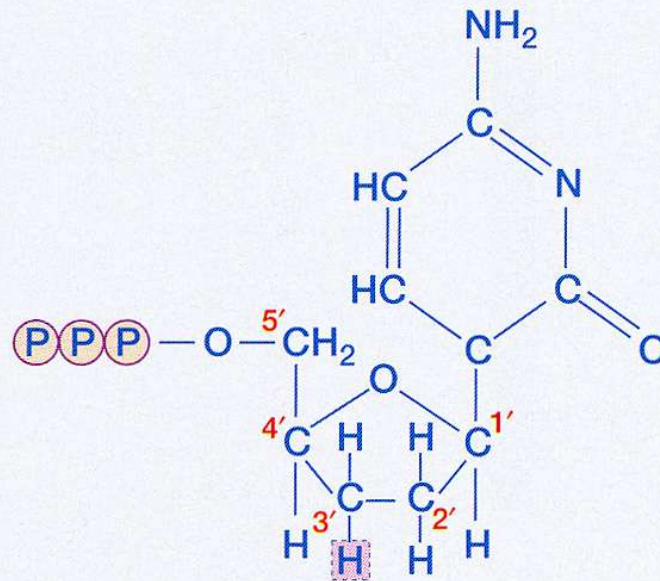

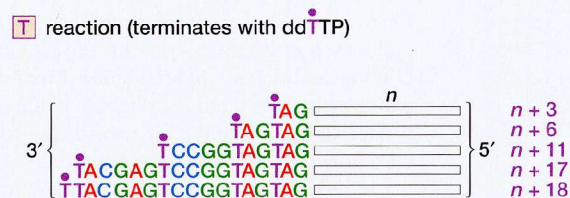
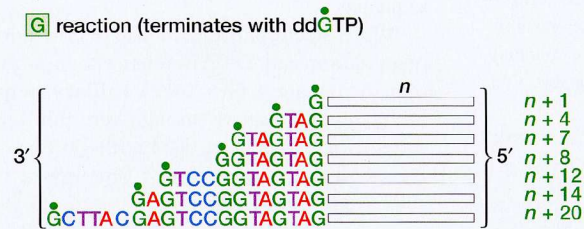
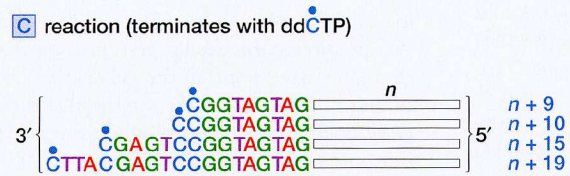
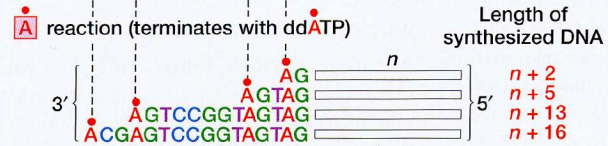
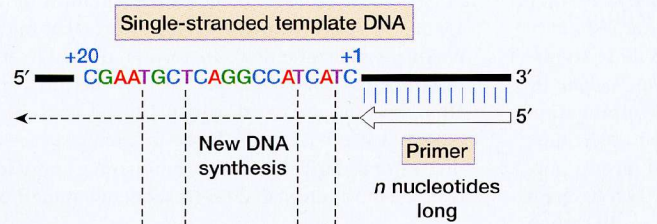


# Sequenzieren



Key:

 Phosphate group



Fractionation according to size

$$n + 20 = \text{G}$$

$$n + 19 = \text{C}$$

$$n + 18 = \text{T}$$

$$n + 17 = \text{T}$$

$$n + 16 = \text{A}$$

$$n + 15 = \text{C}$$

$$n + 14 = \text{G}$$

$$n + 13 = \text{A}$$

$$n + 12 = \text{G}$$

$$n + 11 = \text{T}$$

$$n + 10 = \text{C}$$

$$n + 9 = \text{C}$$

$$n + 8 = \text{G}$$

$$n + 7 = \text{G}$$

$$n + 6 = \text{T}$$

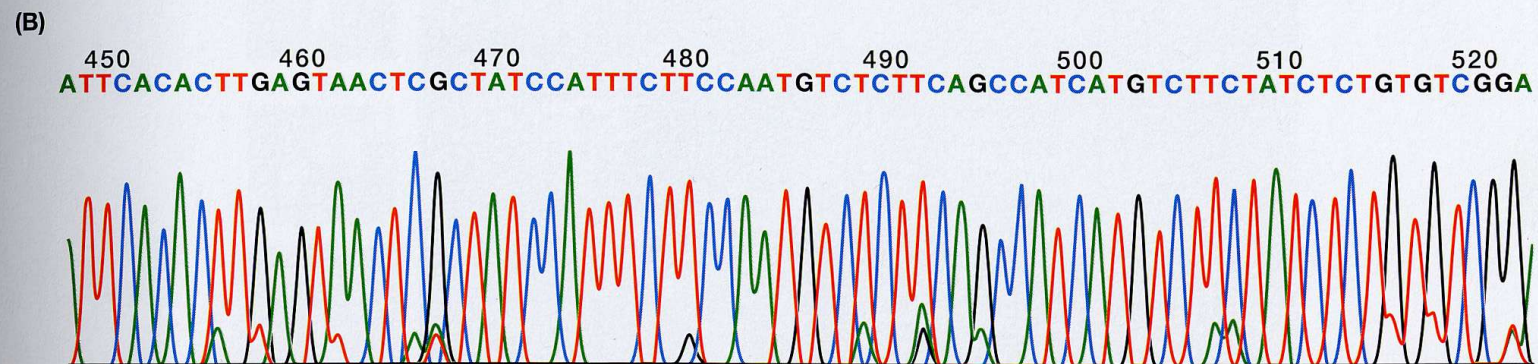
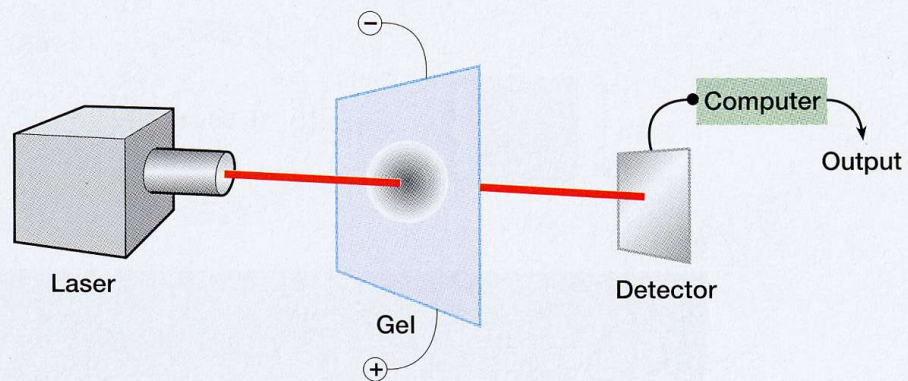
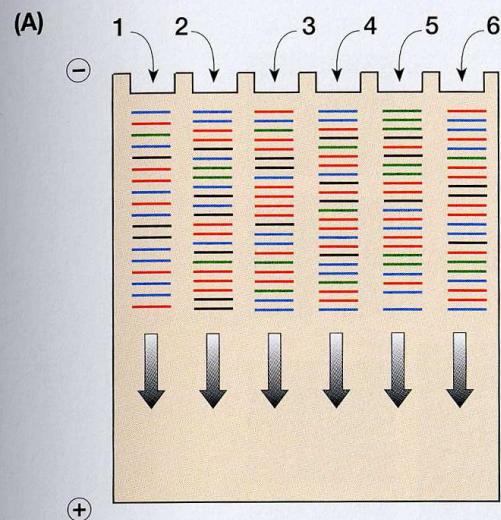
$$n + 5 = \text{A}$$

$$n + 4 = \text{G}$$

$$n + 3 = \text{T}$$

$$n + 2 = \text{A}$$

$$n + 1 = \text{G}$$





# Human Genome Project

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

# Organisation von HGP

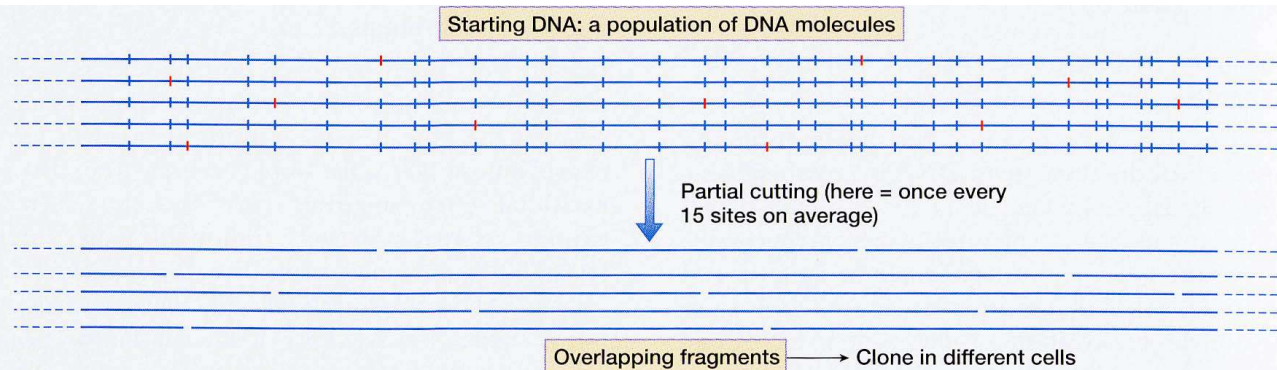
- Schaffung von genetischen und physikalischen Landkarten
- Verbesserung der Techniken
- Durchführung in großen Zentren
- Zentrale Datenbanken

# Schaffung genetischer und physikalischer Landkarten

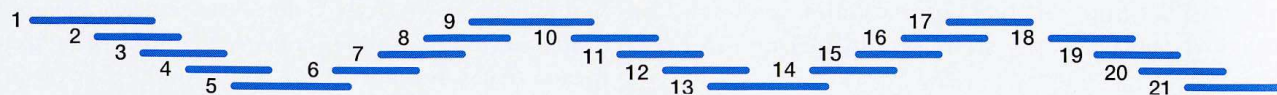
- RFLP's und Mikrosatelliten Marker
- Entwicklung von SNP Karten
- Bibliotheken genomischer DNA
- Clone contigs

# Sequence Tagged Sites

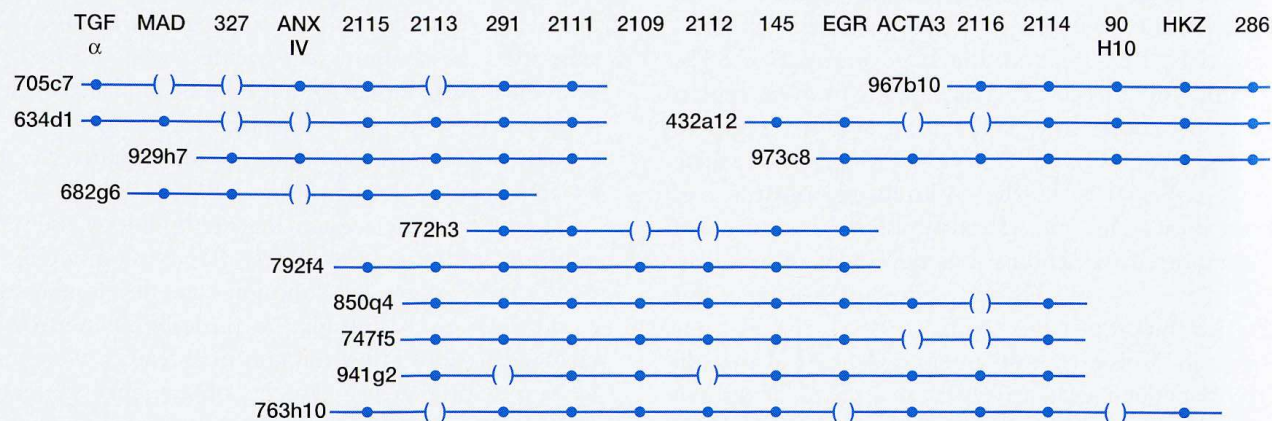
- Wichtige Kartierungswerkzeuge
- 15000 STS bildeten Grundgerüst
- Nichtpolymorphe und polymorphe Marker



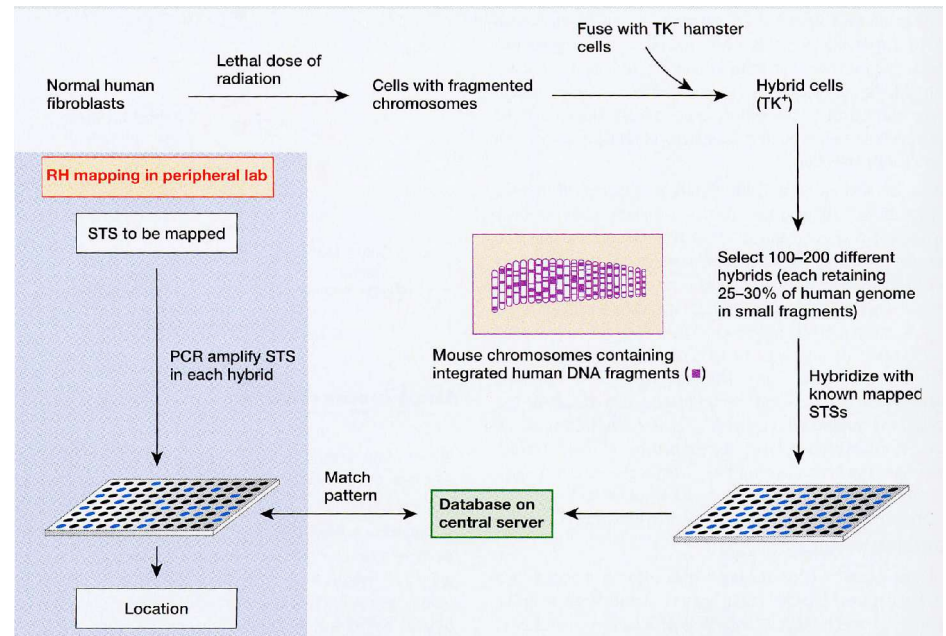
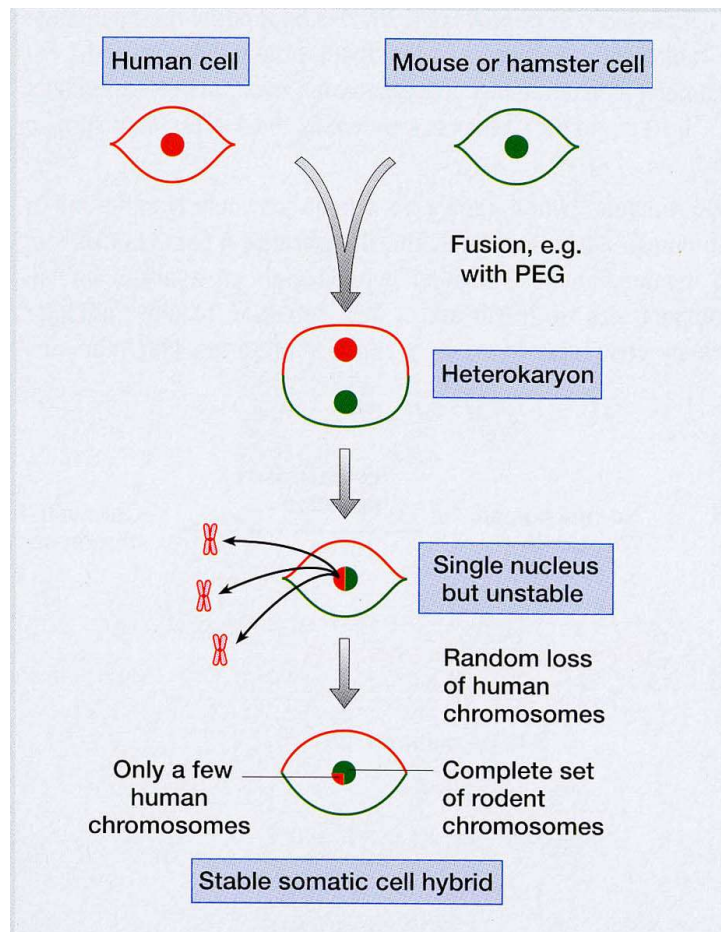
**Upper panel. Generating overlapping DNA fragments by partial restriction cutting.**



**Middle panel. A clone contig – a series of clones with partially overlapping DNA inserts.**

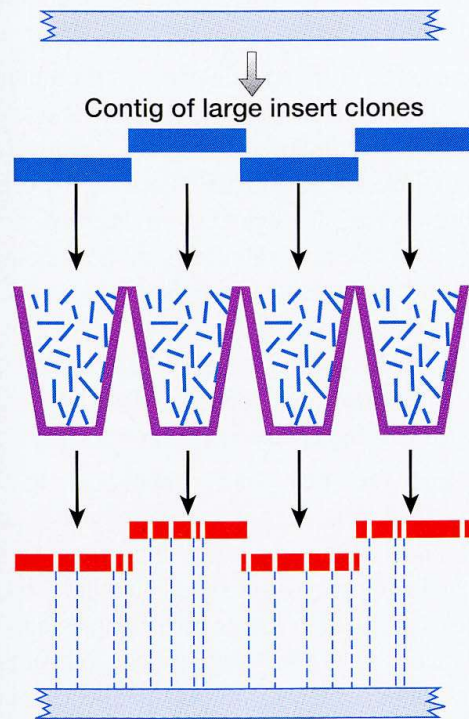






(A)

### Hierarchical shotgun



Random fragmentation

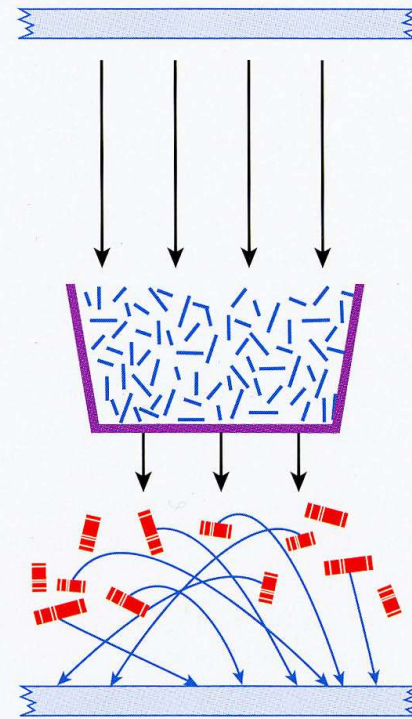
Sequencing and  
assembly

Genome assembly

(B)

### Whole-genome shotgun

Genome



Anchoring

# 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

## viomycin kinase activity

**Accession:** GO:0050394

**Aspect:** molecular\_function

**Synonyms:** None

**Definition:**

Catalysis of the reaction: ATP + viomycin = ADP + O-phosphoviomycin.

**Term Lineage** [Graphical View](#)

GO:0003673 : Gene\_Ontology ( 146200 )  
 ① GO:0003674 : molecular\_function ( 97507 )  
 ① GO:0003824 : catalytic activity ( 32256 )  
 ① GO:0016740 : transferase activity ( 9890 )  
 ① GO:0016772 : transferase activity, transferring phosphorus-containing groups ( 5162 )  
 ① GO:0016773 : phosphotransferase activity, alcohol group as acceptor ( 3682 )  
 ① **GO:0050394 : viomycin kinase activity ( 0 )**

## External References

None.

## Direct Gene Product Associations

### Filter Associations

Datasource	Evidence Code	Species
All	All Curator Approved	All
FlyBase	IMP	A. aeolicus
SGD	IGI	A. fulgidus

Submit Query



# Human Diversity Project

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

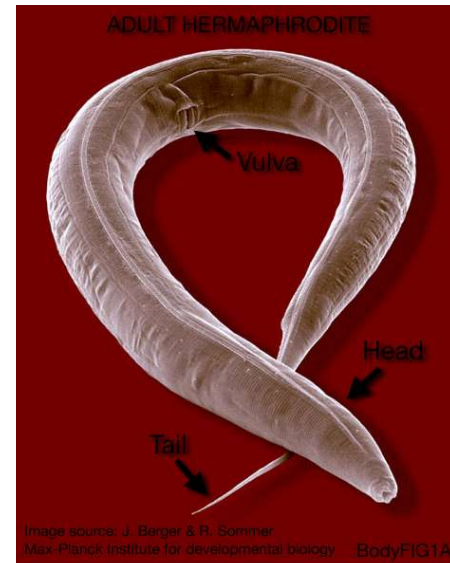
# Genomprojekte für Modellorganismen

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

# Einzeller

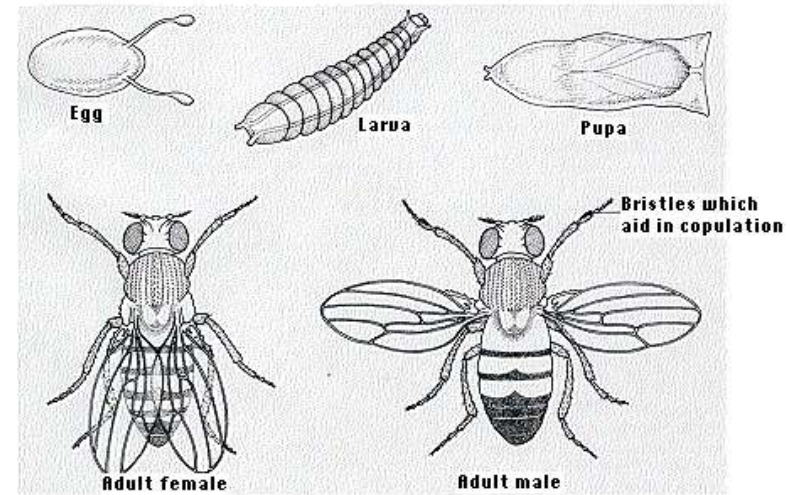
- *Saccharomyces cerevisiae*
- *Schizosaccharomyces pombe*
- *Plasmodium falciparum*

# Caenorhabditis elegans



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

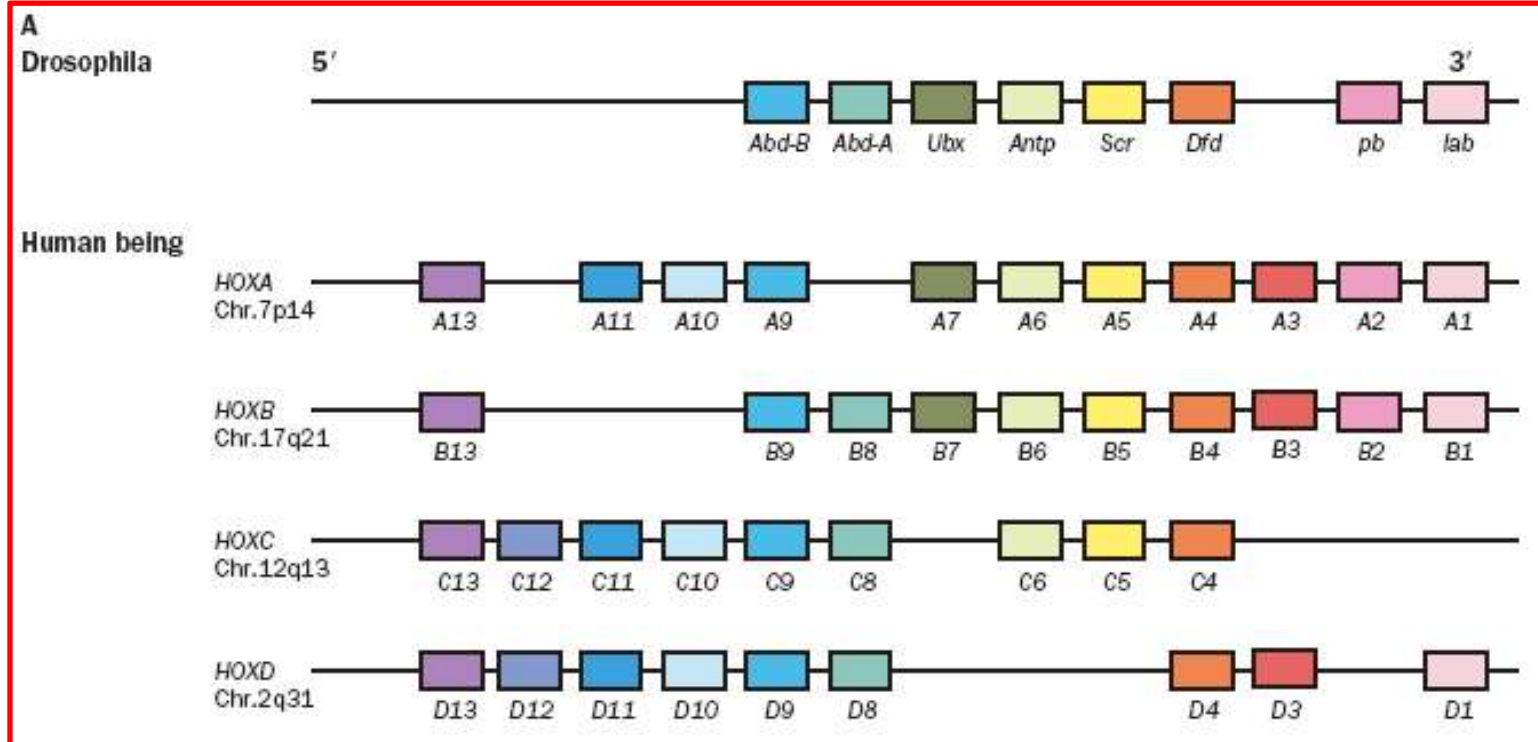
# Drosophila



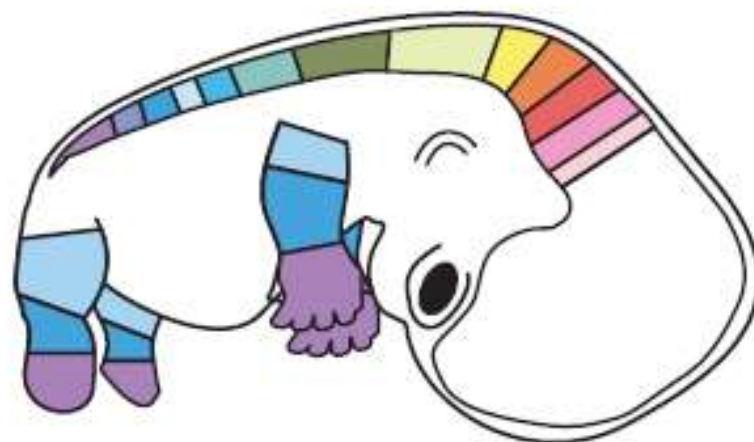
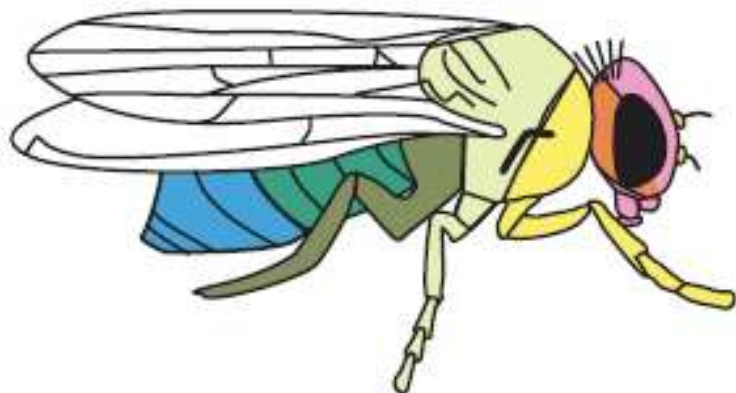
- Fruchtfliege
- Polytäne Riesenchromosomen
- Genom hat 165Mb
- 13601 Gene
- Ähnlichkeiten der Gene zwischen Drosophila und Mensch (Bsp. Hox Cluster)



# Hox Genes:



B

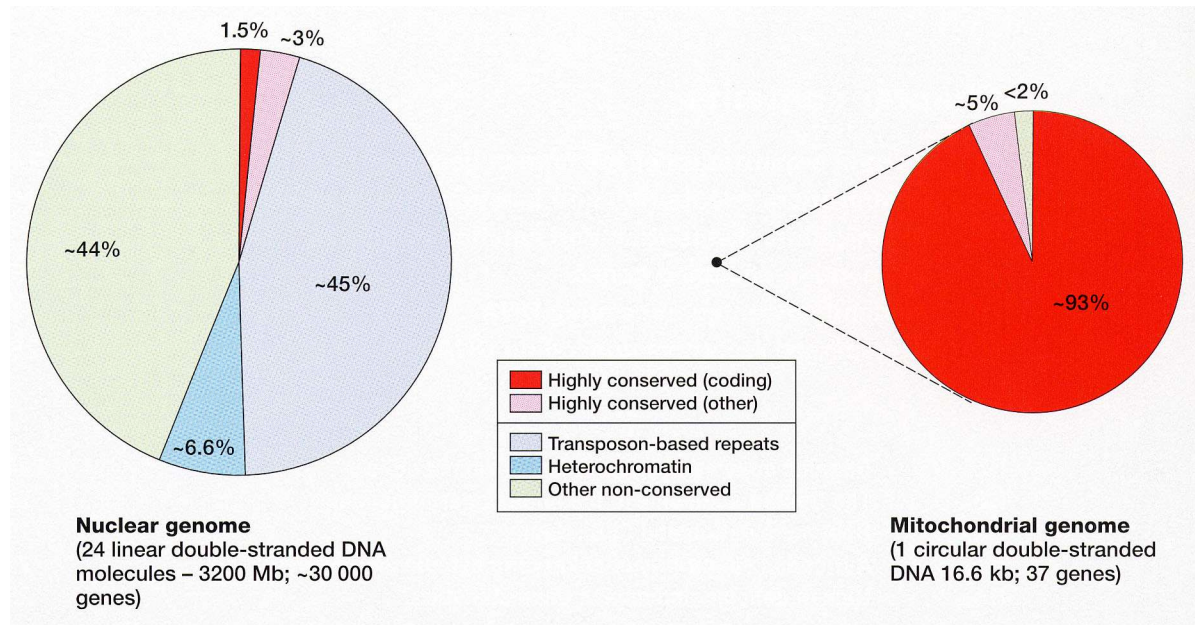


# Andere Genomprojekte

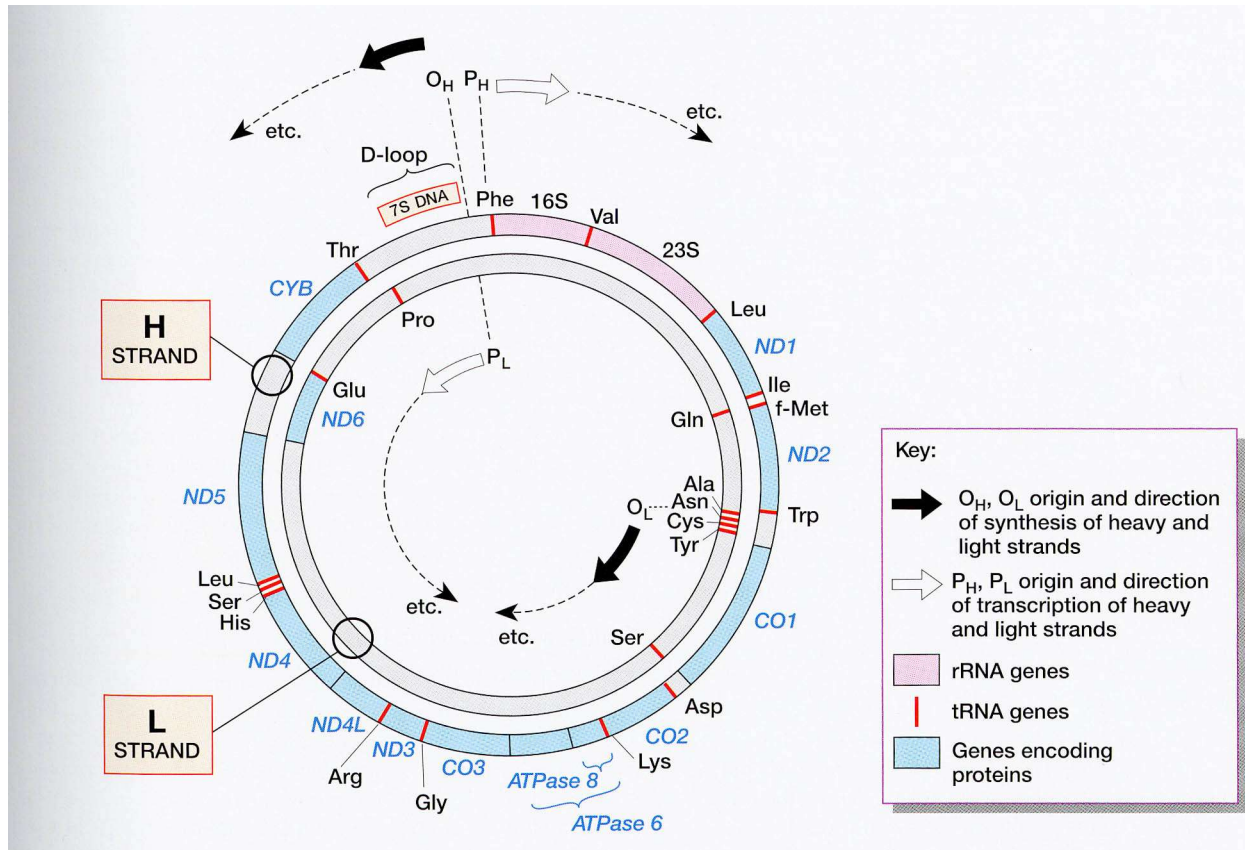
- Huhn
- *Xenopus laevis*
- Ratte
- Maus
- Etc.

# Organisation des humanen Genoms

- Nukleares Genom
- Mitochondriales Genom



# Mitochondriale Genom





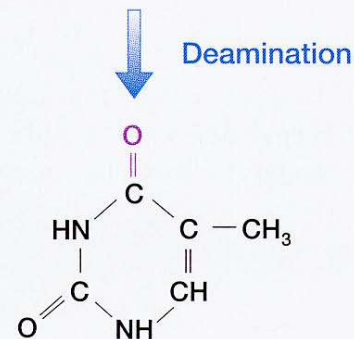
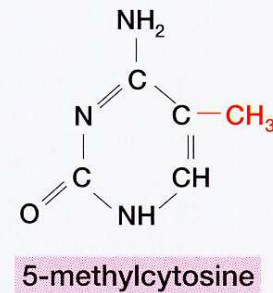
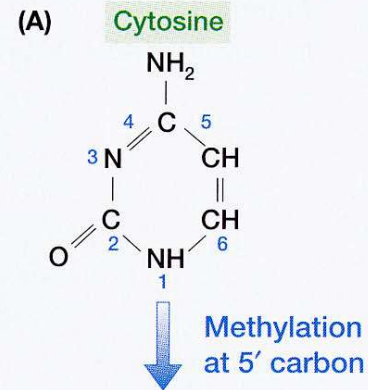
# Mitochondriale Genom

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

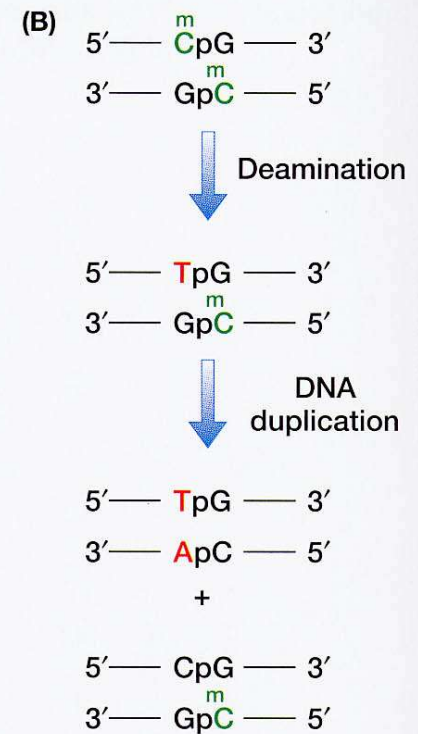
# Nukleare Genom

- 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

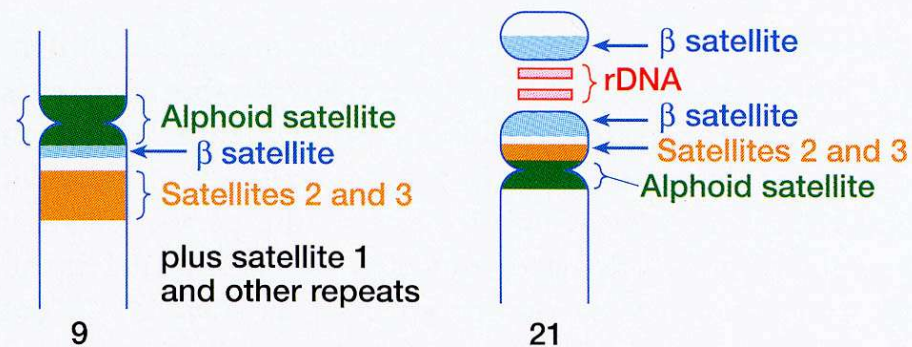


(forms mismatch with G;  
inefficiently recognized by  
DNA repair system)



<b>Genome size</b>	~ 3200 Mb
<b>Nuclear genome</b>	~ 3200 Mb
<b>Mitochondrial genome</b>	37 kb
<b>Euchromatic component</b>	~ 2900–3000 Mb
<b>Constitutive heterochromatin</b>	> 200 Mb ( <i>Table 9.2; Figure 2.15</i> )
<b>Highly conserved fraction</b>	> 100 Mb (>3%)
<b>Coding DNA</b>	~ 50 Mb (~1.5%)
<b>Other (regulatory etc.)</b>	~ 100 Mb (3%)
<b>Segmentally duplicated DNA</b>	>150 Mb (>5%)
<b>Noncoding repetitive DNA</b>	> 50% of genome
<b>Transposon-based repeats</b>	~ 1400 Mb (~ 43%; see <i>Table 9.15</i> )
<b>Gene number</b>	~ 30 000–35 000
<b>Nuclear genome</b>	~ 30 000–35 000 ( <i>Section 9.1.3</i> ).
<b>Mitochondrial genome</b>	37 ( <i>Section 9.1.2</i> ).
<b>Per chromosome</b>	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
<b>Polypeptide-encoding genes</b>	~ 30 000 but considerable uncertainty
<b>RNA genes</b>	~ 3000, but some uncertainty (see <i>Figure 9.4</i> )
<b>Pseudogenes</b>	~20 000
<b>Gene density</b>	~ 1/100 kb in nuclear genome; 1/0.45 kb in mitochondrial genome
<b>Gene size (genomic extent)</b>	Average = 27 kb, but enormous variation (see <i>Figure 9.7</i> ).
<b>Intergenic distance</b>	Average = ca. 75 kb in nuclear genome.
<b>CpG island number</b>	~ 30 000 (in genome sequence filtered to remove noncoding repeats)
<b>Exon number</b>	Average = 9. Generally correlated with gene length, but wide variation.
<b>Largest number</b>	363 (in the titin gene)
<b>Smallest number</b>	1 (that is, no introns – <i>Table 9.5</i> )
<b>Exon size</b>	Average = 122 bp for internal exons with comparatively little length variation, but 3' exons can be considerably longer (Zhang, 1998).
<b>Largest exons</b>	Many kb long, e.g. exon 26 of the apoB gene ( <i>APOB</i> ) is 7.6 kb
<b>Smallest exons</b>	< 10 bp
<b>Intron size</b>	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.
<b>Largest introns</b>	Tens of bp
<b>Smallest introns</b>	
<b>mRNA size</b>	Average of about 2.6 kb, but considerable variation (titin mRNA is > 115 kb long!)
<b>5' UTR</b>	Average of about 0.2–0.3 kb
<b>3' UTR</b>	Average of about 0.77 kb but likely to be an underestimate because of under- reporting of long 3' UTRs
<b>Noncoding RNA size</b>	Highly variable; from ~ 21–22 nucleotides (microRNA) to many kb e.g. <i>XIST</i> (17 kb)
<b>Polypeptide size</b>	Average of about 500–550 amino acids
<b>Largest polypeptide</b>	Titin: 38 138 codons in titin gene (but significant length variation)
<b>Smallest polypeptides</b>	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
<b>Satellite DNA</b> (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
<b>Minisatellite DNA</b> (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
<b>Microsatellite DNA</b> (= simple sequence repeats, <b>SSR</b> ) (arrays typically < 100 bp)	12	Dispersed throughout all chromosomes; some small arrays of very simple sequence