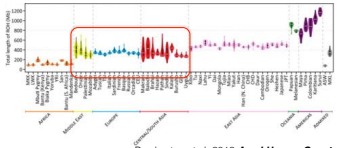


Walker *et al.* ***PlosOne*** 2011

Many documented hunter and gatherer societies have arranged marriages.

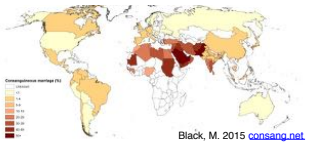
Human culture shaping the genome via marriage patterns

Length of runs of homozygosity



Pemberton et al. 2012 ***Am J Human Genet***

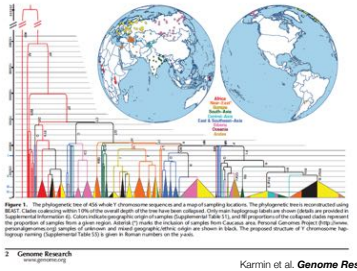
Prevalence of consanguineous marriages



Black, M. 2015 consang.net

Impact of close-relative marriage (uncle niece or first cousins) on homozygosity (length of DNA segments with no evidence of recombination). Marriage patterns are directly mold human genomes.

Cultural Effects on the Gene Pool:

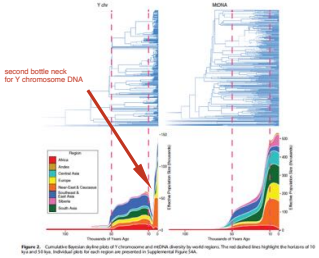


Karmin et al. ***Genome Research*** 2015

Phylogeny of Y-chromosomes showing deeper divergences (highest levels of variation) in Africa

---

Cultural Effects on the Gene Pool:



Strong bottleneck in Y-chromosome ~6000 years ago, no such effect on mt DNA!
Male variance in reproductive success with adoption of agriculture and subsequent conflict/wars between paternal sibships?

Karmin et al. ***Genome Research*** 2015

Extinction of entire Y-chromosome lineage due to lethal conflicts between paternal kinship clans?

The “Gene” Crowd

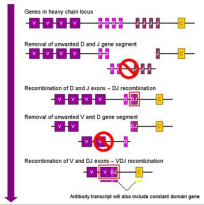
Spencer Tunick, Socalo, Cuidad Mexico 2007

- 22,000 protein coding genes (UCSD undergrads!)
- 1000s of long non-coding RNA “genes”
- 100s of thousands of enhancers in genome (friends and relatives of UCSD undergrads!)
- 50,000 transcribed enhancers

Analogy for the genome

Question:

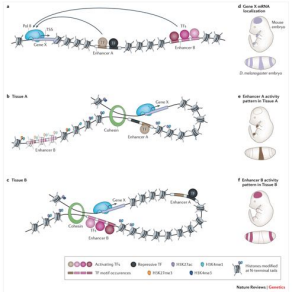
- It is estimated that each of us makes over 10^{12} different antibodies.
- How can a genome code for more proteins than it has genes for?



Individual B-cells and T-cells can diversify their receptors (T-cell receptors and Antobodies) by shuffling cassettes of genes/ somatic recombination.

Enhancers and Transcription Factors

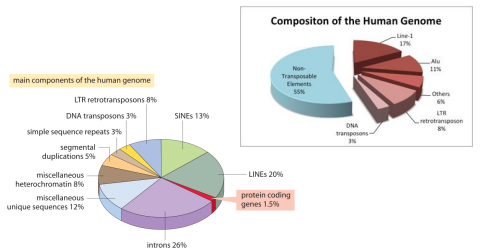
Transcription Factors:
- Proteins
- Pressure
- Cytoarchitecture
- Social partners



Daria Shlyueva, Gerald Stampfel & Alexander Stark ***Nature Reviews Genetics*** 15, 272–286 (2014)

Enhancers are distinct genomic regions (or the DNA sequences thereof) that contain binding site sequences for transcription factors (TFs) and that can upregulate (that is, enhance) the transcription of a target gene from its transcription start site (TSS). Along the linear genomic DNA sequence, enhancers can be located at any distance from their target genes, which makes their identification challenging. b,c | In a given tissue, active enhancers (Enhancer A in part b or Enhancer B in part c) are bound by activating TFs and are brought into proximity of their respective target promoters by looping, which is thought to be mediated by cohesin and other protein complexes. Moreover, active and inactive gene regulatory elements are marked by various biochemical features: active promoters and enhancers are characterized by a depletion of nucleosomes, which is the structural unit of eukaryotic chromatin. Nucleosomes that flank active enhancers show specific histone modifications, for example, histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac). Inactive enhancers might be silenced by different mechanisms, such as by the Polycomb protein-associated repressive H3K27me3 mark (part b) or by repressive TF binding (part c). d–f | Complex patterns of gene expression result from the additive action of different enhancers with cell-type- or tissue-specific

45% of Genome comprised of Transposons (“parasitic DNA”)



Repetitive DNA, was initially excluded from DNA comparisons, thus only 1. 3 % difference reported between human and chimp DNA.
If one includes repetitive DNA (mostly transposable elements) the difference is ~5%. Half of the human genome is made up of “parasitic DNA”/transposons.

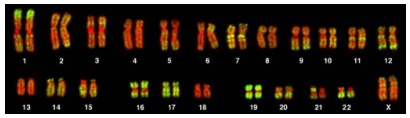
Gene evolution - How many genes?

- 3 billion base pairs in haploid genome
- 22 000 genes
- Alternative splicing (40-60 % of genes)
- Protein coding ~ 20 000? (1.5% of genome)
- Structural RNAs ~ 3 000?
- 2 % highly conserved non protein coding
- Gene structure
- Gene expression
- Post-translational modification
- Microbiome genomes

A few stats about our genomes

Alu elements in primates

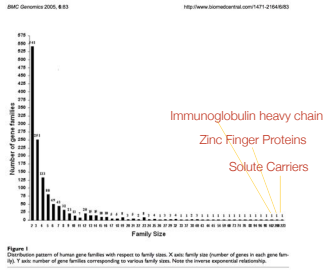
Over 1 million *Alu* element insertions in the human genome!



Karyotype from a female human lymphocyte (46, XX). Chromosomes were hybridized with a probe for Alu elements (green) and counterstained with TOPRO-3 (red). Alu elements were used as a marker for chromosomes and chromosome bands rich in genes.

Alu elements are named after Arthrobacter luteus bacteria. An enzyme from this bacteria cuts DNA at a sequence carried by all these million of copies of a ~ 300 basepair element.

Gene Families and their sizes



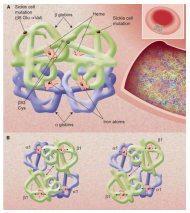
Over half of all genes exist in more than one copy…..

Sharma et al, ***BMC Genomics*** 2005;6:83

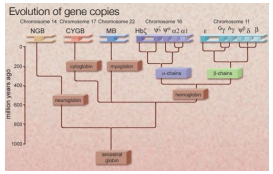
Multiple copies of the same gene allow for adaptation by tweaking the function of each copy…..

Hemoglobin Evolution “genes as LEGO blocks”

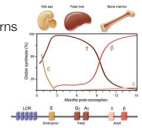
STRUCTURE of PROTEIN



Evolution of gene copies



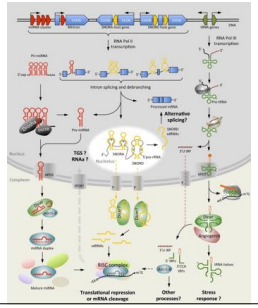
Expression patterns



Pesce *et al.* ***EMBO Rep.*** 2002;3:1146-1151)

The X-ray determined structure of the hemoglobin molecule and a representation of its very high concentration in the erythrocyte. (A) The arrangement of the -helices (shown as tubes) in each unit—one on the left and one, 180° rotated, on the right—is shown, as are the 4 heme groups with their iron atoms where gas molecules bind. The site of the sickle mutations on mutant -chains as well as the 93 conserved cysteine residues is also shown. Hemoglobin molecules in the red blood cell, shown in an inset on the right, are very tightly packed (at a concentration of approximately 34 g/dL) and have little access to solvent; this allows efficient oxygen transport by each cell but also affects the chemical behavior of the molecules, such as promoting sickle cell hemoglobin polymerization upon slight deoxygenation. (B) A representation of the quaternary structural changes in the hemoglobin tetramer, in a top-down view, in the transition from the oxy conformation (left) to the deoxy conformation (right). The iron atoms shift relative to the planes of the heme groups and a central cavity between the -chains opens, facilitating 2,3 BPG binding. These diagrams are based on drawings of Irving M. Geis. Figure 2. A diagram of the proposed evolutionary relationships of the human globin proteins as inferred from sequence analyses. NGB, neuroglobin; CYGB, cytoglobin; MB, myoglobin.

RNA complications:



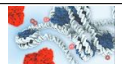
tRNA (transfer)
rRNA (ribosomal)
micro RNAs
sdRNA (sub-genomic)
SnoRNA (small nucleolar)
siRNA (small interfering)
piRNA (piwi interacting)
CRISPER RNA (Clustered Regularly Interspaced Short Palindromic Repeats)

Martens-Uzunova et al. ***Cancer Letters*** 2013

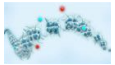
Cross-talk between the pathways of biogenesis and function of miRNAs, snoRNAs, tRNAs, sdRNAs and tRFs. MiRNAs are encoded in clustered genomic loci or in the introns of other genes and are transcribed by RNA polymerase II (RNA Pol Pri-miRNA transcripts are processed to individual pre-miRNAs in the nucleus by Drosha. SnoRNAs and intron-encoded miRNAs are produced after splicing, debranching, and exonucleolytic trimming. Pre-miRNAs are exported to the cytoplasm by Exportin 5 (XPO5) and further processed by Dicer to mature miRNAs that enter the RNA-induced silencing complex (RISC). Cytoplasmically matured miRNAs in complex with proteins from the Ago family (AGO) may be imported back in the nucleus possibly by Importin 8 (IPO8) to participate in the processes of transcriptional gene silencing (TGS) or RNA activation (RNAa). SnoRNAs assemble with snoRNP-core proteins (not shown) and enter the nucleolus where they participate in the chemical modification of ribosomal RNA (rRNA) and other RNA species. SnoRNAs may be exported to the cytoplasm by unknown transporter proteins, where they are cleaved possibly by Dicer to short ~22 nt long sdRNAs and are loaded into RISC. Alternatively, snoRNAs may also be cleaved by unknown nucleases in the nucleus or nucleolus, to sdRNAs with a different size. Longer sdRNAs of ~27 nt do not exit the nucleus, but instead participate in the regulation of alternative splicing. tRNAs are transcribed from individual tRNA genes by RNA polymerase III (RNA Pol III). Pre-tRNA transcripts are processed by the endonucleases RNase P and RNase Z to remove 5′- and 3′-trailer sequences, and after chemical modification, CCA addition, and aminoacylation, are exported to the cytoplasm by Exportin-t (XPOT) to participate in protein synthesis. 3′U tRFs are produced by RNase Z after trimming of the 3′-trailer sequence. Stress factors may induce cleavage in the anticodon loop of mature tRNAs to tRNA halves performed by the endonuclease Angiogenin. Shorter 5′tRFs and 3′CCA tRFs may be produced from 5′- and 3′-ends of mature tRNAs by Dicer and associate with AGO proteins to participate in various processes of transcriptional and post-transcriptional regulation. sdRNA = subgenomic RNA SnoRNA=small nucleolar RNA MiRNA

Genome “syntax”

Chromatin Remodeling
“genome packaging” and its effects on gene expression via access for transcritpion factors, enhancers and transcriptional machinery



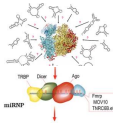
Histone Modification
annotation of histone and effects on gene expression. methylation, acetylation, ubiquitination, O-GlcNAcylation



DNA Methylation
annotation of DNA, silencing of paternal or maternal allele, or both.

non-coding RNA
(micro, piwi, nc, circular RNA etc.. interact with ribosomal proteins, transcription factors, messenger RNA)

RNA-binding Proteins
Approximately 1000 RBP in nucleus, cytoplaams and mitochondria regulate splicing, translation, degaradation



Information in the genome includes several layers above the simple DNA sequence: packaging, modification of histones, modification of DNA, microRNA and microRNA binding proteins all affect regulation of gene expression.

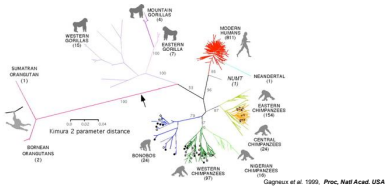
Information from comparative genomics: Demographic History vs Biology

- Whole genome sequencing of multiple individuals:
  - Reconstruction of phylogeny
  - Reconstruction of past population dynamics (demography)
  - Reconstruction of past episodes of selection
  - Comparative genomics to look for shared features, uniquely derived features and their effects on each species

To reconstruct the evolutionary history, there more parts of the genome is available, the better the reconstruction will become.
Currently, it is realistic to obtain whole genome data fro most species. Reconstructing population dynamics relies on the use of parts of the genome that are “selectively neutral”. Most of the genome actually does not see to be under strong natural selection and the majority of mutations are neutral (no advantage and no down side). Detecting past natural selection in present day genomes is difficult. Most commonly, one looks for sequence that appear more different between closely related species than one would expect under “neutral evolution” which is time since last common ancestor times the mutation rate. Once identified, uniquely evolved parts of the genome can be studied in model animals or cell culture for their effects on biology.

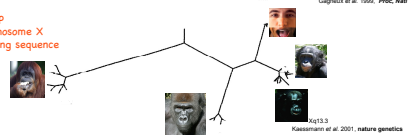
Hominid phylogenies, mitochondrial and X-linked DNA

350 bp
of mitochondrial
sequence



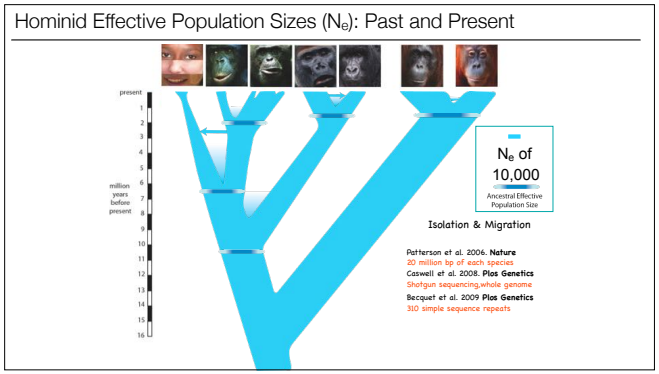
Gagneux *et al.* 1999. ***Proc. Natl Acad. USA***

11,000 bp
of chromosome X
non-coding sequence



Kaessmann *et al.* 2001. **nature genetics**

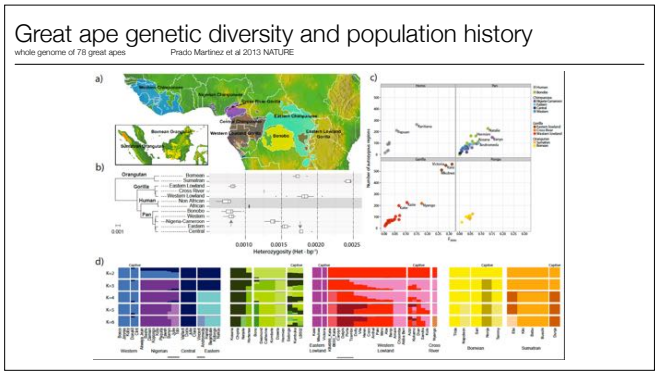
Years ago, I collaborated with a large group of people to compare the DNA sequences of a small stretch of mitochondrial DNA. We reported that each of the grat ape species showed much more genetic variation than 800 humans from populations from all around the world. A few years later, Svante Pääbo’s group sequenced a stretch of DNA 30 times longer on noncoding parts of the X chromosome and found a similar pattern. Now we have whole genomes for all these players including Neanderthals.



N sub e Definition:

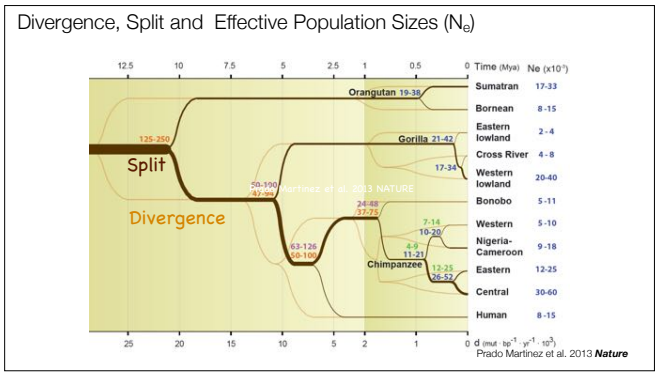
The number of breeding individuals in an idealized population whose genetic diversity is subject to the same effects of genetic drift and inbreeding as the population under consideration.

Uncertainty due to recombination, selection? Different parts of the genome have differing histories. The more variance there is in coalescence times (number of generations back to the last common ancestor) across the genome, the larger the ancestral populations were.



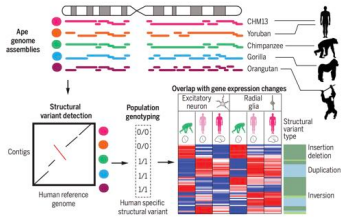
Samples, Heterozygosity and genetic diversity. a. Geographical distribution of great ape populations across Indonesia and Africa sequenced in this study. The formation of the islands of Borneo and Sumatra resulted in the speciation of the two orangutan populations. The Sanaga River forms a natural boundary between Nigeria-Cameroon and Central chimpanzee populations while the Congo River separates the bonobo population from the Central and Eastern chimpanzees. Eastern lowland gorillas and Western lowland gorillas are both separated by a large geographical distance. b. Heterozygosity estimates of each of the individual species and subspecies are superimposed onto a neighbor-joining tree from genome-wide genetic distance estimates. Arrows indicate heterozygosities previously reported[4] for Western and Central chimpanzee populations respectively. An almost fourfold range of heterozygosity is observed among different great ape populations. c. Runs of Homozigosity among great apes. Relationship between the coefficient of inbreeding (F_{ROH}) and the number of autozygous >1Mbp segments. Bonobos, and Eastern lowland gorillas show an excess of inbreeding compared to the other great apes, suggesting small population sizes or fragmented population. d. Genetic structure based on clustering algorithm of great apes. All individuals (columns) are grouped in a different cluster (K=2 to K=6, rows) colored by species and according to their common genetic structure. Some groups, such as Western lowland gorillas, present a transitional clustering pattern, while other groups, such as chimpanzee, show a clear distinct pattern according to the known subspecies. Most captive individuals, labeled on top, present a complex admixture from different wild populations. A signature of admixture, for example, is clearly observed in the

---



Inferred population history. Population splits and effective population sizes (Ne) during great ape evolution. Split times (dark brown) and divergence times (light brown) are plotted as a function of divergence (d) on the bottom and time on top. Time is given using a single mutation rates ($\mu = 1 \cdot 10^{-9}$ mut/(bp·year)). The ancestral and the current effective population sizes are also given depending on μ; the methods used in different periods of time COALHMM and PSMC are colored in orange and blue, respectively. The chimpanzee split times are estimated using the ABC method. The x-axis is rescaled for divergences larger than $2 \cdot 10^{-3}$ to provide more resolution in recent splits.

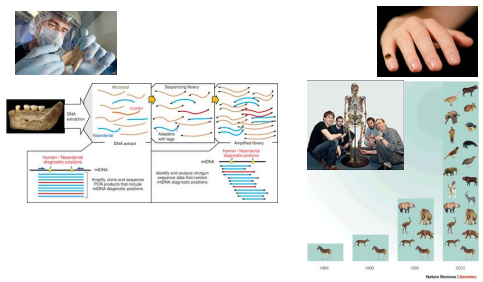
High coverage ape genomes (65X)



Kronenberg let al., ***Science***, 2018

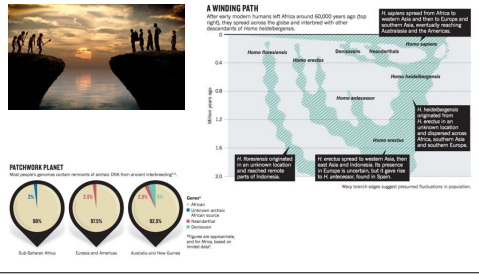
SMRT assemblies and SV analyses. (Top) Contiguity of the de novo assemblies. (Bottom, left to right) For each ape, SV detection was done against the human reference genome as represented by a dot plot of an inversion). Human-specific SVs, identified by comparing ape SVs and population genotyping (0/0, homozygous reference), were compared to single-cell gene expression differences [range: low (dark blue) to high (dark red)] in primary and organoid tissues. Each heatmap row is a gene that intersects an insertion or deletion (green), duplication (cyan), or inversion (light green).

Ancient DNA



Several research group have been sequencing ancient DNA with increasing success. To dat, the oldest hominin DNA sequenced is from Spain (Atapuerca), where DNA from *Homo heidelbergensis*, ancestor to Neanderthals from 400,000 years ago was obtained. The partial finger bone from a young female yielded a good quality genome, that revealed a new group of extinct hominids called Denisovans, after the South Siberian Cave in which it was found.

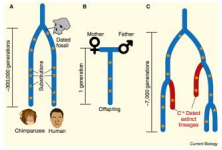
Archaic admixture



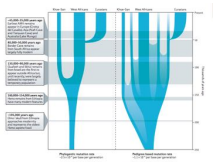
Comparisons between modern human genome sequences and those of two extinct hominids, Neanderthals that went extinct about 45 kya and Denisovans who's date of extinction is not known have shown that the genomes of present day humans outside Africa contain a few % from these extinct forms. Inside Africa, there is evidence of a similar admixture of archaic forms.

Mutation rates

- inferred from species comparisons: number of observed mutations, divergence estimates and generation times
- parent offspring trios: approx. 50 new mutations per individual, most of paternal origin.
- paleogenetics: comparing DNA sequences dating back 40, 30, 20 10 and 2 thousand years.



Green & Shapiro, ***Current Biology*** 2014



Veeramah & Hammer, ***Nature Reviews Genetics***, 2014

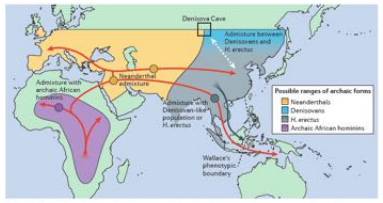
Key fossil evidence for nearly or fully anatomically modern humans (AMHs) is described on the left, and approximate date range estimates are indicated by the grey shading. West Africans are relatively genetically homogenous modern-day Niger–Kordofanian- and Nilo-Saharan-speaking populations that are often represented by the Yoruba of Nigeria. Eurasians encompass all modern-day non-African populations. Divergence times that are estimated using the faster phylogenetic mutation rate under the assumption of relatively instantaneous population splitting are mostly consistent with the fossil evidence. To preserve the correspondence between fossil dates and population divergence times under the slower pedigree-based mutation rate, this model assumes long-term gene flow among subdivided ancestral populations (represented by the gradient of blue shading), which leads to older divergence time estimates.

Types of Genetic Changes

	size of DNA sequences involved:
Chromosomal rearrangements (structural variation)	10^7 bp
Deletions, inversions, duplications, fusion, Fission, translocation	
Segmental duplication	10^4 - 10^6 bp
intra- and inter chromosomal	
Endogenous retroviruses	10^5 bp
Transposable elements	10^2 - 10^4 bp
LINE, SINE	
Simple Sequence Repeats (Microsatellites)	1 - 5 bp
SNP Single Nucleotide Polymorphisms	1 bp

Mutations are changes to the genome. These can be gigantic changes involving fusion or inversion of million of basepairs during chromosome rearrangements , or much smaller events all the way down to single basepair mutations, occurring as copying errors during DNA replication.

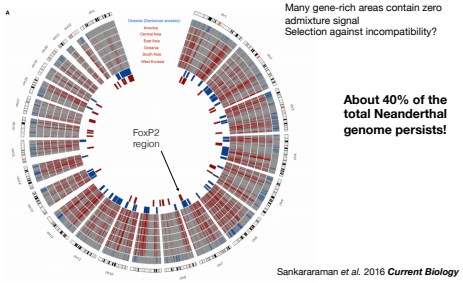
Hybridization happened



Veeramah & Hammer, ***Nature Reviews Genetics***, 2014

A possible model of archaic introgression based on the latest analysis using second-generation sequencing. Red arrows indicate initial colonization events across the Old World after the origination of anatomically modern humans (AMHs) in Africa, including two movements into Asia. Approximate positions of introgression events are represented by colored circles and are not intended to be accurate. This model portrays the hypothesis that portions of the Denisovan genome entered the human gene pool through hybridization with more widespread populations of archaic hominins (such as Homo erectus), which also interbred with the Denisovan population. The black arrow shows a more recent expansion of Asian farming populations (that is, <10,000 years ago) that did not carry introgressed Denisovan alleles and that replaced much of the indigenous resident population up to Wallace's phenotypic boundary (shown by the dashed line), which lies just east of Wallace's biogeographical line. This hypothesis may explain the lack of evidence for Denisovan introgression outside islands in Southeast Asia and Oceania.

“Deserts” of Archaic admixture

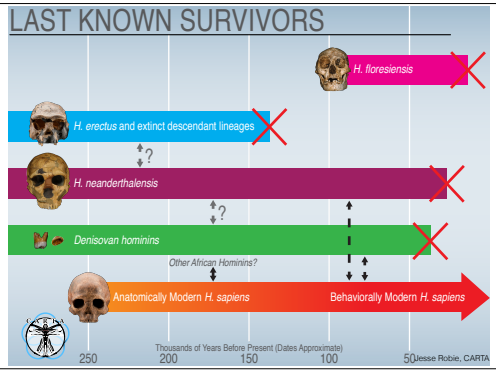


Distribution of Denisovan and Neanderthal DNA across the human chromosomes: The remaining archaic hominin DNA are concentrated in “gene deserts” areas with relatively few protein coding sequences and away from other know functional elements in the genome. This would indicate that there could have bee selection to purge such introgressed DNA. After .5 million years of independent evolution, one can safely expect that some of the Neanderthal/Denisovan genes, might have evolved too much to still “harmoniously” function with modern human genes.

Adaptive gifts from archaic hominids:

EPAS1 high altitude in Tibetans
BNC2 pigmentation in Europeans
POU2F3 keratinicyte proliferation in Europeans
HLA-C 15:05 allele disease resistance in Asia&Europe
TLR1/6/10 (Toll like receptors) innate immunity Asia & Europe
SLC16A11: Lipid metabolism (protection from starvation)Asia

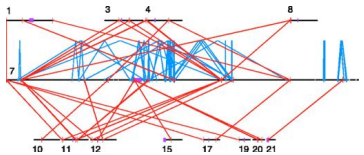
There are however, several genes that have been actively retained in the genomes of modern humans, some just in relatively few local populations:
EPAS1 in Tibetans and Sherpas in the Himalayan plateau appears to be a Denisovan variant that is highly adaptive for high altitude.
Several disease -resistance genes have also been co-opted after hybridization.



The big picture remains one, where we Homo sapiens have taken over the entire planet and remain the last hominin standing. Our precise role if any in the demise of several other hominins, the Neanderthals, Denisovans and *Homo floresiensis* is hotly debated.

Segmental Duplications – Crucibles of Primate Evolution?

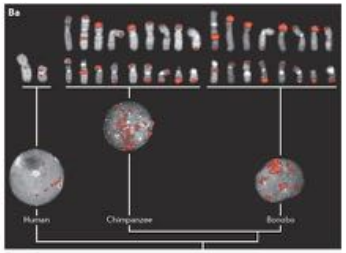
Recent (>40kb) segmental duplications on human chromosome 7



E. E. Eichler *et al.*, ***Science*** 301, 793 -797 (2003)

A schematic of recent segmental duplications (copy and pasting within or across chromosomes of large chunks of DNA) on human chromosome 7. The distribution of both interchromosomal (red) and intrachromosomal (blue) duplications is shown for human chromosome 7 (drawn ×50 to scale). Duplications (>90% sequence identity and >40 kb in length) correspond to duplications/gene conversion events that occurred over the last ~30 million years of human genome evolution.

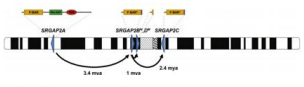
Segmental Duplication: differences in the Panid lineage: hyperexpansion of a 40kb segment

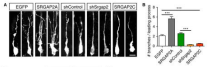


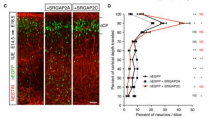
Bailey & Eichler ***Nature Reviews Genetics*** 2006

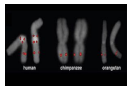
By using DNA probes that are specific for segmental duplication in chimpanzees and bonobos one can stain the many places where these chunks of DNA landed after copying themselves… This is an example where the human genome was much less affected.

Evolution of Human-specific Neural SRGAP2 Genes by Incomplete Segmental Duplication







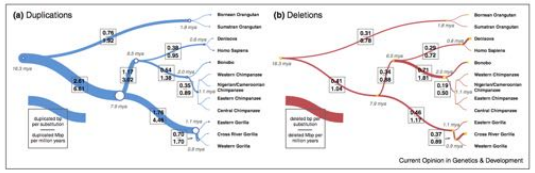


Dennis et al. CELL 2012, Eichler Group UW

Charrier et al. CELL 2012,Polleux, Scripps

Top: SLIT-ROBO Rho GTPase-activating protein 2 (srGAP2) also known as formin-binding protein 2 (FNBP2) is a protein that in humans is encoded by the SRGAP2 Schematic depicts location and orientation (blue triangles) of SRGAP2 paralogs on human chromosome 1 with putative protein products indicated above each based on cDNA sequencing. Asterisks indicate a 49 amino acid truncation of the F-BAR domain. Note that the orientation ofSRGAP2D remains uncertain, as the contig containing this paralog has not yet been anchored. Arrows trace the evolutionary history of SRGAP2duplication events. Copy number polymorphism and expression analyses suggest both paralogs at 1q21.1 (SRGAP2B and SRGAP2D) are pseudogenes, whereas the 1q32.1 (SRGAP2A) and 1p12 (SRGAP2C) paralogs are likely to encode functional proteins.Bottom. SRGAP2C Expression in Radially Migrating Mouse Cortical Neurons Phenocopies Srgap2 Knockdown (A) Confocal images of optically isolated neurons showing representative morphologies of radially migrating cortical neurons in E18.5 embryos following in uteroelectroporation (IUE) at E14.5 of the indicated constructs. sh, short hairpin. Scale bar, 10 mm. (B) Mean number of branches (±SEM) of the leading process of neurons as represented in (A). n = 3 animals/condition, 100–150 neurons/condition. (C) Low magnification confocal images of E18.5 cortical slices showing migration of in utero electroporated neurons expressing nuclear-EGFP (nEGFP) alone or together with SRGAP2A or SRGAP2C. Staining with anti-GFP shows the position of the electroporated neurons, and anti-NESTIN marks the radial glial scaffold.dCP, dense Cortical Plate. (D) Quantification of neuron distribution in cortical slices as illustrated in (C) (mean ±SEM). n = 3 animals/condition, 9–10 slices/condition. In (B) and (D), *p < 0.05; **p < 0.01; ***p < 0.001; NS (not significant, p > 0.05); Mann-Whitney test.

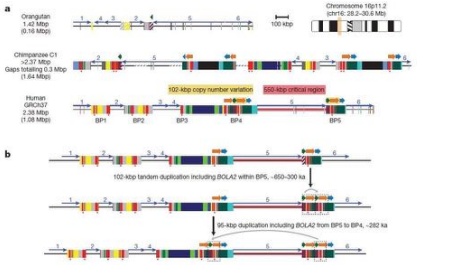
Primate rates of duplication and deletion



Dennis and Eichler 2016 ***Curr. Opinion in genetics and Development***

Primate rates of duplication and deletion. Rates of fixed (a) duplications and (b) deletions are shown as a function of the number of substitutions along each branch of the great ape phylogeny. Branch widths are scaled proportionally to the number of duplicated base pairs per substituted base pair based on analysis of 97 human/ape genomes. A burst of duplicated base pairs appears to have occurred in the common ancestral branch leading to humans and African great apes, where duplicated base pairs were added at 2.6-fold the rate of substitution. In contrast, the rate of deletion in the great ape lineage is more clocklike along all branches (mean of 0.32 deleted base pairs per substitution) with the exception of the chimpanzee–bonobo ancestral lineage, where an approximate twofold increase in the rate of deletion is observed (0.71 deleted base pairs per substitution).

Complex set of segmental Duplications on Chromosome 16

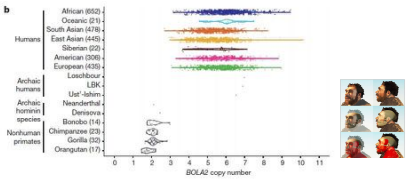


Nuttle *et al.* ***Nature*** 2016

a) Schematic depicts the genomic organization of chromosome 16p11.2 for one orangutan and two chimpanzee haplotypes along with the human reference haplotype (GRCh37 chr16:28195661–30573128; see ideogram for approximate chromosomal location). Blocks of segmental duplications within this locus mediate recurrent rearrangements in humans; thus, these blocks have been defined as breakpoint regions BP1–BP5). The ~550 kbp critical region (pink) and a >1 Mbp chimpanzee-specific inversion polymorphism (orange) are highlighted. Tiling paths of sequenced clones are indicated above each haplotype, with chimpanzee clones that could not be fully resolved marked with asterisks. Colored boxes and thick arrows indicate the extent and orientation of segmental duplications (with different colors denoting duplicons from different ancestral genomic loci, and hashed boxes indicating sequence duplicated in humans but not in the species represented). Thin numbered arrows show orientations of gene-rich regions of unique sequence. Numbers (left) indicate the size of each orthologous haplotype, with the number of segmentally duplicated base pairs shown in parentheses. Note that, for chimpanzee, these sizes are lower bounds due to gaps in the contigs (dotted line sections) and the contigs not reaching unique sequence beyond BP1 (i.e., unique region 1). b) Schematic depicts distinct human structural haplotypes over the chromosome 16p11.2 critical region and flanking sequences (three complete haplotypes extending from unique sequence distal to BP3 to unique sequence proximal to BP5 and one partial haplotype including BP3–BP4 and BP5 sequence contigs). High-quality sequence for each haplotype was generated by sequencing a total of 40 BACs and 15 fosmids from three different human genomic libraries. Regions of copy number variation (highlighted in yellow along the first two haplotypes) occur on both sides of the critical region and involve the same 102 kbp unit in direct orientation, including a 30 kbp block containing BOLA2 and two other genes and a 72 kbp block harboring a partial segmental duplication of SMG1 (SMG1P). Expansion and contraction of this cassette underlie hundreds of kbp of structural diversity between human haplotypes. BOLA2 paralog-specific copy number genotype data suggest that H1 and H3 likely represent the most common haplotype structures in humans.

Complex set of segmental Duplications on Chromosome 16

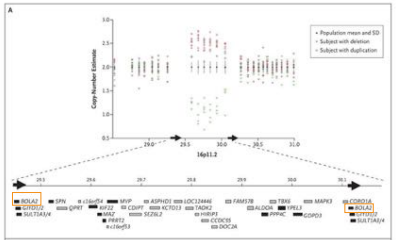
281 kya this event simultaneously increased copy number of gene associated with iron homeostasis and predisposed our species to recurrent rearrangement associated with disease.



Nuttle *et al.* ***Nature*** 2016

Diploid copy number estimates (points) for BOLA2 based on sequence read depth12 are shown for 2,359 humans, three archaic humans13,14, a Neanderthal2, a Denisovan3, and 86 nonhuman primates, with violin plots overlaid. c) Paralog-specific BOLA2 copy number genotypes (points, jittered around their integer values) were inferred from Whole Genome Sequencing read depth over informative markers for 222 individuals sequenced to high coverage. Colors correspond to different populations as in panel b.

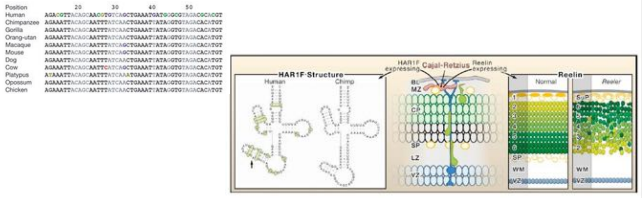
Association between Microdeletion and Microduplication at 16p11.2 and Autism



Weiss LA et al. ***N Engl J Med*** 2008;358:667-675.

Deviation from BOLA2 copy number are associated with psychiatric disease. This is one of the strongest evidence to date, that Neanderthal DNA could have been a liability to neurotypical development in modern humans.

HAR 1 human accelerated region 1



18/118 human specific changes to an RNA coding region

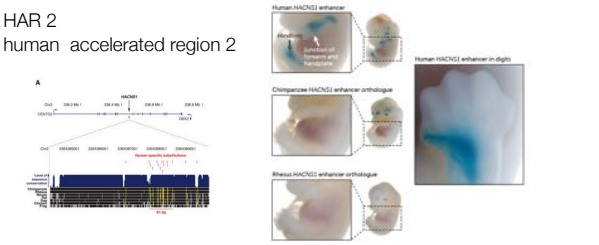
expression during fetal brain development in cortex

Pollard et al. 2006 ***NATURE***
Haeussler Group UCSC

Speed demon sequences: Human accelerated regions HARs:
These DNA sequences are virtually identical from chicken to chimpanzee, but then show many changes in humans
HAR1 is expressed doing embryonal brain development and appear to affect the formation ion the six cortical layers in the brain.
It does not encode a protein but rather two different small functional RNAs, one in each direction of reading….

Human Accelerated Non-coding regions
human sequence expressed in mouse embryos dramatic affect on limb development

HAR 2
human accelerated region 2

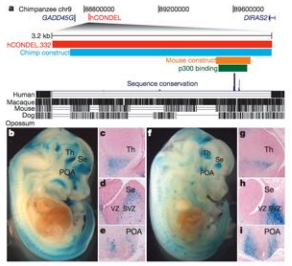


Prabhakar S et al ***Science***, 2009, Noonan Group, Yale

Expression patterns obtained from HACNS1 and its chimpanzee ortholog in E13.5 mouse embryos. Three embryos resulting from independent transgene integration events are shown for each construct. Close-up views of forelimb and hindlimb expression in a representative embryo for each construct are shown at left, and arrows indicate positions where limb expression is present or absent. B. Dorsal view of reporter gene expression in the distal anterior forelimb of a HACNS1 E13.5 transgenic embryo. Arrows indicate the most anterior digit.

Enhancer deletion of tumor suppressor gene GADD45G

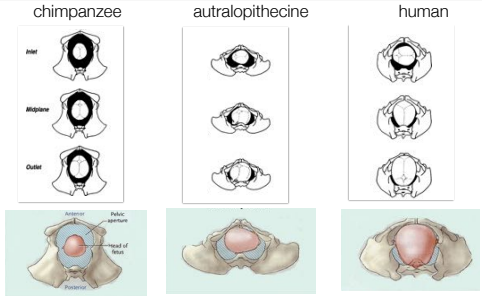
hCONDEL in enhancer of tumor suppressor gene Growth arrest and DNA-damage-inducible, gamma (GADD45G) results in prolonged neuronal cell division which may contribute to thalamic and cortical expansion in humans.



McLean et al. ***Nature***, 2011. Kingsley Group, Stanford

Transgenic analysis of a chimpanzee and mouse forebrain enhancer missing from a tumour suppressor gene in humans. a, Top panel: 1.3Mb region of the chimpanzee chromosome 9. The red bar illustrates a 3,181 bp human-specific deletion removing a conserved chimpanzee enhancer located downstream ofGADD45G. Bottom panel: multiple species comparisonof the deleted region, showing sequences aligned between chimpanzee and other mammals. The green bar represents a mouse forebrain-specific p300
binding site18, and the blue and orange bars represent chimpanzee and mouse sequences tested for enhancer activity in transgenic mice. The chimpanzee (b–e) and mouse sequence (f–i) both drive consistent lacZ expression in E14.5 mouse embryos in the ventral thalamus (c,g), the SVZ of the septum (d,h), and the preoptic area (e, i). Increased production of neuronal subtypes from these regions may contribute to thalamic and cortical expansion in humans27–30. All sections are sagittal with anterior to right. POA, preoptic area; Se, septum; SVZ,
subventricular zone; Th, thalamus; VZ, ventricular zone.

Obstetrics: birthing a big-brained baby

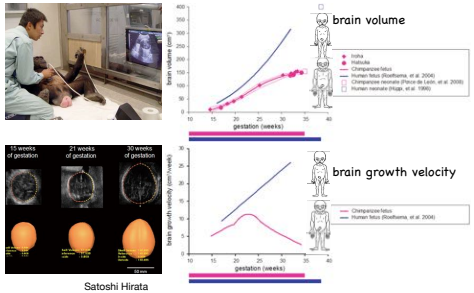


Humans have hit the limit: bipedality imposes an upper limit to baby head size.

Human birth has become very risky, as human babies have to rotate their heads in order to clear the pelvic passage of their mothers.

Regular C-sections are now lifting this limit, prediction: in population where c0section rates are very high, there will soon be many more babies with heads that cannot possibly clear the pelvic opening of their mothers (cephalo-pelvic disproportion)

How to grow a big brain



Satoshi Hirata

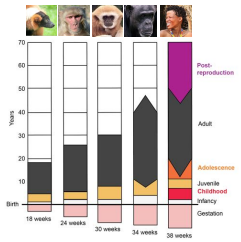
Sakai et al. **Current Biology**, 2012

Gestational age-related changes in brain volume in chimpanzee (Hatsuka and Iroha) and human fetuses.

Gestational age-related changes in the growth velocity of brain volume in chimpanzee and human fetuses

Chimpanzee brains start slowing down their growth in mid-pregnancy, humans on the other hand continue a high fetal rate for a full year after birth.

Childhood, Adolescence & Post-reproductive Survival



- Adapted to cultural opportunities?
- Nutritional opportunities?
- Facilitated by stronger pair bonds between parents?
- Facilitated by allomothering?

redrawn from Schultz 1963
Boaz & Almquist 1999

Odd life history compared to other primates

Humans have delayed development: humans invented childhood (slow body but rapid brain growth), adolescent growth spurt, and prolonged post-reproductive survival. But evolved shorter inter-birth intervals than apes!

Derived Human Growth Schedule

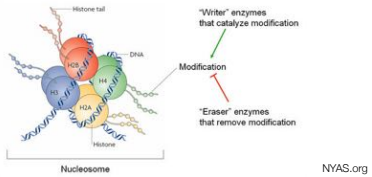
- Delay allows increased transmission of behavior and concepts.
- Human minds are effective copying devices and idea generators.
- Language is one of the major target of imitation and idea transmission.
- Delayed development: biological assimilation of culture?
- Paradoxically shorter Inter-birth-Interval than apes.

Minds as copying machines and idea generators

Humans over-imitate, focusing as much on the way than on the goal, chimps go for the goal.

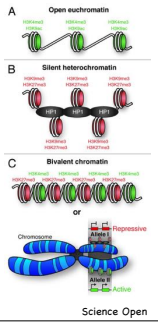
Ratcheting culture: build on old ideas with new ideas….

"Histone Code" & Epigenetics



Wrapping of DNA around histones and modification of DNA and histones affect gene expression (methylation of DNA, acetylation, phosphorylation and methylation of histones)

“Histone Code” & Epigenetics

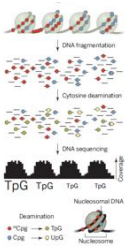


Histone deacetylase in royal jelly

Spannhof et al. EMBO Reports 2011

One of the best illustrations of an epigenetic effect: roll jelly made by glands in the worker bees can turn any fertilized egg from a worker bee into a queen. The secretions contains a deacetylase enzyme that can modify the histone code (chemical modifications on the histone tails, that can prevent the activity of the DNA wrapped around these…

Ancient Epigenetics



DNA damage in ancient genomes may provide insight into past regulatory changes in humans and other species

Ludovic and Willerslev, ***Science*** 2014

There are the first few studies on ancient epigenetics…

Summary

- Individual genomes are unique genetic mosaics.
- Each piece of DNA has its distinct history.
- Complete genome sequences for many apes allow:
  - Reconstruction of past population histories
  - Finding changes that define the human species.
  - Humans are enriched for changes affecting brain development, incl. genes involved in uniquely human diseases.
  - Fossil DNA data, experiments in model animals and cell culture allow testing for biological effects.
  - Experimental approaches to neurodevelopment are severely limited for ethical reasons.