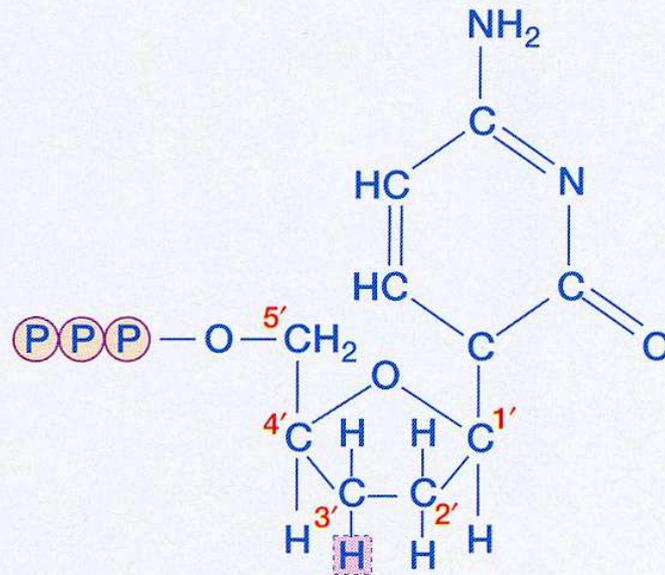
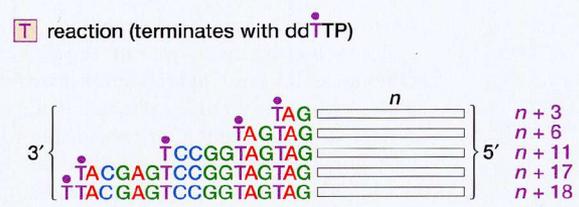
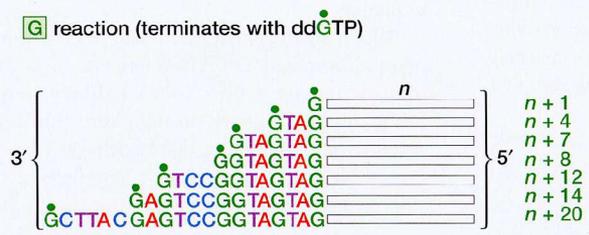
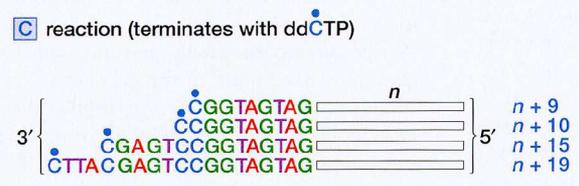
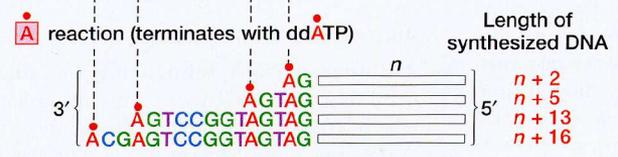
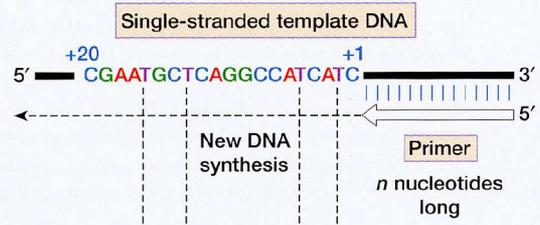


Sequenzieren

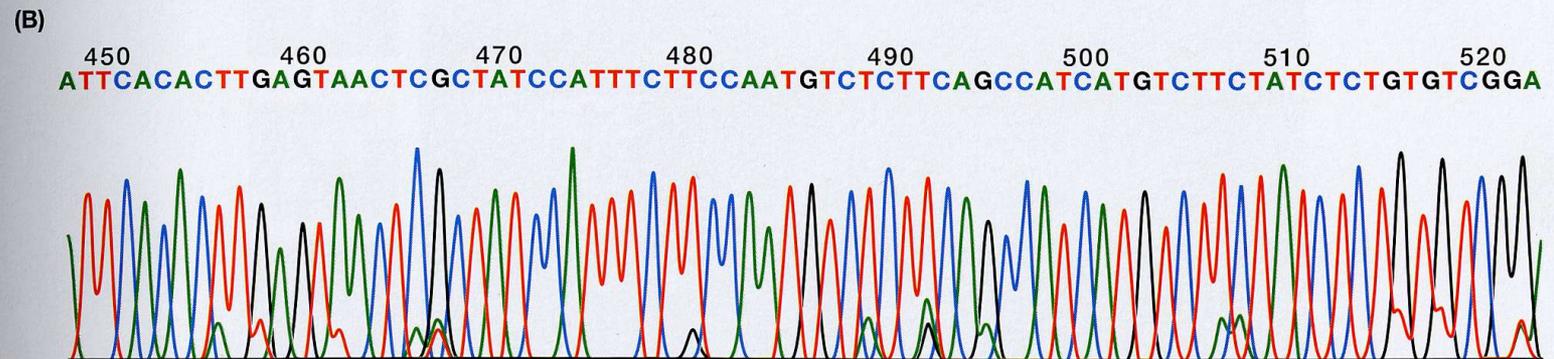
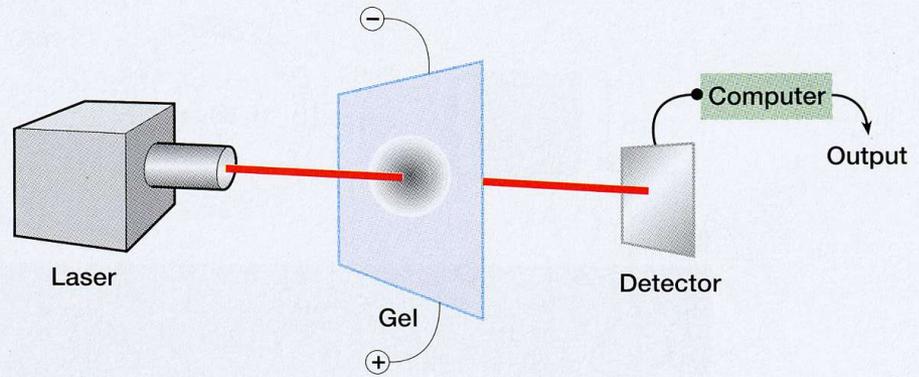
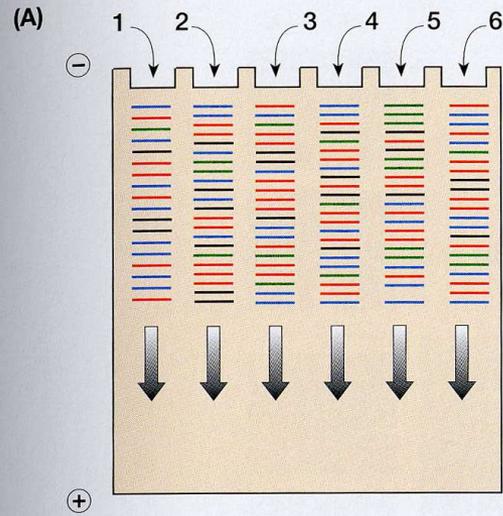


Key:
Ⓟ Phosphate group



Fractionation according to size

- $n + 20 = G$
 $n + 19 = C$
 $n + 18 = T$
 $n + 17 = T$
 $n + 16 = A$
 $n + 15 = C$
 $n + 14 = G$
 $n + 13 = A$
 $n + 12 = G$
 $n + 11 = T$
 $n + 10 = C$
 $n + 9 = C$
 $n + 8 = G$
 $n + 7 = G$
 $n + 6 = T$
 $n + 5 = A$
 $n + 4 = G$
 $n + 3 = T$
 $n + 2 = A$
 $n + 1 = G$
-



Human Genome Project

- Entwicklung geeigneter Technologien
- Fünf Modellorganismen
- ELSI: ethische, legale und soziale Bedeutung

Organisation von HGP

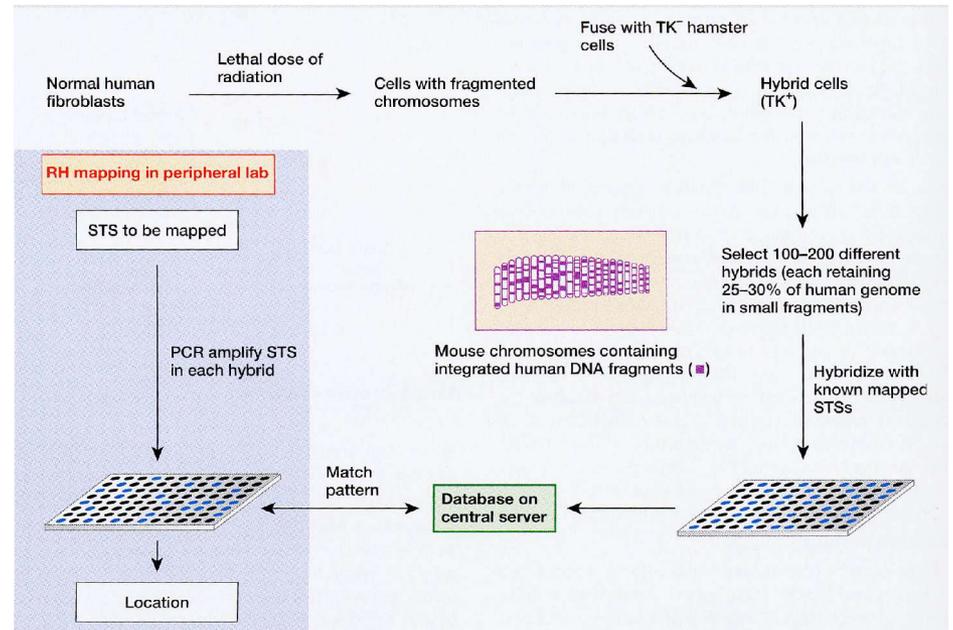
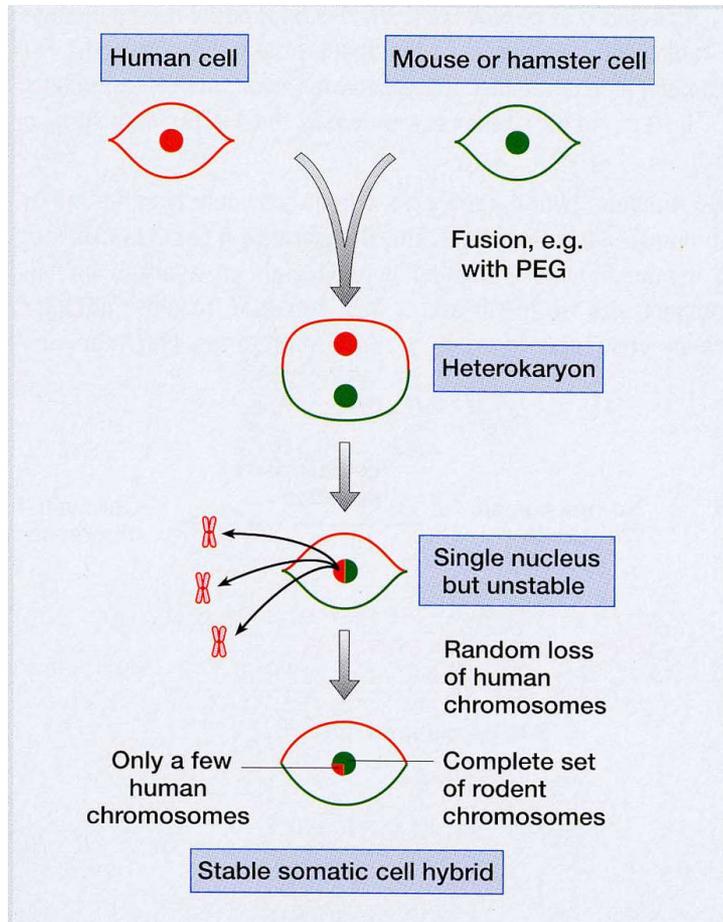
- Schaffung von genetischen und physikalischen Landkarten
- Verbesserung der Techniken
- Durchführung in großen Zentren
- Zentrale Datenbanken

Schaffung genetischer und physikalischer Landkarten

- RFLP's und Mikrosatelliten Marker
- Entwicklung von SNP Karten
- Bibliotheken genomischer DNA
- Clone contigs

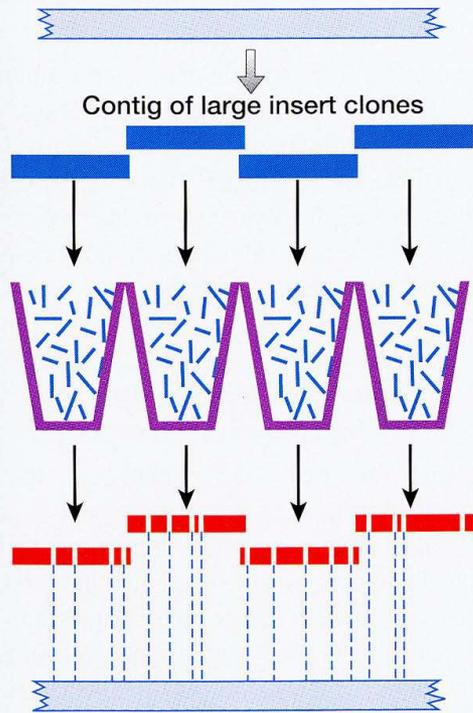
Sequence Tagged Sites

- Wichtige Kartierungswerkzeuge
- 15000 STS bildeten Grundgerüst
- Nichtpolymorphe und polymorphe Marker



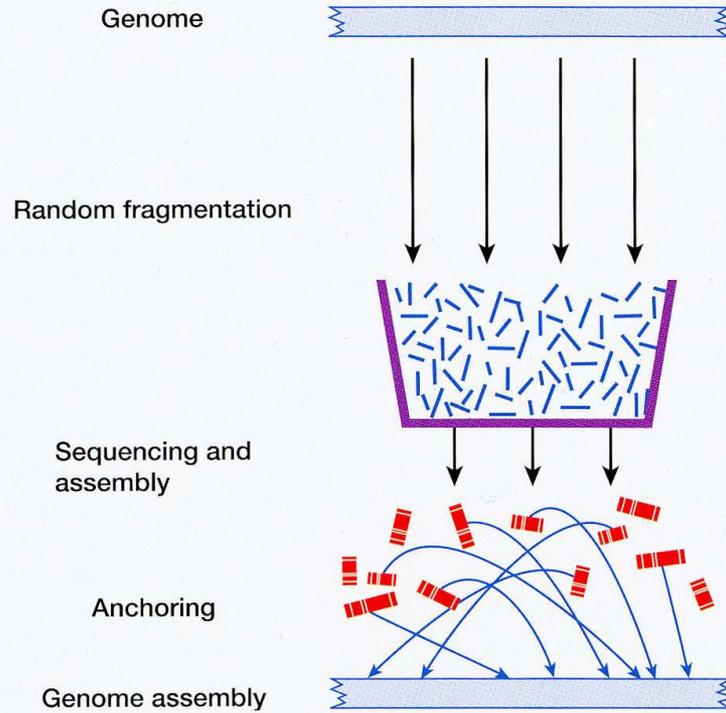
(A)

Hierarchical shotgun



(B)

Whole-genome shotgun



Genome

Random fragmentation

Sequencing and assembly

Anchoring

Genome assembly

Wieviele Gene hat der Mensch?

- Frühe Annahme: 60-100000
- Derzeitige Annahme: 30-35000
- Lediglich 50% mehr als C.Elegans
- Probleme bei der in-silico Voraussage

Gene Ontology (GO)

- Entwicklung eines systematischen Vokabulars
- Drei separate Ontologien:
biologische Prozesse
zelluläre Komponente
molekulare Funktion

Human Diversity Project

- Verschiedene ethnische Gruppen
- Bedeutsam für: Anthropologie, Forensik und vor allem Medizin
- Problembehaftet: Eugenik etc.

Genomprojekte für Modellorganismen

- Prokaryonten:
 - Haemophilus influenza
 - Mycoplasma genitalium
 - E.coli

Einzeiler

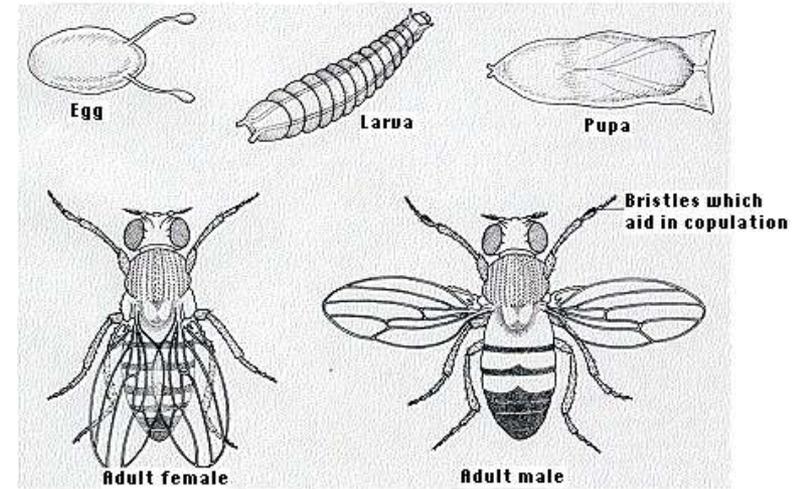
- *Saccharomyces cerevisiae*
- *Schizosaccharomyces pombe*
- *Plasmodium falciparum*

Caenorhabditis elegans



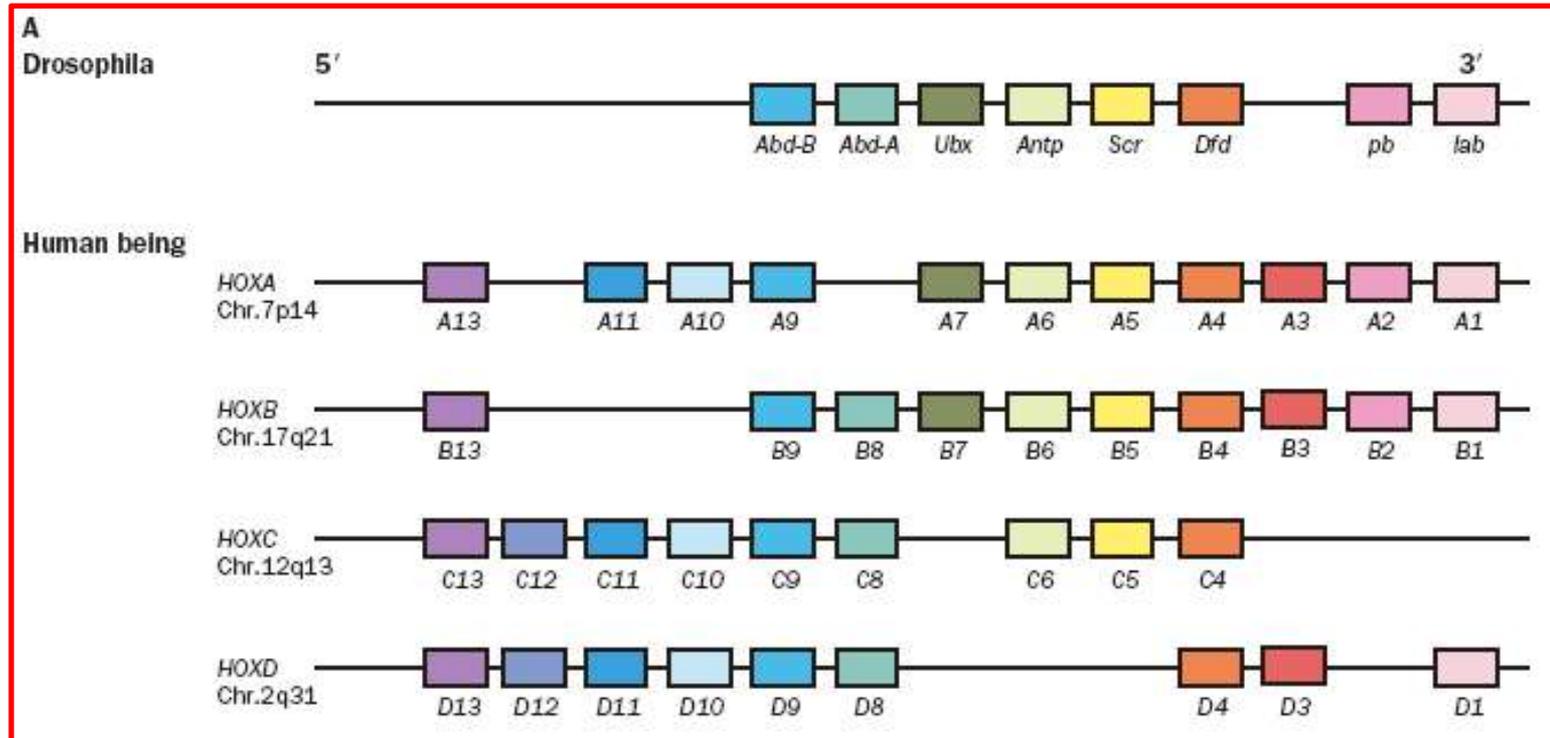
- Bedeutendes Modell für Entwicklung
- 959 oder 1031 somatische Zellen
- 100Mb Genom
- Ca. 19000 Polypeptid codierende und 1000 nicht translatierte Gene

Drosophila

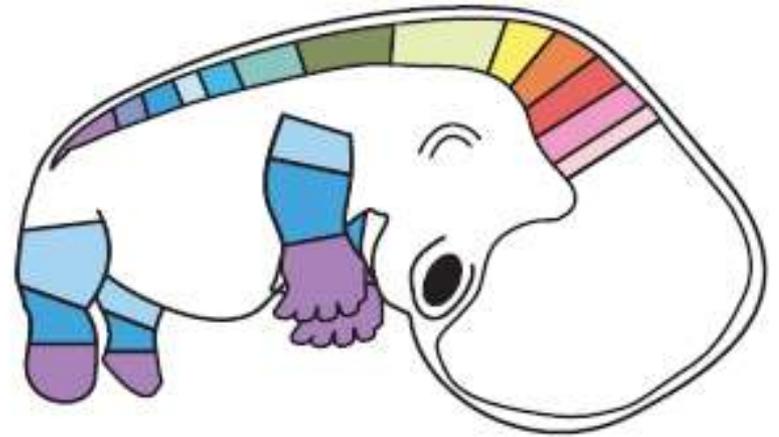
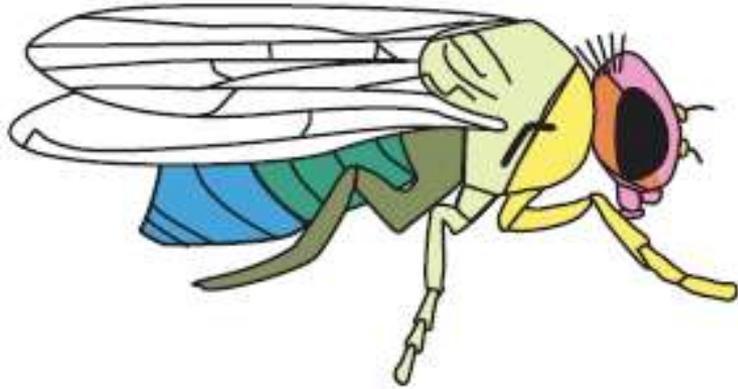


- Fruchtfliege
- Polytäne Riesenchromosomen
- Genom hat 165Mb
- 13601 Gene
- Ähnlichkeiten der Gene zwischen Drosophila und Mensch (Bsp. Hox Cluster)

Hox Genes:



B

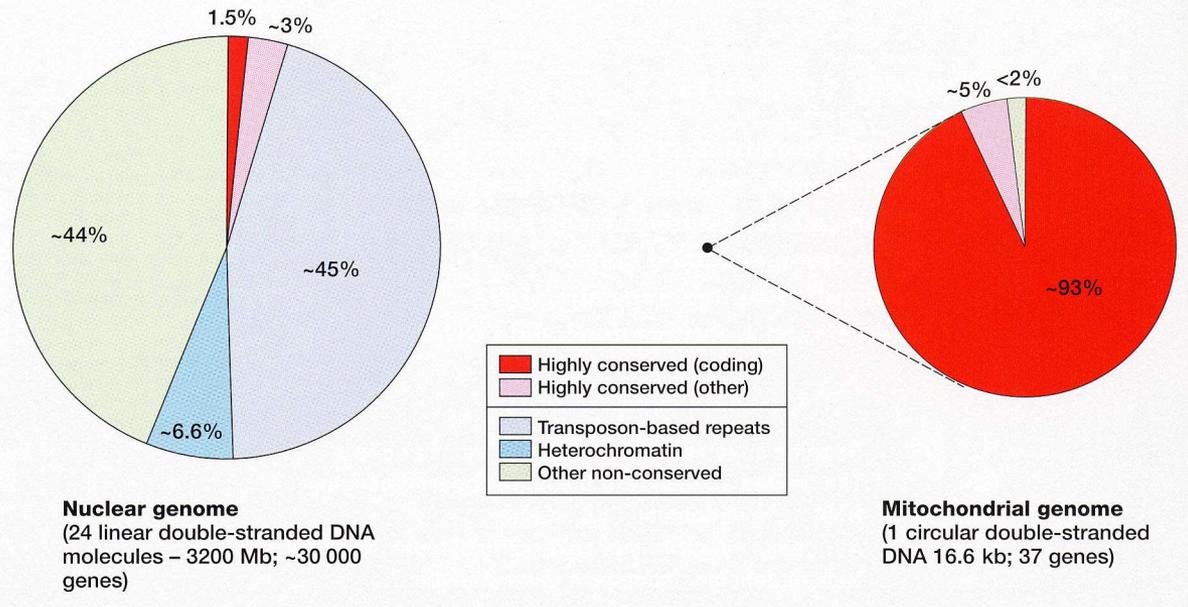


Andere Genomprojekte

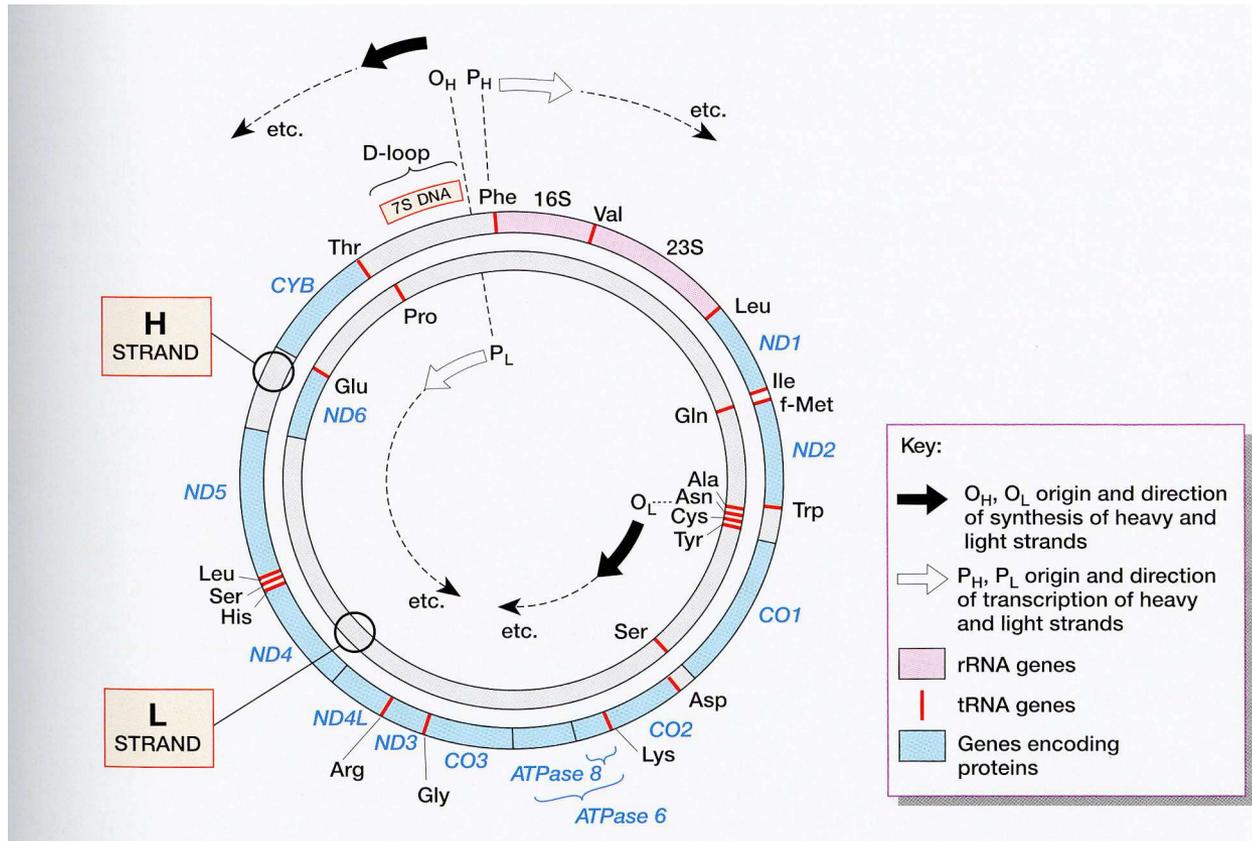
- Huhn
- *Xenopus laevis*
- Ratte
- Maus
- Etc.

Organisation des humanen Genoms

- Nukleares Genom
- Mitochondriales Genom



Mitochondriale Genom



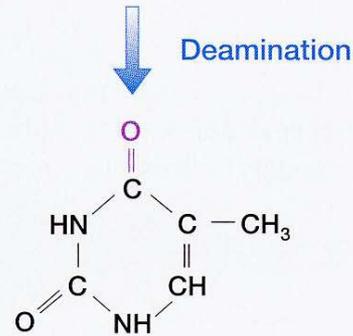
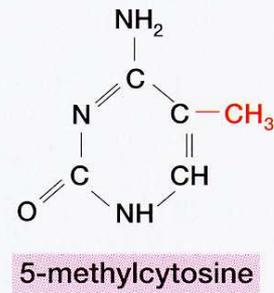
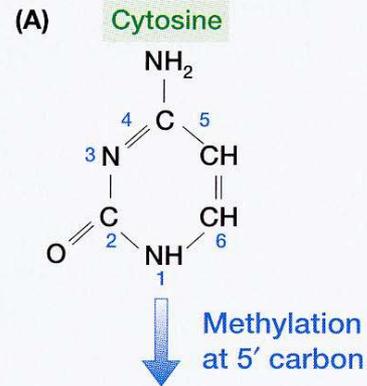
Mitochondriale Genom

- Ca. 93% ist coding sequence
- keine Introns
- Displacement (D) loop region
 - Wichtig für Replikation und Transkription

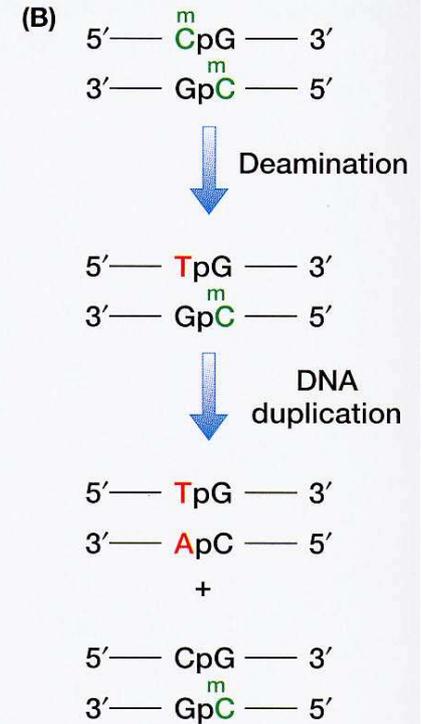
Nukleare Genom

- 24 lineare doppelsträngige DNA Moleküle
- HGP beschränkte sich auf Euchromatin
- 41%GC Gehalt, kann stark schwanken
- Korreliert mit Giemsa Bandenmuster
- unverhältnismäßige Verteilung des CpG Gehalts

CpG Islands



(forms mismatch with G;
inefficiently recognized by
DNA repair system)



Genome size	~ 3200 Mb
Nuclear genome	~ 3200 Mb
Mitochondrial genome	37 kb
Euchromatic component	~ 2900–3000 Mb
Constitutive heterochromatin	> 200 Mb (<i>Table 9.2; Figure 2.15</i>)
Highly conserved fraction	> 100 Mb (>3%)
Coding DNA	~ 50 Mb (~1.5%)
Other (regulatory etc.)	~ 100 Mb (3%)
Segmentally duplicated DNA	>150 Mb (>5%)
Noncoding repetitive DNA	> 50% of genome
Transposon-based repeats	~ 1400 Mb (~ 43%; see <i>Table 9.15</i>)
Gene number	~ 30 000–35 000
Nuclear genome	~ 30 000–35 000 (<i>Section 9.1.3</i>).
Mitochondrial genome	37 (<i>Section 9.1.2</i>).
Per chromosome	Average of ~ 1400; but depends on chromosome length and type (see <i>Figure 8.4</i>); ~ 60 per band in a 550-band chromosome preparation
Polypeptide-encoding genes	~ 30 000 but considerable uncertainty
RNA genes	~ 3000, but some uncertainty (see <i>Figure 9.4</i>)
Pseudogenes	~20 000
Gene density	~ 1/100 kb in nuclear genome; 1/0.45 kb in mitochondrial genome
Gene size (genomic extent)	Average = 27 kb, but enormous variation (see <i>Figure 9.7</i>).
Intergenic distance	Average = ca. 75 kb in nuclear genome.
CpG island number	~ 30 000 (in genome sequence filtered to remove noncoding repeats)
Exon number	Average = 9. Generally correlated with gene length, but wide variation.
Largest number	363 (in the titin gene)
Smallest number	1 (that is, no introns – <i>Table 9.5</i>)
Exon size	Average = 122 bp for internal exons with comparatively little length variation, but 3' exons can be considerably longer (Zhang, 1998).
Largest exons	Many kb long, e.g. exon 26 of the apoB gene (<i>APOB</i>) is 7.6 kb
Smallest exons	< 10 bp
Intron size	Enormous variation; strong direct correlation with gene size (see <i>Table 9.6</i>): Hundreds of kb e.g. intron 8 of the human <i>WWOX</i> gene is ~ 800 kb.
Largest introns	
Smallest introns	Tens of bp
mRNA size	Average of about 2.6 kb, but considerable variation (titin mRNA is > 115 kb long!)
5' UTR	Average of about 0.2–0.3 kb
3' UTR	Average of about 0.77 kb but likely to be an underestimate because of under- reporting of long 3' UTRs
Noncoding RNA size	Highly variable; from ~ 21–22 nucleotides (microRNA) to many kb e.g. <i>XIST</i> (17 kb)
Polypeptide size	Average of about 500–550 amino acids
Largest polypeptide	Titin: 38 138 codons in titin gene (but significant length variation)
Smallest polypeptides	Tens of amino acids e.g. various small hormones etc.

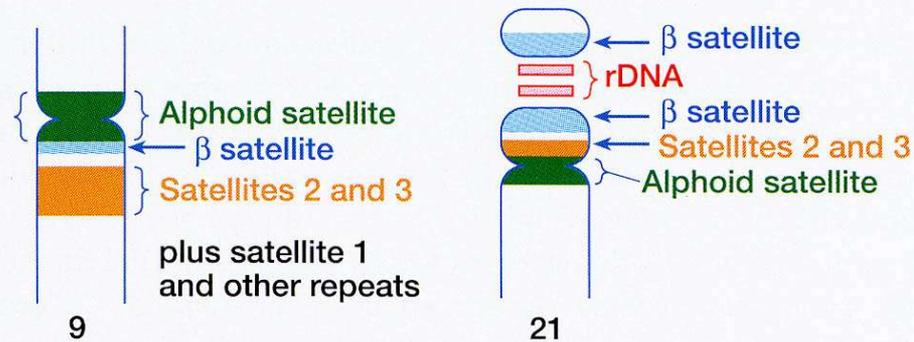


Table 9.14: Major classes of tandemly repeated human DNA

Class	Size of repeat unit (bp)	Major chromosomal location(s); transcriptional status
Satellite DNA (arrays often within 100 kb to several Mb size range)	5–171	Especially at centromeres; not transcribed
α (alphoid DNA)	171	Centromeric heterochromatin of all chromosomes
β (<i>Sau3A</i> family)	68	Notably the centromeric heterochromatin of 1, 9, 13, 14, 15, 21, 22 and Y
Satellite 1 (AT-rich)	25–48	Centromeric heterochromatin of most chromosomes and other heterochromatic regions
Satellites 2 and 3	5	Most, possibly all, chromosomes
Minisatellite DNA (arrays often within the 0.1–20 kb range)	9–64	At or close to telomeres of all chromosomes; vast majority not transcribed
Telomeric family	6	All telomeres
Hypervariable family	9–64	All chromosomes, often near telomeres
Microsatellite DNA (= simple sequence repeats, SSR) (arrays typically < 100 bp)	12	Dispersed throughout all chromosomes; some small arrays of very simple sequence