EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Language Evolution WiSe 2023/2024

Lecture 5: Human Evolution III
Genetics

07/11/2023, Christian Bentz

Overview

Section 1: Recap

Taxonomic Considerations

How are species classified?

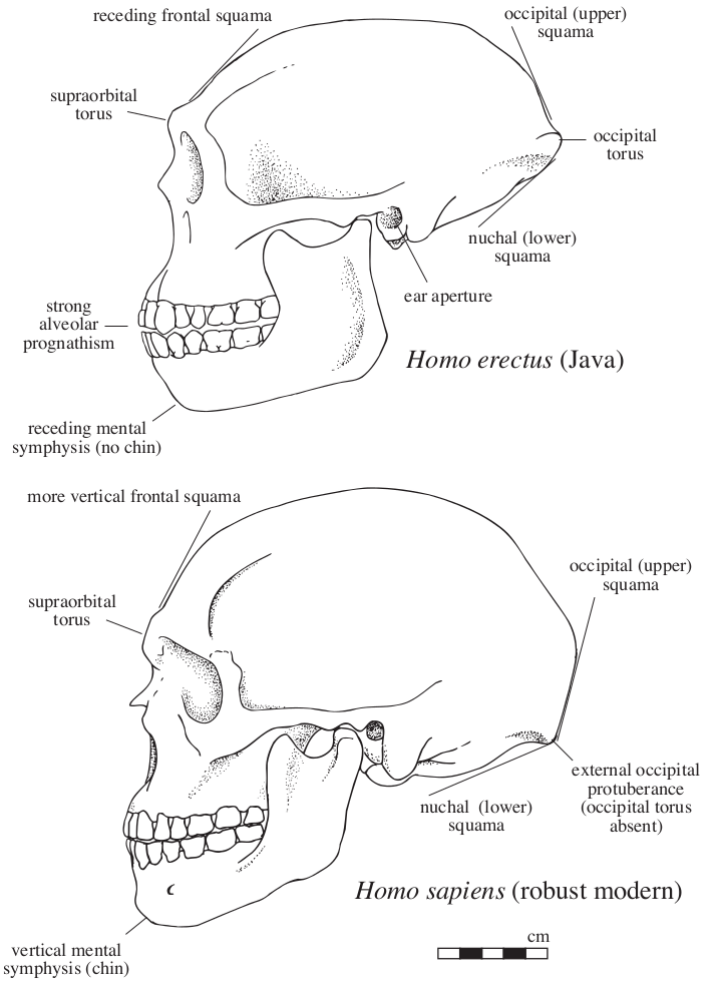
- **Morphometrics**: Analysing and comparing the morphological shape of fossils.
- **Behavior**: Analysing archaeological assemblages (mostly stone tools).
- **Genetics**: Analyses of different parts of the (available) DNA, applying phylogenetic methods from evolutionary biology.

Cranial Morphology

Typical features for distinction:

- **Globularity** (globular vs. flat)
- **Occipital torus** (rounded vs. angled)
- **Supraorbital torus** (robust vs. less pronounced)
- **Mandibular chin** (vertical vs. receding)
- **Facial angle** (orthognathic vs. prognathic)

Note: Some anatomical variation due to sexual dimorphism.



Klein (2009).

Taxonomic Considerations

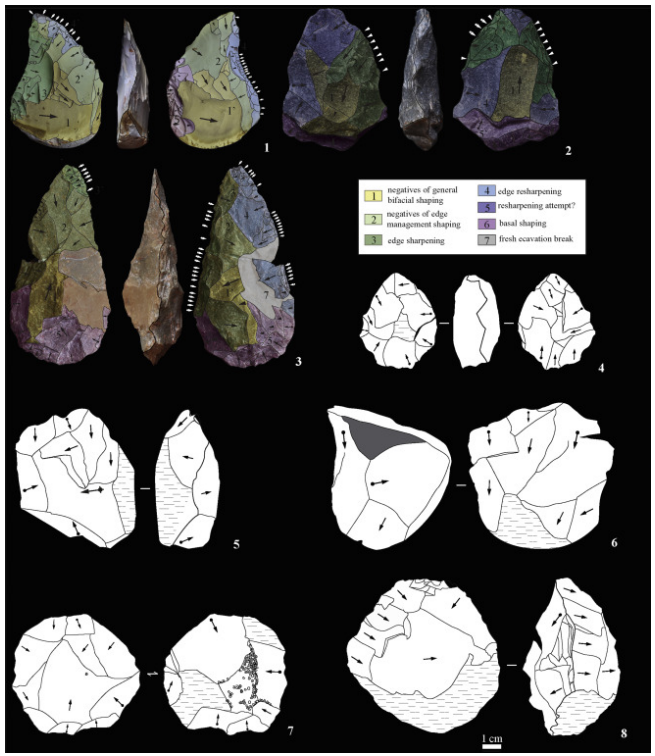
How are species classified?

- **Morphometrics**: Analysing and comparing the morphological shape of fossils.
- **Behavior**: Analysing archaeological assemblages (mostly stone tools).
- **Genetics**: Analyses of different parts of the (available) DNA, applying phylogenetic methods from evolutionary biology.

Acheulean industry
(c. 1.7 Mya - 130 Kya)

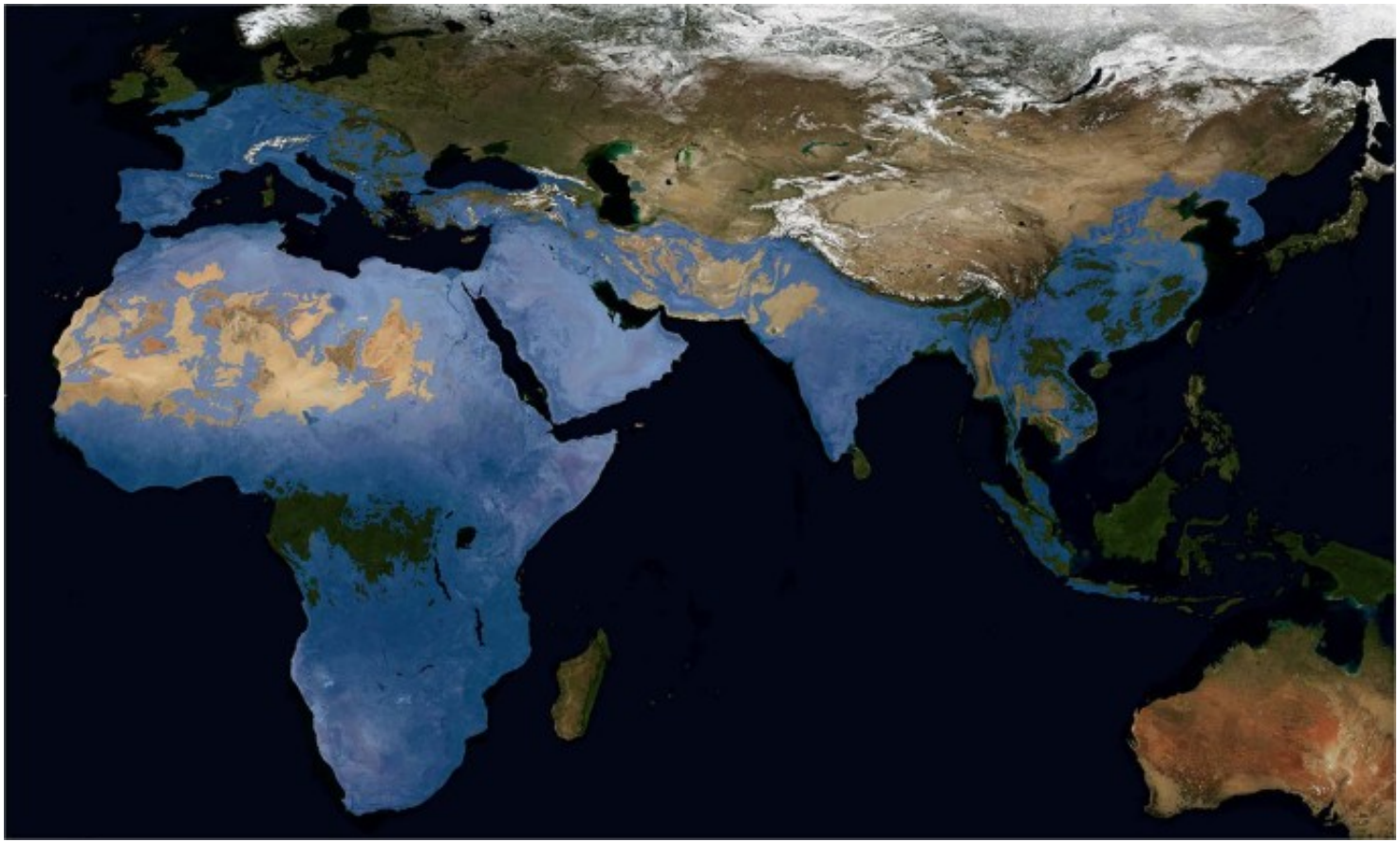
Assemblages are dominated by **bifacial handaxes** of different sizes. These were produced by systematically flaking off parts of the core and retouching, until an often symmetrical "tear-drop" shaped tool is achieved. Versatile tools of different shapes and sizes with many purposes (including hunting).

Daura et al. (2013). A 400,000-year-old Acheulean assemblage associated with the Aroeira-3 human cranium.



Left: Handaxe from Europe; Middle: Handaxe from Bose, China; Right: Handaxe from India.
https://humanorigins.si.edu/evidence/behavior/stone-tools

Distribution



Key et al. (2023). Modelling the end of the Acheulean at global and continental levels suggests widespread persistence into the Middle Palaeolithic.

Taxonomic Considerations

How are species classified?

- **Morphometrics**: Analysing and comparing the morphological shape of fossils.
- **Behavior**: Analysing archaeological assemblages (mostly stone tools).
- **Genetics**: Analyses of different parts of the (available) DNA, applying phylogenetic methods from evolutionary biology.

Section 2: Basics of Genetics

Some useful Resources

Glossary of Terms:
`https://www.genome.gov/genetics-glossary`

Human Genome Project:
`https://www.genome.gov/human-genome-project`

Genome Data Viewer:
`https://www.ncbi.nlm.nih.gov/genome/gdv/`

Videos and info:
`https://www.yourgenome.org/`

The Cell (Eukaryote)

The genetic information resides in

- the nucleus (**DNA**),
- mitochondria (**mtDNA**),
- chloroplasts (**cpDNA**).

Note: Chloroplasts only exist in plant cells for photosynthesis, i.e. are not relevant here.

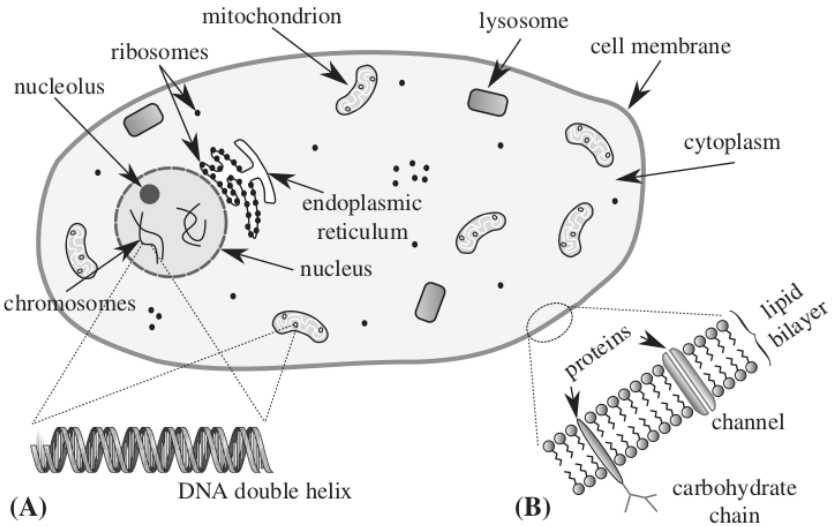
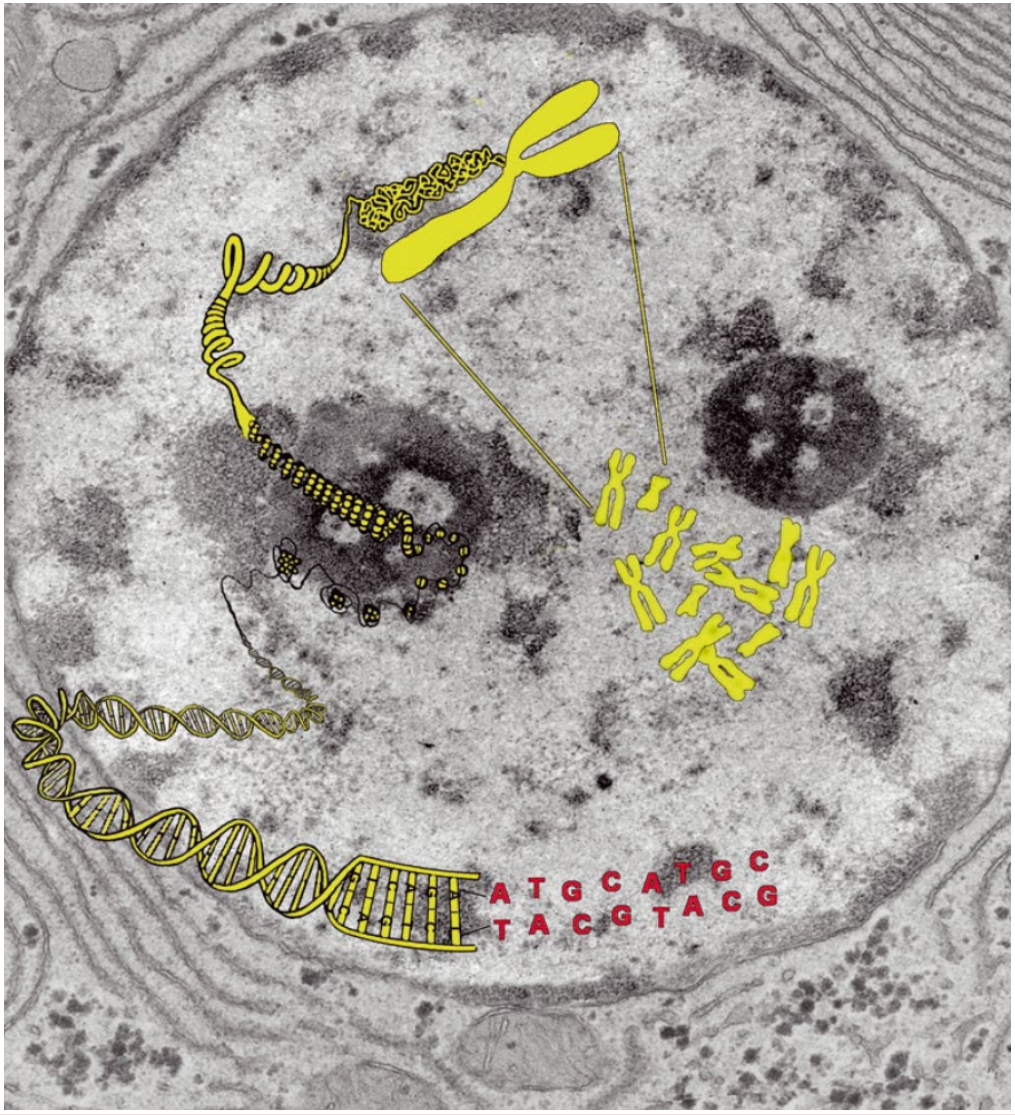


Figure 3.1 The structure of a generalized eukaryotic (animal) cell. The cell is bounded by a membrane (insert **B**) composed of two layers of lipids with various types of embedded proteins allowing matter, energy and information exchanges between the cell and its environment. In the cytoplasm there are various structures such as the nucleus containing most of the genetic material (DNA double helix, insert **A**) structured into several discrete linear chromosomes, mitochondria (with their own tiny genetic material structured as a circular molecule), lysosomes, the endoplasmic reticulum and ribosomes.

Dediu (2015). An introduction to genetics for language scientists, p. 47.

The Nucleus
From DNA to Chromosomes



DNA (DeoxyriboNucleic Acid)

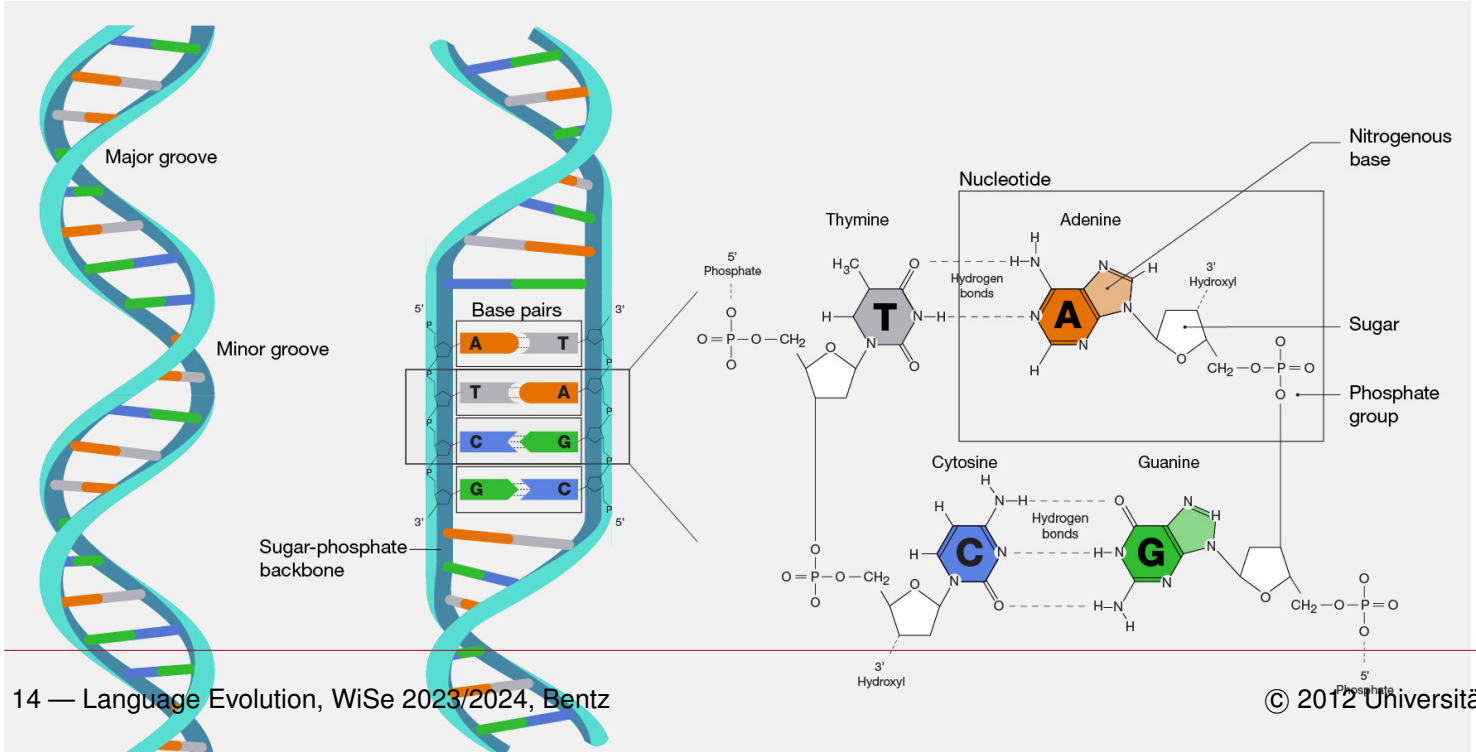
Base pair

Consits of two nucleotides bound together.

Nucleotides

A base (**A**denine, **G**uanine, **C**ytosine, **T**hymine) plus a sugar.

See also video: `https://www.youtube.com/watch?v=o_-6JXLYS-k`



Mitosis (DNA Replication)

A cell can divide itself in a process called **mitosis**. This creates a new cell with identical DNA (bare any mutations).

Dediu (2015). An introduction to genetics for language scientists, p. 52.

For a video on DNA replication see:
`https://www.youtube.com/watch?v=TNKWgcFPHqw`

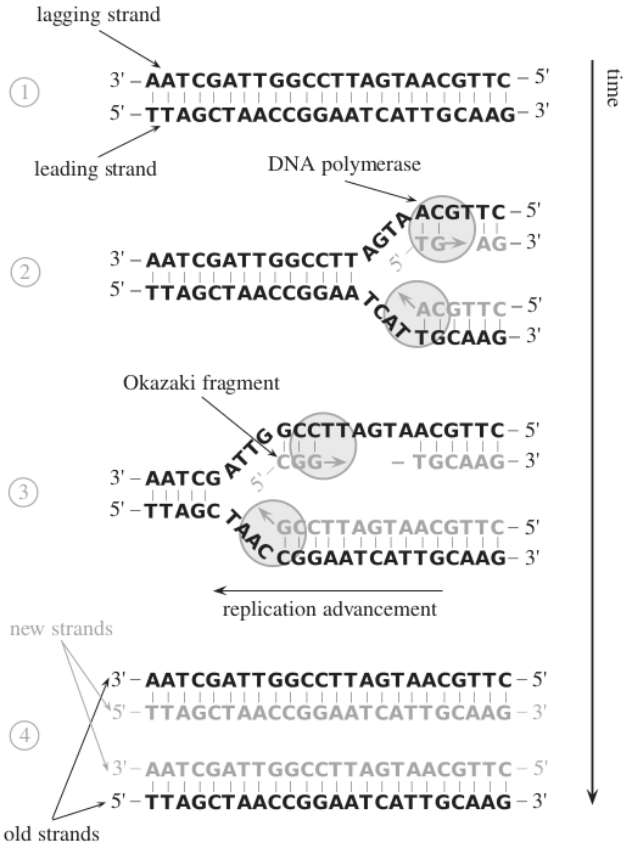


Figure 3.3 A simplified depiction of DNA replication. In black are the old DNA strands and in grey the new ones. The grey circles represent DNA polymerase while grey arrows stand for the direction in which the new DNA strands grow. Time step 1 shows the old DNA double-stranded molecule before replication begins, while time steps 2 and 3 show the advance of the replication fork and the elongation of the leading and lagging strands. The final time step (4) shows the two daughter double-stranded DNA molecules, each composed of one old (black) and one new (grey) DNA strand.

Genes

The term **gene** refers to regions on the DNA strand whose information is transcribed and translated into amino acids which are the building blocks of *proteins*. Humans have ca. **20,000** genes coding for proteins. The regions relevant for transcription are called **exons** (**ex**pressed regi**ons**), and are interspersed with **introns** (**intr**agenic regi**ons**).

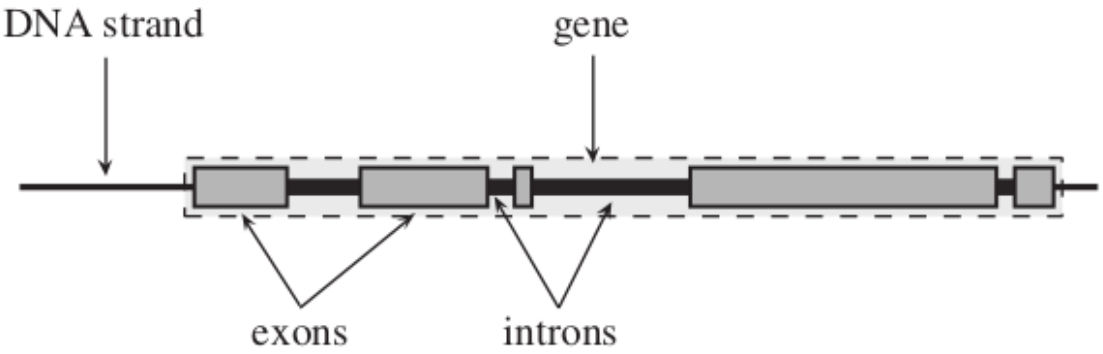


Figure 3.14 Schematic representation of a gene composed of several exons and introns.

Dediu (2015). An introduction to genetics for language scientists, p. 68.

Chromosomes

The DNA double helix is folded and wrapped around proteins (histones), and then packaged into **chromosomes**. The DNA strand in a chromosome is very long, one of the longest measuring ca. 8.5 cm. In comparison, the size of a cell is ca. 10 μm, with a μm being one millionth of a metre.

Dediu (2015). An introduction to genetics for language scientists, p. 53.

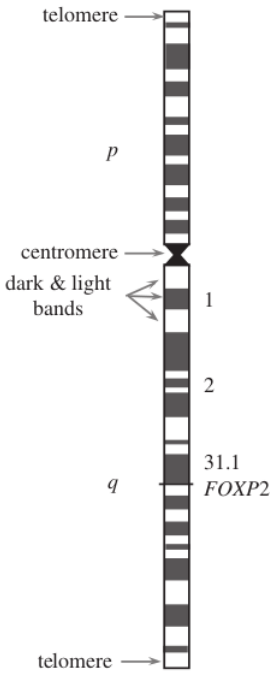
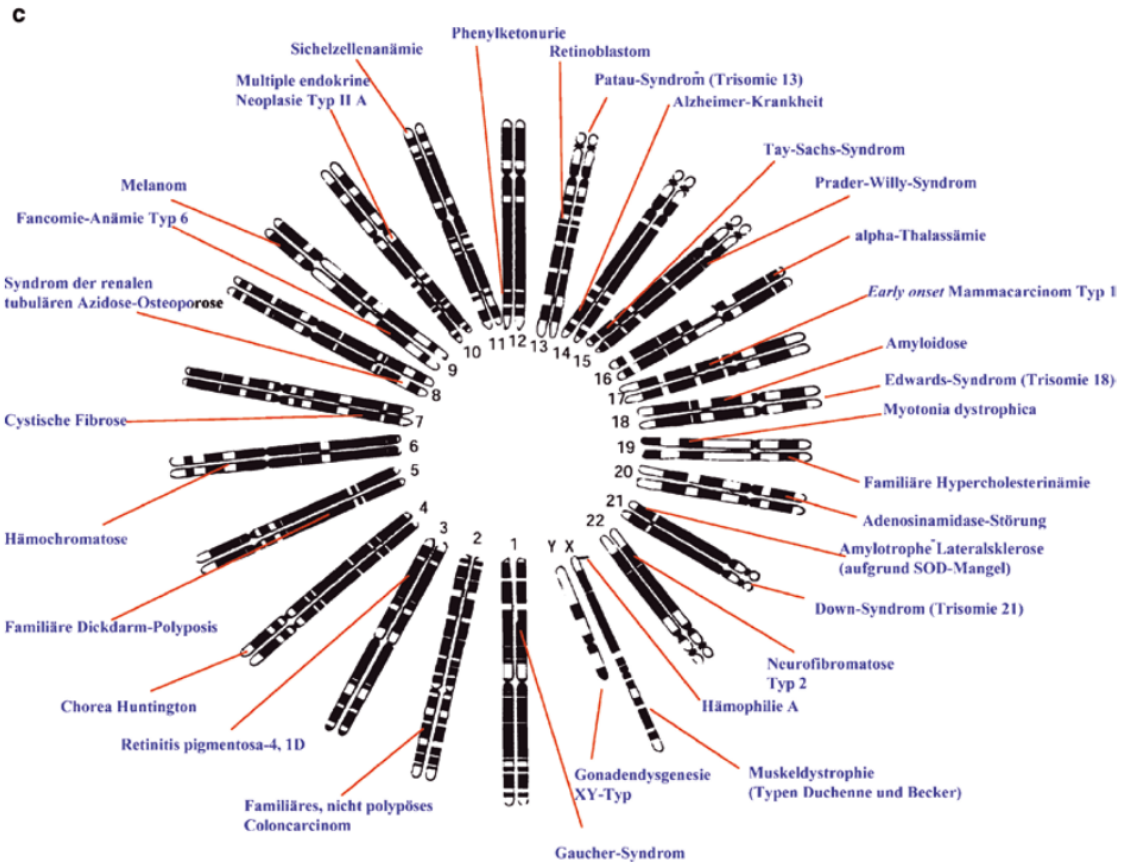


Figure 3.5 A graphical representation of chromosome 7, also showing the position of the *FOXP2* gene denoted as 7q31.1.

Chromosomes

There are overall **46 human chromosomes**, coming in **23 pairs**. The pairing of chromosomes is the reason why they are called **diploid**. 22 of these pairs are called *autosomes* (numbered 1 to 22 in decreasing order of size), and the last pairs is are the **sex chromosomes (X, Y)**.

Dediu (2015). An introduction to genetics for language scientists, p. 53.



Storch et al. (2013). Evolutionsbiologie, p. 257.

Exercise

Go the the *Gene Data Viewer* at `https://www.ncbi.nlm.nih.gov/genome/gdv/` and find the FOXP2 gene for the modern human, chimpanzee, and mouse reference genomes. Try to answer the following questions: How many *exons* does the FOXP2 gene have in each species? What is the length of the FOXP2 gene in terms of *nucleotides* for these three species. What could the differences stem from?

Section 3: Sexual Reproduction

Meiosis

Sexually adult individuals built **gametes**, i.e. cells specialized for reproduction, in a process called **meiosis**. In the case of females these gametes are called *ova*, and for males *sperm*. They contain half the genetic material of each individual, that is, they contain 23 non-paired chromosomes, which are **randomly** allocated (aka Mendel's *First Law of segregation*). The gametes are therefore called **haploid**.

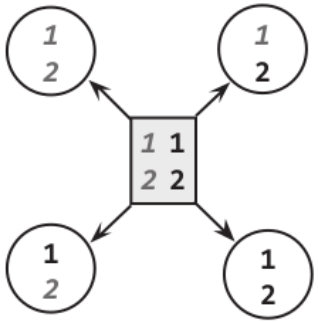


Figure 3.6 Gametogenesis: going from a diploid adult to the haploid gametes. For simplicity, we are focusing on just two chromosomes, 1 and 2, and the two individual members of the chromosome pairs in the adult (square in the middle) are differentiated using *italic* and grey colour. These individual chromosomes are transmitted to the resulting gametes (circles) independently, resulting in four possible such gametes.

Note: This is a simplified depiction of gametogenesis with just 2 out of 23 chromosomes. Hence, there are 4 possibilities here. With all 23 chromosomes there are $2^{23} = 8,388,608$ possibilities.

Dediu (2015). An introduction to genetics for language scientists, p. 56.

Recombination

Besides random allocation of the homologous chromosomes of a diploid parent in *meiosis* (gametogenesis), they are also **recombined**. In a nutshell, this means that the chromosomes inherited from the parents of the parent are crossed over to yield a new mix of the DNA strands. This gives rise to a huge number of possible recombinations of DNA substrands.

Dediu (2015). An introduction to genetics for language scientists, p. 58-59.

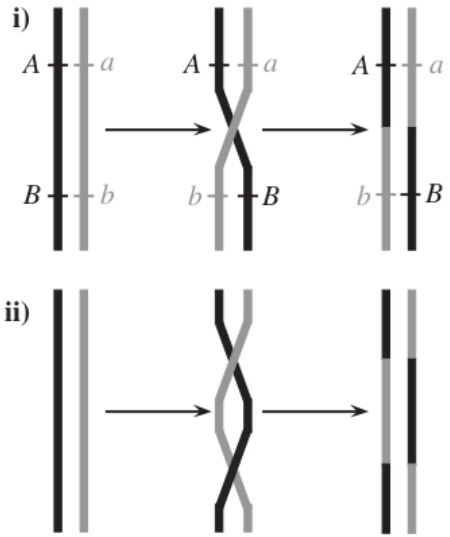


Figure 3.8 A highly simplified representation of chromosomal crossover. Left: the two original homologous parental chromosomes (black and grey) are paired. Middle: they cross over once (top) or twice (bottom) to result in two new recombinant chromosomes (right).

Note: The differences in genetic **loci**, here denoted as A, B, a, b, could correspond to a point mutation of a single nucleotide, or several nucleotides, whole genes, and even observable effects such as “this locus causes blue eyes in the phenotype”.

Sexual Reproduction

In the process of **fertilization** a sperm and an ova fuse and produce a so-called *zygote*, which is then again **diploid**. There are $2^{23} \times 2 = 16,777,216$ different possible unique zygotes. One of these will develop into offspring.

Dediu (2015). An introduction to genetics for language scientists, p. 57.

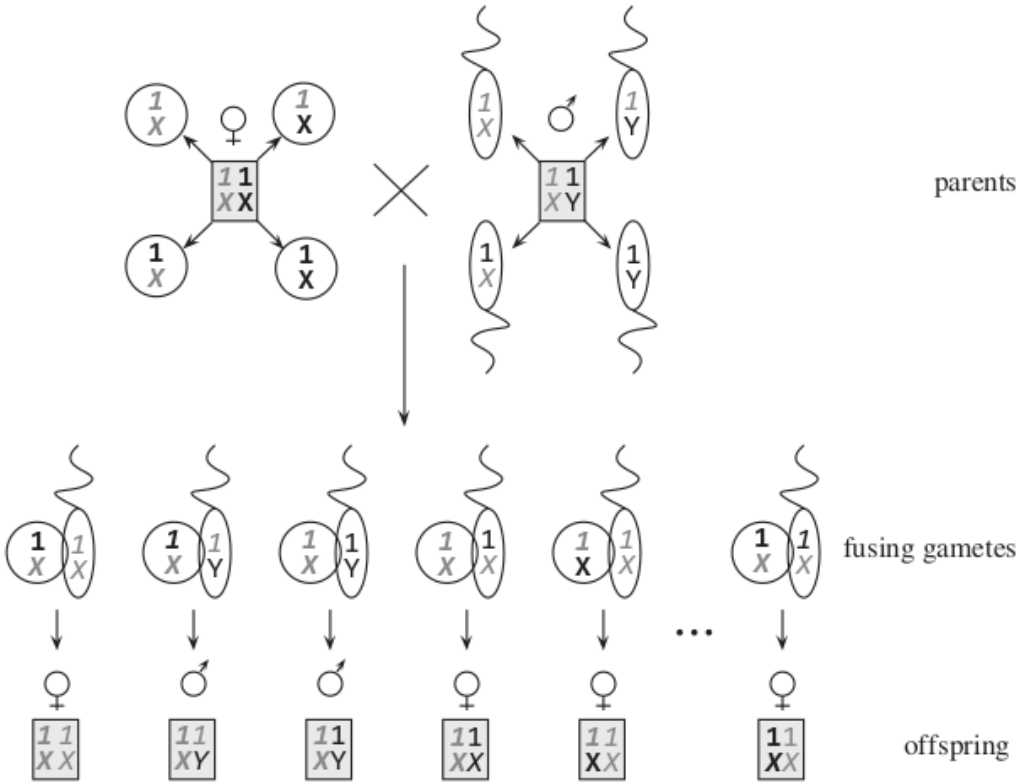


Figure 3.7 The parents (male and female, top) produce gametes which fuse (middle) to produce offspring (bottom). Here we focus on autosome 1 and on the sex chromosomes. For illustration purposes, the female parent's chromosomes are **bold** while the male parent's chromosomes are regular, and within each, one member of the pair is *grey italic* and the other black upright. Only six of the possible 16 zygotes are shown.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Faculty of Philosophy
General Linguistics

Section 5: Gene Expression

Gene Transcription

The expressed regions (exons), i.e. genes, on a DNA strand are firstly *transcribed* into messenger **RNA (mRNA)**. This is very similar to DNA replication, except that Thymine (T) is replaced by Uracil (U). The mRNA travels from the nucleus to the cytoplasm and is there translated into amino acids.

Dediu (2015). An introduction to genetics for language scientists, p. 63.

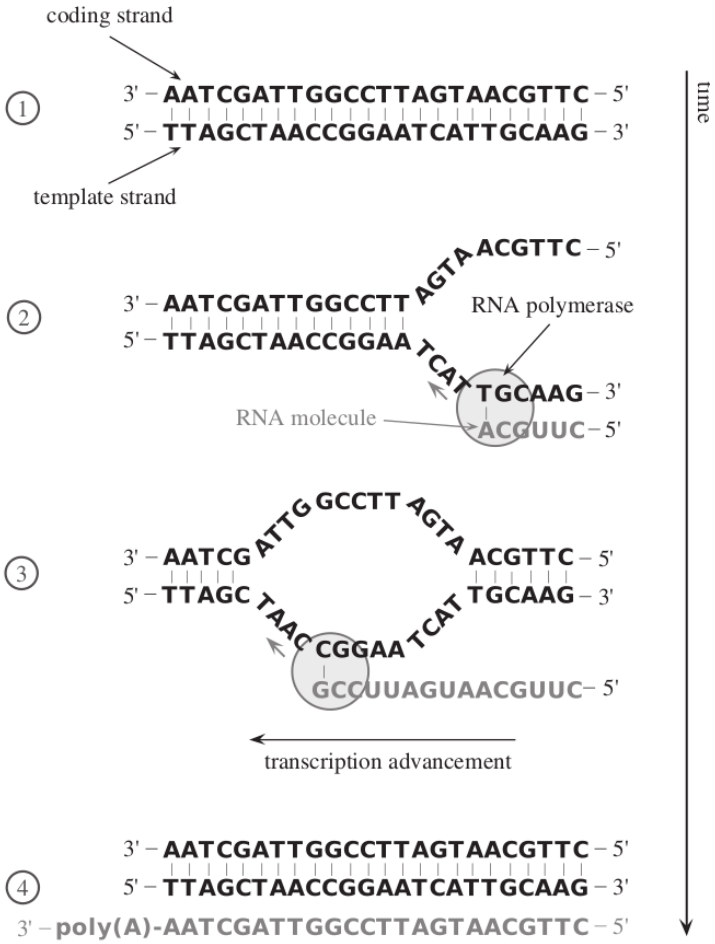


Figure 3.11 Schematic representation transcription.

Gene Translation

The mRNA is translated (in blocks of three bases) into **20 amino acids**, the building blocks of proteins. **Proteins** then take on a multitude of functions in the body: transporting oxygen (haemoglobin), shaping cells and the body, movement of muscle fibres, signalling (neurotransmitters and hormones).

Dediu (2015). An introduction to genetics for language scientists, p. 64-67.

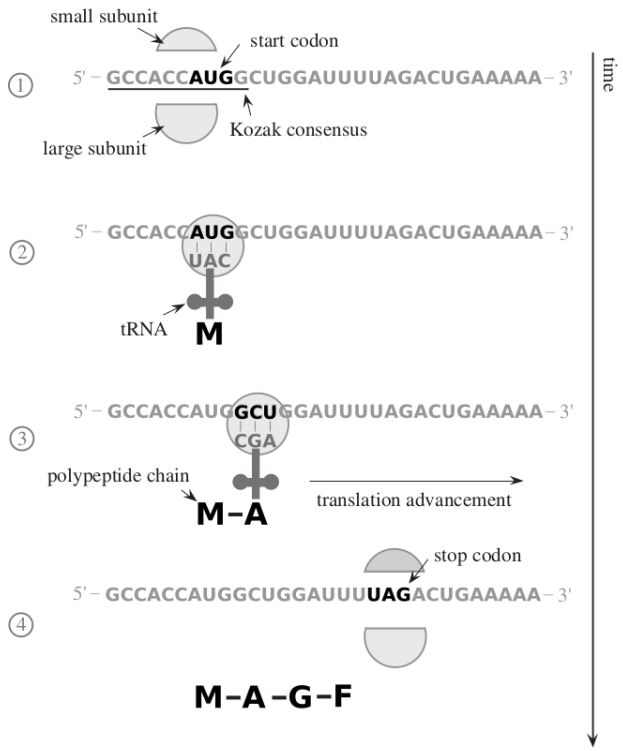


Figure 3.12 Schematic representation of translation. The ribosome is composed of two subunits (large and small) that assemble and begin the translation process with the start codon **AUG**. Translation advances in three-nucleotide units (codons) in the 5′ → 3′ direction (here, left → right) and consists in the elongation of the polypeptide chain by adding the amino acid carried by the tRNA corresponding to the current codon.

For a brief introduction to gene transcription and translation see video at:
`https://www.youtube.com/watch?v=gG7uCskUOrA`

Overview: Amino acids

Table 3.1 *The 20 amino acids with their full names and 3- and 1-letter symbols, ordered by full name. A ◂ marks the* essential *amino acids, i.e., those amino acids that the body cannot synthesize and must be ingested through food.*

Full name	3-letter symbol	1-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine ◂	His	H
Isoleucine ◂	Ile	I
Leucine ◂	Leu	L
Lysine ◂	Lys	K
Methionine ◂	Met	M
Phenylalanine ◂	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine ◂	Thr	T
Tryptophan ◂	Trp	W
Tyrosine	Tyr	Y
Valine ◂	Val	V

Dediu (2015), p. 62.

Overview: Codons and Amino Acids

Table 3.2 *The genetic code showing for each possible triplet or codon (tri-nucleotide "word") the corresponding amino acid (represented by its one-letter symbol) or ■ for the STOP codons. The first column gives the first nucleotide ("letter") of the codons, the first row gives the second nucleotide, and the last column the third one.*

1st	2nd								3rd
	U		C		A		G		
U	UUU	F	UCU	S	UAU	Y	UGU	C	U
	UUC		UCC		UAC		UGC		C
	UUA	L	UCA		UAA	■	UGA	■	A
	UUG		UCG		UAG		UGG	W	G
C	CUU	L	CCU	P	CAU	H	CGU	R	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Q	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	I	ACU	T	AAU	N	AGU	S	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	K	AGA	R	A
	AUG	M	ACG		AAG		AGG		G
G	GUU	V	GCU	A	GAU	D	GGU	G	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	E	GGA		A
	GUG		GCG		GAG		GGG		G

Dediu (2015), p. 64.

Section 5: Mutations

Types of Mutations

Mutations come in different types. Depending on the type and **locus** of a mutation, the actual impact on gene expression and the phenotype of an organism can be anything from negligible to fatal.

- **Point mutations**: a single nucleotide is replaced for another in an exon on the DNA strand (e.g. A $\rightarrow$ C).
- **Frameshift mutation**: A single nucleotide is deleted or inserted.
- **Chromosomal mutation**: Whole chromosomes or parts of chromosomes can be deleted, duplicated, inverted or translocated.

Point Mutation

In *mitosis* or *meiosis*, one nucleotide might be erroneously replaced by another in a **point mutation** (with very low probability: ca. one per one billion replicated nucleotides).

Importantly: A point mutation creates a new alternative version of the DNA sequence, called an **allele**.

Dediu (2015). An introduction to genetics for language scientists, p. 53.

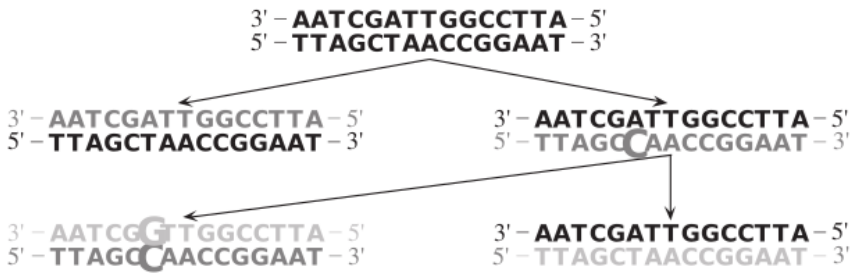


Figure 3.4 A point mutation. Top: the original double-stranded DNA molecule, undergoing semi-conservative replication to produce an identical DNA daughter molecule (left) and another DNA daughter molecule (right) carrying a **T** → **C** mutation (exaggerated size). As before, black, dark and light grey represent the old and new strands. Subsequently, this DNA produces two daughter molecules, one identical to the original (right) and one mutant (left), which has replaced the original pair $\frac{A}{T}$ by the new $\frac{G}{C}$ pair.

Note: A point mutation might have no effect at all on gene expression. For example, if it changes a codon from CUC → CUA, then both will code for the amino acid Leucine (L). In other cases, e.g. CUC → GUC, it will code for a different amino acid Valine (V), and hence have an effect on gene expression.

Frameshift Mutation

If nucleotides are deleted or added in a **frameshift mutation**, then the frame for transcription and translation changes. This likely has a drastic effect on gene expression, as the original amino acid coding will be seriously altered.

Dediu (2015). An introduction to genetics for language scientists, p. 67.

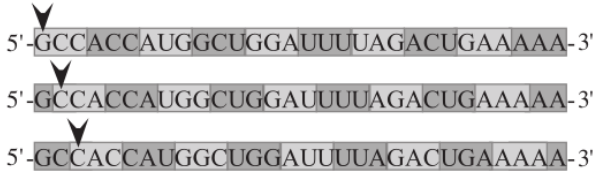


Figure 3.13 Representation of the three possible reading frames in the three rows of the figure. The arrow shows the first letter of a codon and, for visual contrast only, the neighbouring codons are highlighted using light and dark grey.

Note: This type of mutation is involved in a series of diseases, e.g. the Tay-Sachs disease (`https://en.wikipedia.org/wiki/Tay-Sachs_disease`)

Chromosomal Mutation

If whole chromosomes, or larger parts of these, are deleted, duplicated, translocated, or inverted in **chromosomal mutation**, this has invariably a strong impact on gene expression and the phenotype of an organism.

Dediu (2015). An introduction to genetics for language scientists, p. 67.

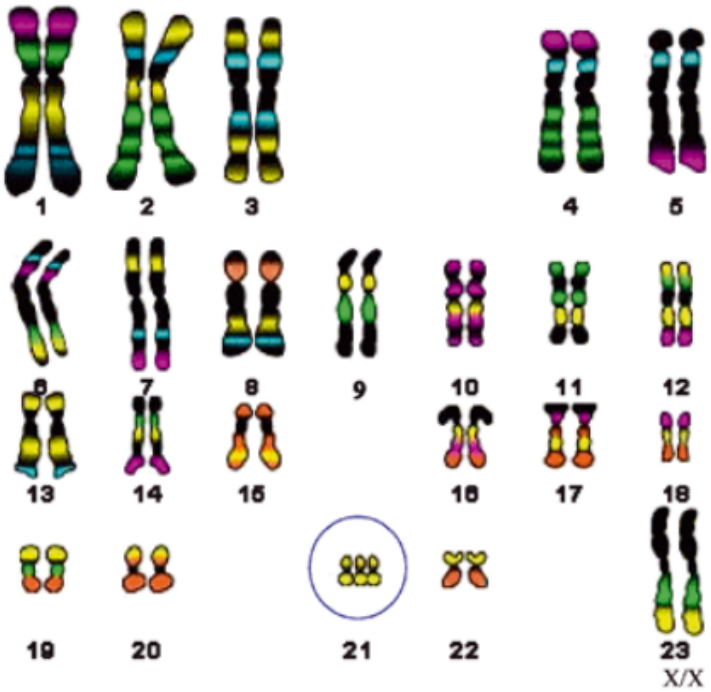


Figure: Chromosomes in the metaphase (part of mitosis) of a person with Trisomy 21 (aka Down Syndrome).
Storch et al. (2013), p. 257.

Exercise

Take the visualization of different versions of the FOXP2 gene below (Krause et al., 2007). For these sequences of nucleotides built a *distance matrix* of pairwise differences in nucleotides. Based on this distance matrix, built a *hierarchical tree* according to the so-called *complete-linkage clustering*. In other words, always merge the species together with the smallest nucleotide difference. When you compare an already existing cluster of species with another species (or cluster of species) then the *maximum distance* between any of the species is considered for clustering. You don't need to give branch lengths.

`https://en.wikipedia.org/wiki/Complete-linkage_clustering`

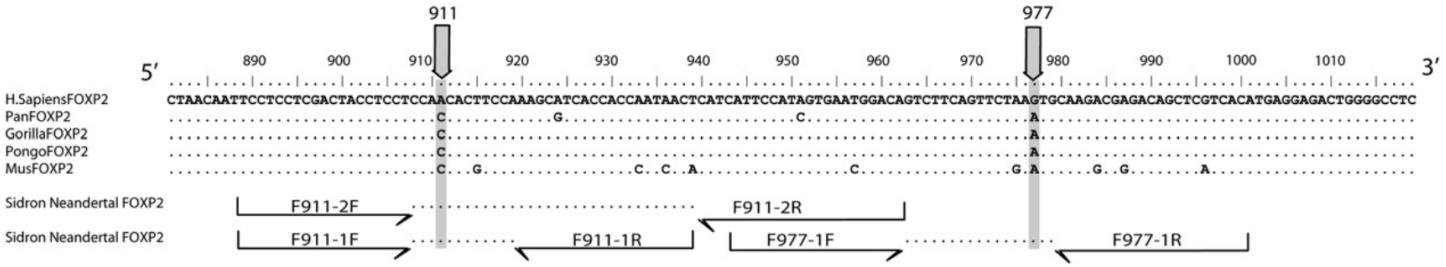


Figure 1. Sequence Alignment of Nucleotide Positions 880–1020 from the *FOXP2* Gene

The two nonsynonymous nucleotide substitutions on the human lineage are indicated by arrows. Identical positions in the alignment are given as dots. The three primer pairs used to retrieve the two substitutions from the El Sidrón Neandertals are indicated by arrows.

Solutions: Matrix

	Hom	Pan	Gor	Pon	Mus	Nea
Hom	0	4	2	2	11	0
Pan		0	2	2	9	4
Gor			0	0	9	2
Pon				0	9	2
Mus					0	11
Nea						0

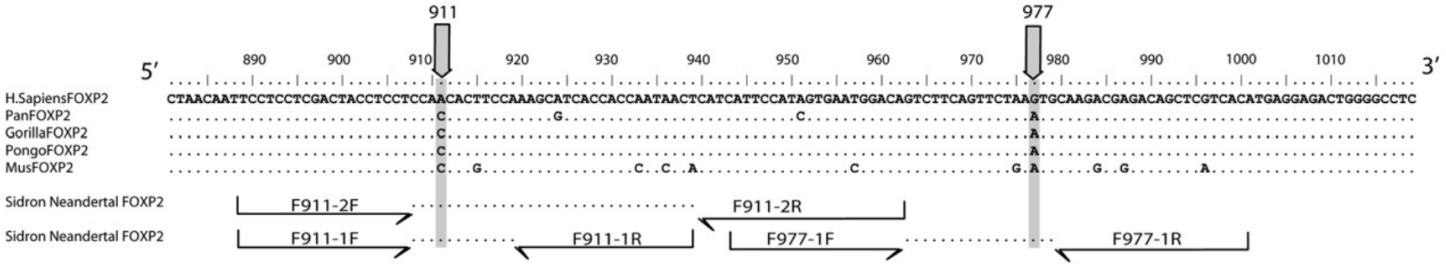


Figure 1. Sequence Alignment of Nucleotide Positions 880–1020 from the *FOXP2* Gene

The two nonsynonymous nucleotide substitutions on the human lineage are indicated by arrows. Identical positions in the alignment are given as dots. The three primer pairs used to retrieve the two substitutions from the El Sidrón Neandertals are indicated by arrows.

Solutions: Hierarchical Tree

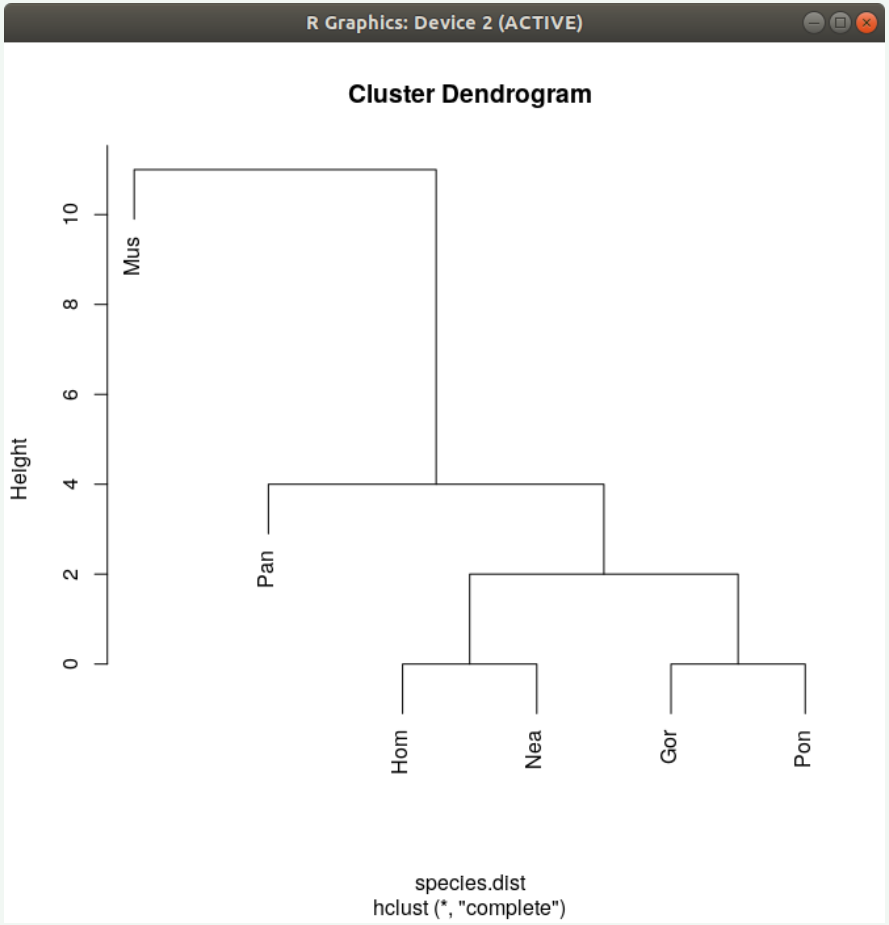



photo by Abdullah Mohiuddin
CC BY-NC-SA Hajime Yamauchi
GGCCTGACACCTTCGCACC

Summary

Summary

- Modern humans have **c. 20k genes**, i.e. regions on the DNA strands which code for *proteins* and hence shape the phenotype.
- These are packaged into **46 (23 pairs) of chromosomes**.
- Genetic diversity is ensured by **random segregation** (splitting of chromatid sisters), and **recombination** of homologous chromosomes in gametogenesis.
- Different **types of mutations** (pointwise, frameshift, chromosomal) will yield new phenotypes under selection (or neutral).


photo by Abdullah Mohiuddin
CC BY-NC-SA Hajime Yamauchi
GGCCTGACACCTTCGCACC

References

References

Dediu, D. (2015). *An introduction to genetics for language scientists. Current concepts, methods, and findings*. Cambridge: Cambridge University Press.

Krause et al. (2007). The Derived FOXP2 Variant of Modern Humans Was Shared with Neandertals. *Current Biology*.

Storch et al. (2013). *Evolutionsbiologie*. Springer Verlag.

Thank You.

Contact:

Faculty of Philosophy
General Linguistics
Dr. Christian Bentz
SFS Keplerstraße 2, Room 168
chris@christianbentz.de
Office hours:
During term: Wednesdays 10-11am
Out of term: arrange via e-mail