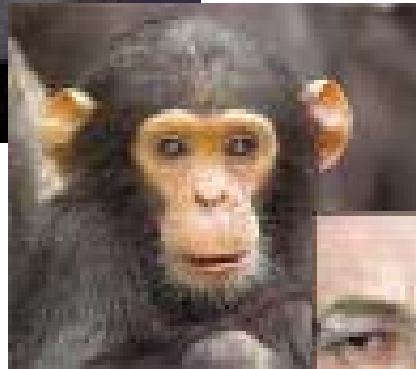
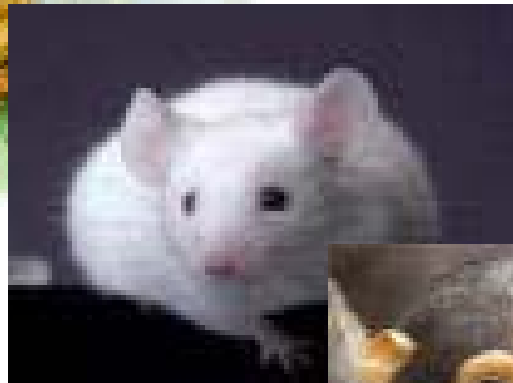
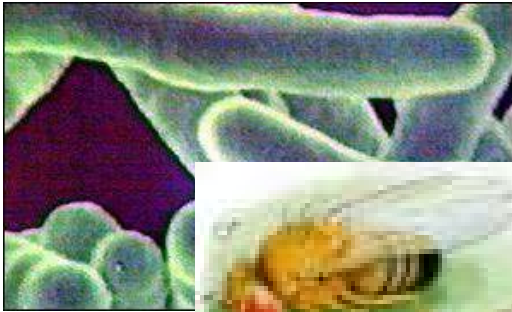


# Organisation und Evolution des Genoms



# Organisation und Evolution des Genoms

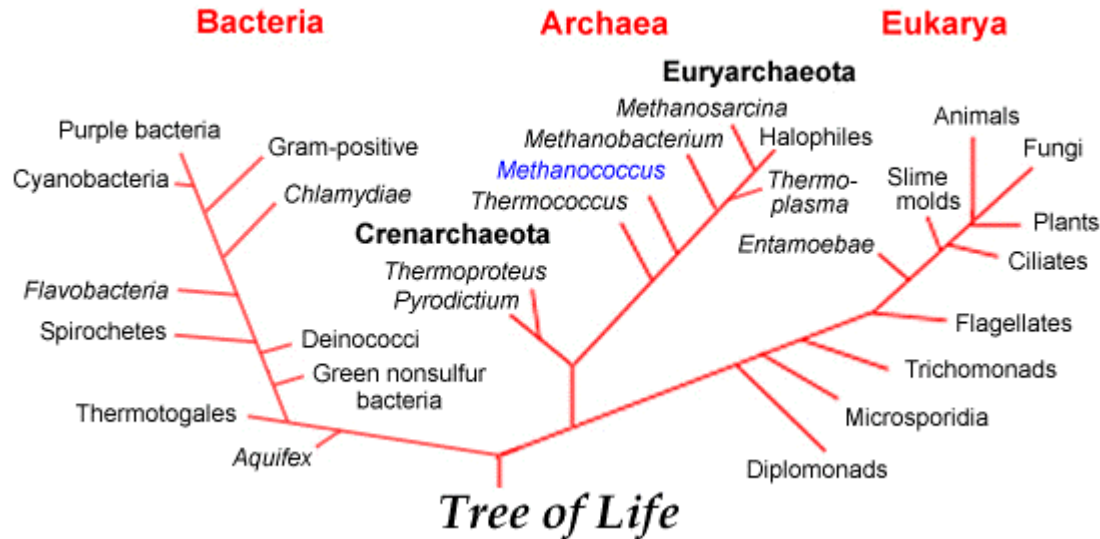
Definition Genom: vollständige DNA-Sequenz eines Organismus

# I. Einfachstes Genom: Prokaryoten

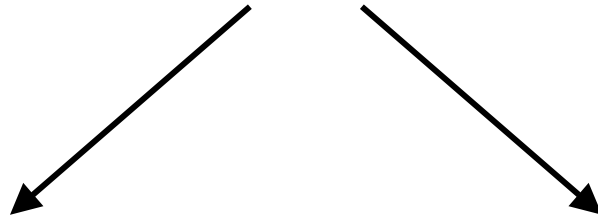
Zwei Gruppen, evolutionär unterschiedlicher Ursprung  
Genomgröße: wenige Mb, hunderte bis tausende Gene  
Einzelnes kleines zirkuläres Chromosom, DNA nicht verpackt

1. Eubakterien = Bakterien

2. Archaea (früher: Archaeobakterien)

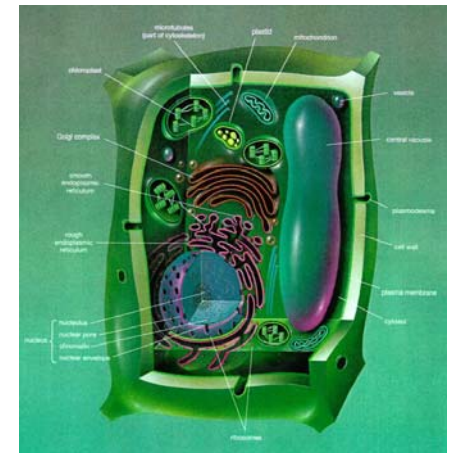
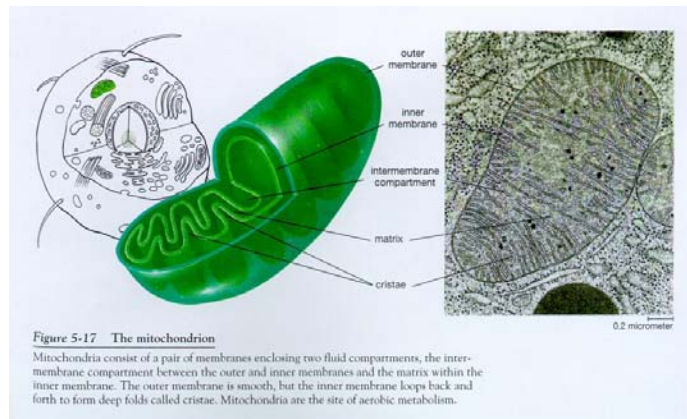


## II. Eukaryoten-Genom



A. Kern (nukleäres Genom)

B. Mitochondrien bzw. Chloroplasten



## B. Wie ist das mitochondriale Genom organisiert ?

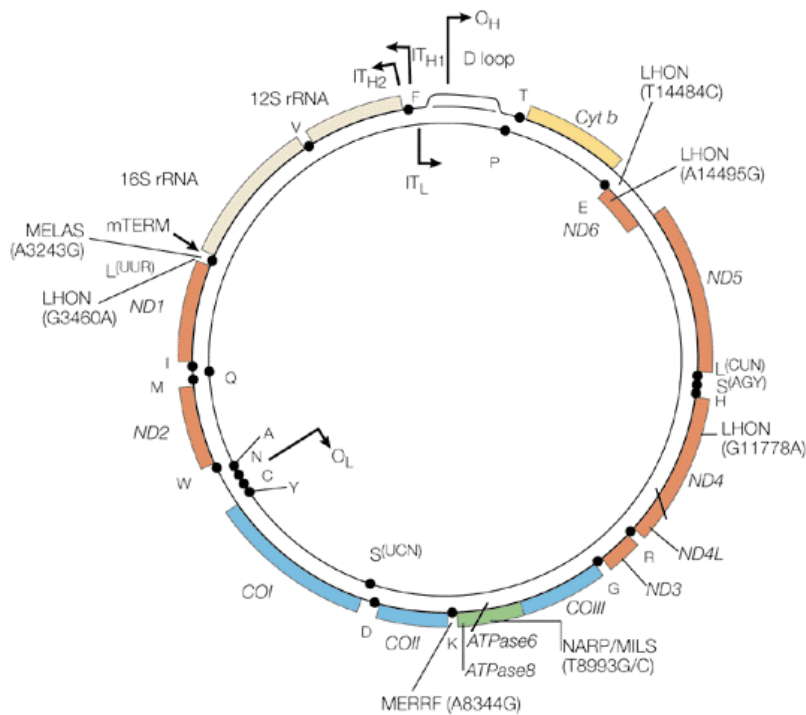
Zirkuläre ds DNA  
16569 bp

H (heavy) Strang: >> G  
L (light) Strang: >> C

37 Gene, 28 kodiert durch H Strang,  
9 durch L Strang  
24 RNAs (22 tRNA, 2 rRNA)  
13 Proteine (Atmungskette)

93% kodierende Sequenz (vs. 3% nukleär)  
Überlappende kodierende Sequenzen  
Keine Introns

Wesentliche nicht-kodierende Region: D-loop  
Enthält vorrangigen Promotor



# A. Nukleäres Genom der Eukaryoten

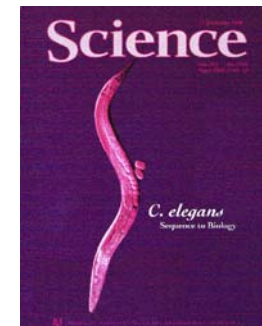
Hefe:

1 Zelle  
14 Mb  
6200 Gene  
20% kodierende DNA



C.elegans

1000 Zellen  
100 Mb  
20000 Gene  
8 % kodierende DNA



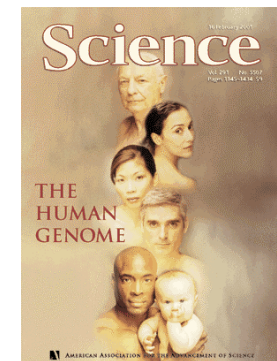
Drosophila

117 Mb  
13500 Gene  
17 % kodierende DNA



Mensch

$10^{14}$  Zellen  
3,2 Gb  
31.000  
3% kodierende DNA



A1. Wie ist das nukleäre Genom organisiert ?

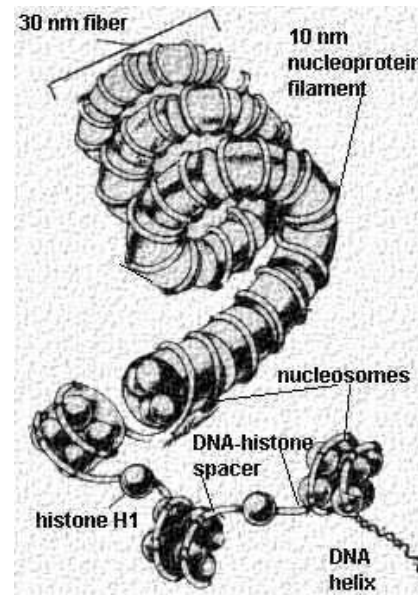
A2. Evolution des nukleären Genoms

# 1. Wie ist das nukleäre Genom organisiert ?

## 1.1 in mehrere lineare Chromosomen !

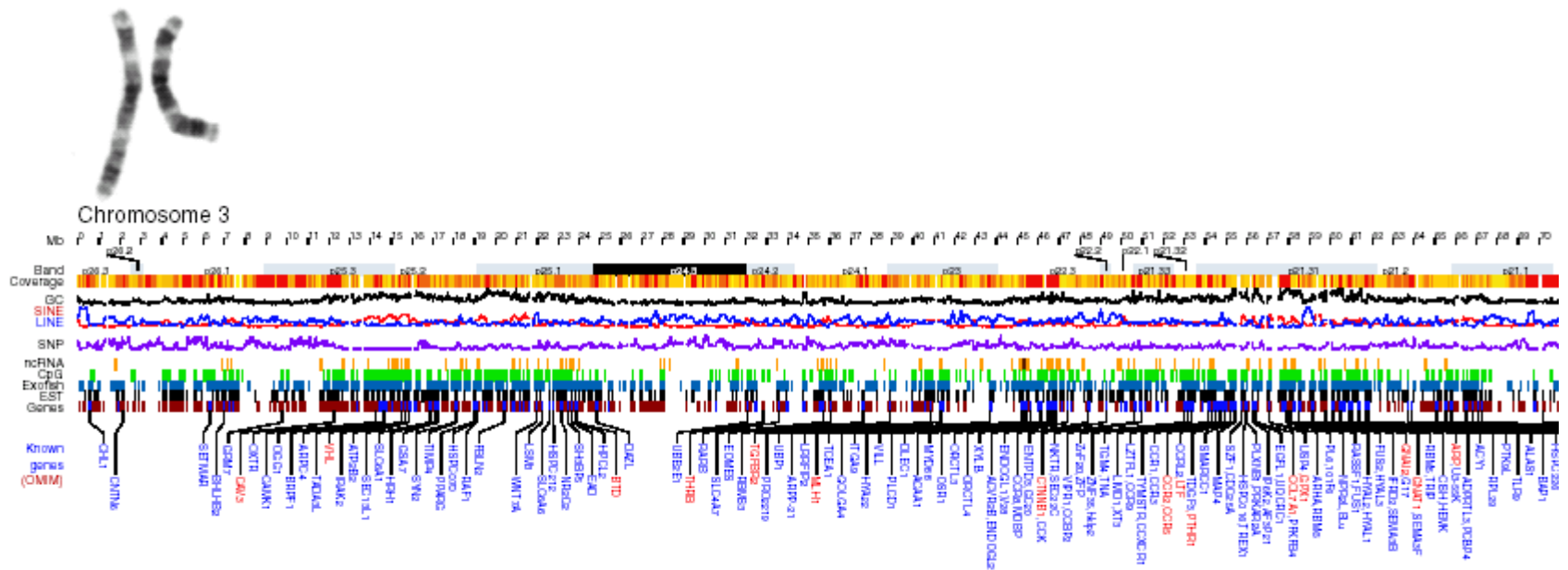
<i>S. cerevisiae</i> (Hefe)	16	14 Mb
<i>C. elegans</i> (Wurm)	11/12	100 Mb
<i>D. melanogaster</i> (Fruchtfliege)	4	170 Mb
<i>M. musculus</i> (Maus)	20	(2500 Mb Euchromatin)
<i>H. sapiens</i>	23	3200 Mb (2950 Mb Euchromatin)

## 1.2 DNA ist „verpackt“





# 1.3 Zwei Texte aus 4 Buchstaben: die DNA-Sequenzen von Mensch und Maus (Nature 15. Feb. 2001/Science Nature 5. Dez 2002)



Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

**Band**, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

Chromosomen sind in Banden organisiert !

Giemsa-Färbung: häufige Chromosomenfärbung

G-Banden:

Schwarz = Giemsa-positiv

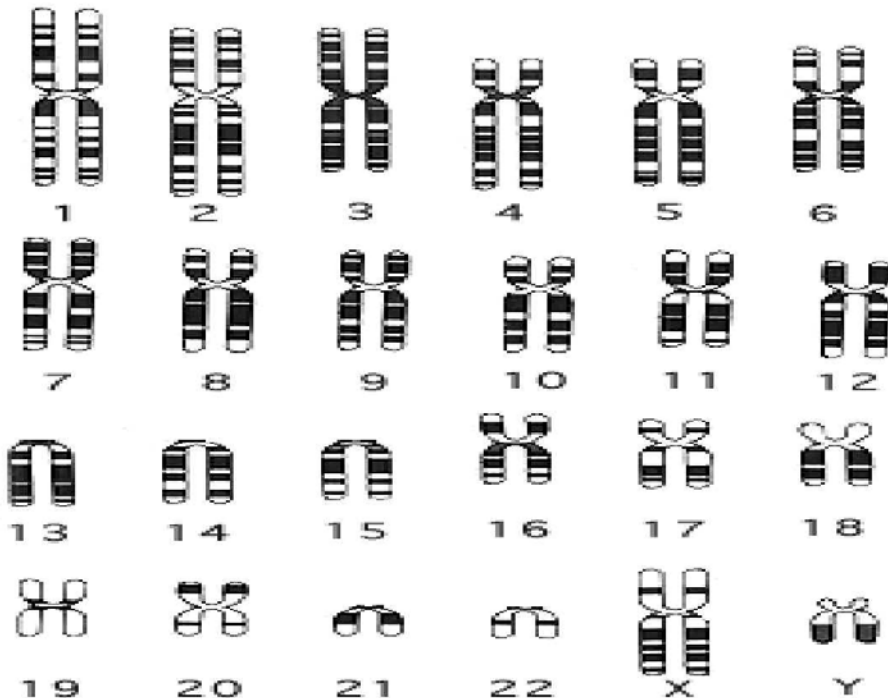
Chromatin kondensiert

Replizieren spät in der S-Phase

Weniger aktive Transkription

Weniger Gene

AT-reich



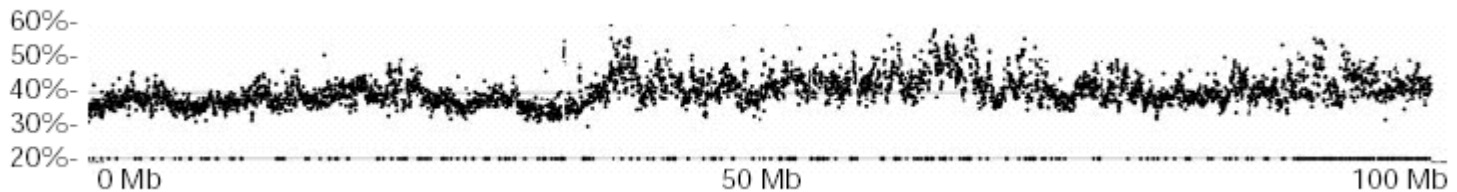
Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

Durchschnittl. 41% GC (Mensch) 42% (Maus)

Variation im GC-Gehalt über das Genom:

Mensch: deutlich, ca. 30% - 65%

Maus: weniger Variation um Durchschnitt

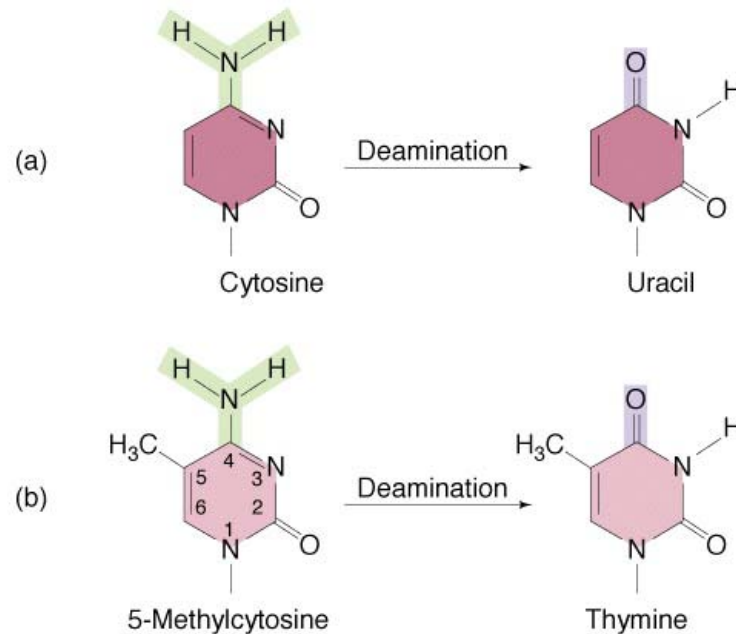


Gibt es eine Korrelation zwischen GC-Gehalt und verschiedenen biologischen Eigenschaften ?

- Giemsa-Banden
- Repetitive Sequenzen
- Gendichte

Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

Dinukleotid CpG: Deutlich unterrepräsentiert als statistisch erwartet. Warum ?

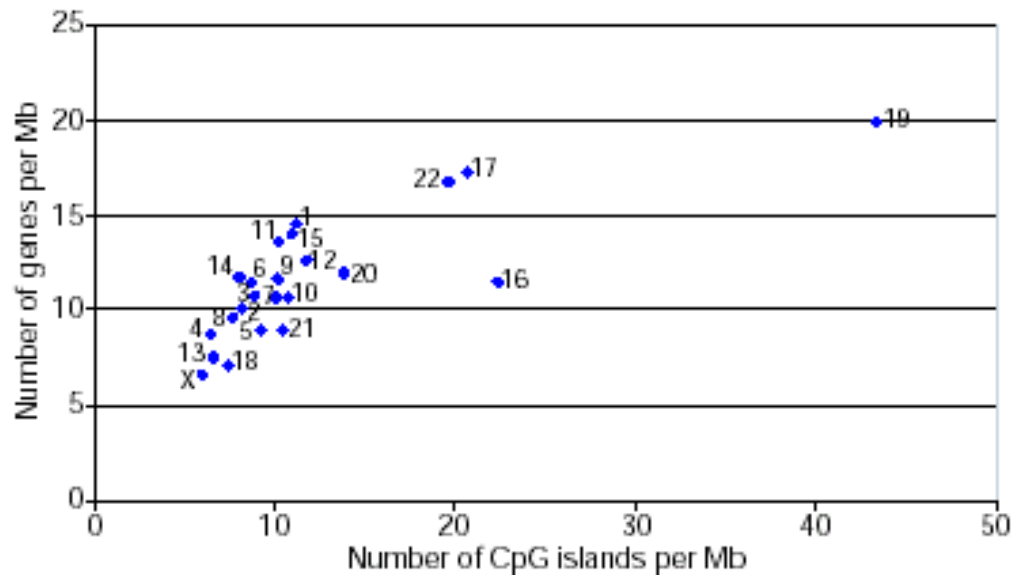


Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

„CpG islands“: Definition

Häufig im 5'-Bereich von Genen => Regulation

Mensch: 27000, Maus 15500



Band, GC, **SINE, LINE**, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

Repetitive Sequenzen –  
Genomgröße korreliert nicht mit Organismuskomplexität !

Humanes Genom 200x größer als Hefe, aber 200x kleiner als Amöbe

Nukleäres Genom

Gensequenzen

Sequenzen ausserhalb von Genen

Kodierend  
< 5% Mensch

Nicht-  
Kodierend  
Introns etc.

Einzelne / gering  
repetitive Elemente

Häufige Wiederholungen

Tandem  
Repeats

Verstreute  
(Interspersed)  
Repeats  
46% Mensch  
37,5% Maus

Band, GC, **SINE, LINE**, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

## Tandem Repeats:

Megasatelliten:

3 – 4.7 kb lange Blöcke

Satelliten:

5 – 171 bp, **Centromer !**

Minisatelliten:

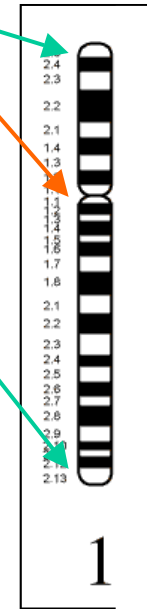
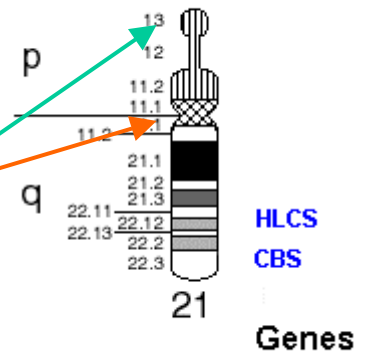
6 – 64 bp, **Telomer !**

Microsatelliten:

1 – 4 bp, alle Chromosomen

Funktion unklar

Chromosome 21

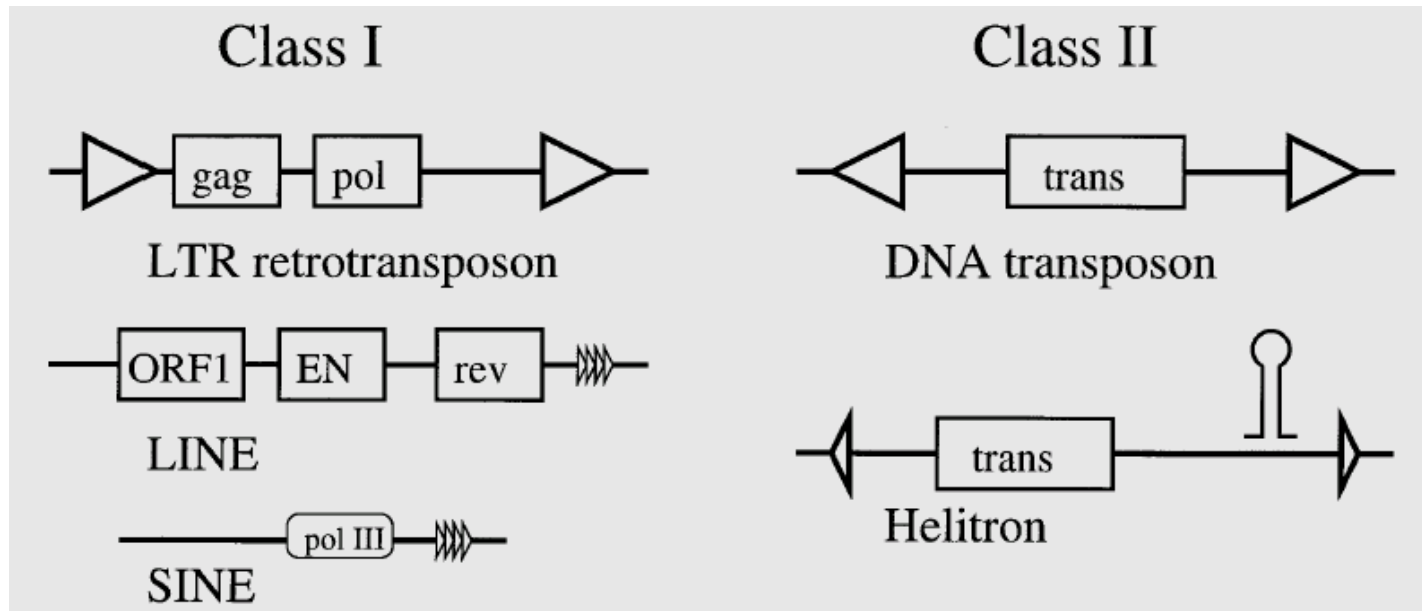


Band, GC, **SINE, LINE**, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

## Interspersed Repeats:

1. SINES : Short interspersed nuclear elements
2. LINES: Long interspersed nuclear elements
3. LTRs: Long terminal repeats
4. DNA Transposons

Reverse Transkription,  
RNA-Zwischenstufe  
(Retroposition)





Band, GC, **SINE, LINE**, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

## SINEs: Short interspersed nuclear elements

z.B. Alu Element: menschliches Genom

280 bp lang

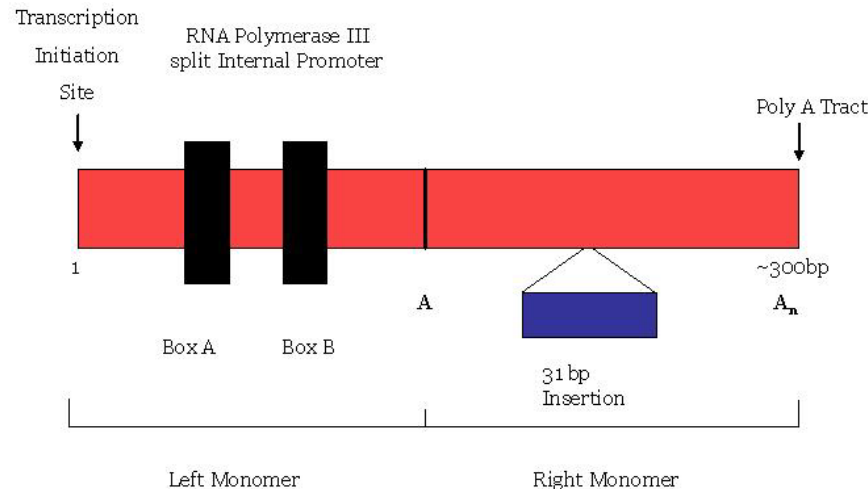
1 Mio Kopien, 1 pro 3 kb

hoher GC-Gehalt

Giemsa-negative Banden

wahrscheinlich über Retrotransposition verbreitetes

7SL RNA-Pseudogen



Band, GC, **SINE, LINE**, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

## Long interspersed nuclear elements

LINE-1 , LINE-2 Familien

LINE-1: Maus und Mensch  
viele aktiv transponierend  
6.1 kb, >100.000 repeats  
Giemsa-positive Banden (AT-reich)



Figure 1. Structure of a Full-Length 6 kb Human L1 Element

Abbreviations: TSD, variable-length target site duplication; 5' UTR, 5' untranslated region; ORF1, first open reading frame; ORF2, second open reading frame; EN, endonuclease domain; RT, reverse transcriptase domain; 3' UTR, 3' untranslated region; AATAAA, hexanucleotide poly(A) signal; and A<sub>n</sub>, the poly(A) tract abuts the hexanucleotide signal in humans.

Band, GC, SINE, LINE, SNP, **ncRNA**, CpG, Exofish, **EST, Genes**, OMIM

Gen: a) Nichtkodierende RNA-Gene

tRNAs

rRNAs

snoRNAs: small nucleolar RNAs: rRNA Prozessierung

snRNA: small nuclear RNA: Spliceosome

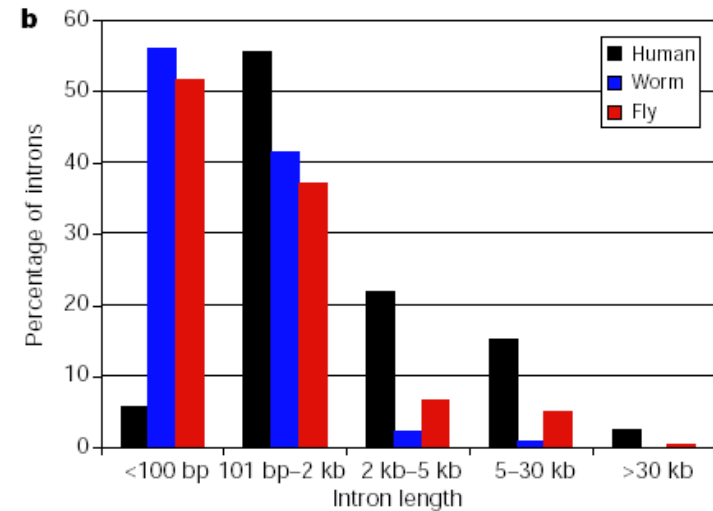
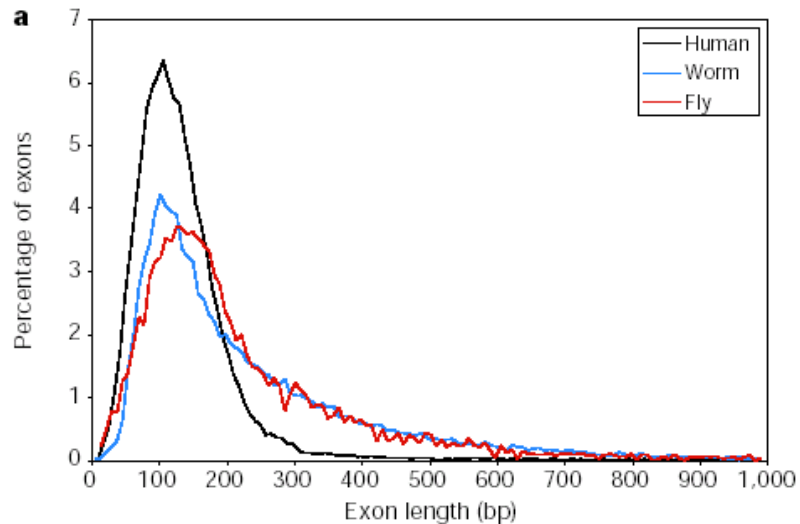
XIST

....

keine ORFs, oft sehr klein, kein polyA

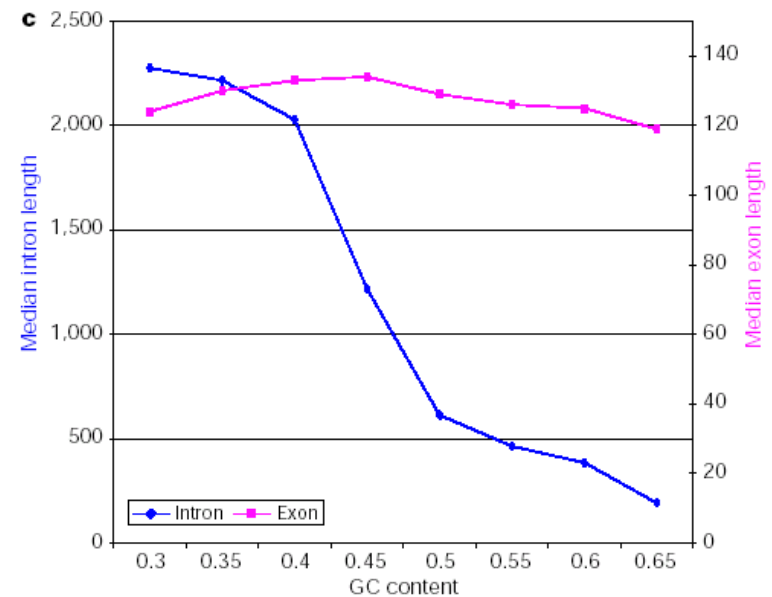
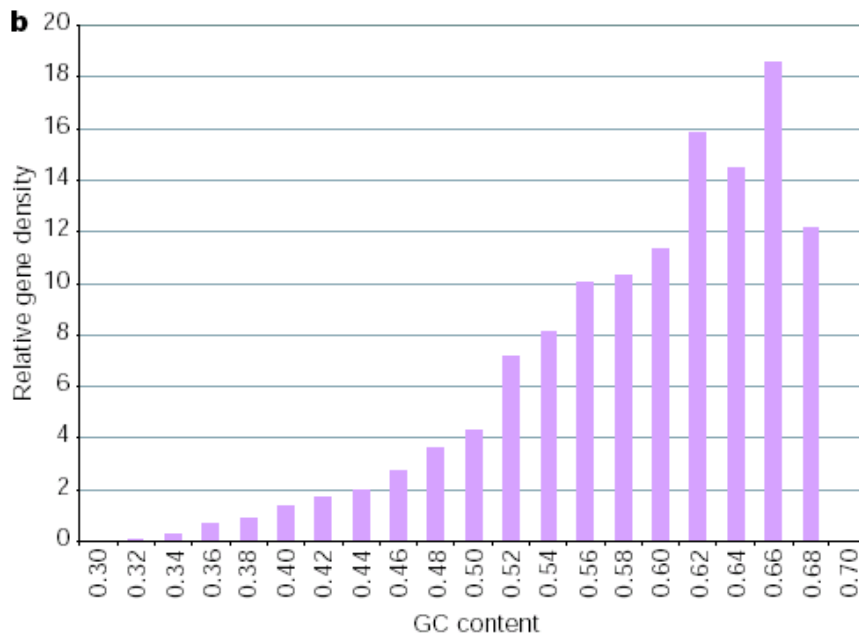
Band, GC, SINE, LINE, SNP, **ncRNA**, CpG, Exofish, **EST, Genes**, OMIM

## b) Protein-kodierende Gene



Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, OMIM

## b) Protein-kodierende Gene



Pseudogen: = inaktive Genkopie

DNA-Region mit hoher Ähnlichkeit zu bekanntem Gen  
nicht funktionell, da Signal für Transkription fehlt (Promotor)  
oder aufgrund von Mutationen, die Translation verhindern

Prozessierte Pseudogene: entstanden durch Retrotransposition  
von (partiell) gespleisster RNA in Genom  
häufig: Fehlen von manchen oder allen Introns

Unprozessierte Pseudogene: entstanden durch Duplikation oder  
Degeneration eines vorhandenen Gens

## Rekorde im humanen Genom !

Was ist das größte Gen ? DMD = Dystrophin-Gen (2.4 Mb)

Was ist die längste kodierende Sequenz ? Titin, 80780 bp

Welches Gen hat die meisten Exons ? Titin, 178 Exons

Welches Gen enthält das längste Exon ? Titin, 17106 bp

# Zusammenfassung Humanes Genom

Anzahl Gene: 32.000 Mensch / Maus: 30.000

Durschnittl. kodierende Sequenzlänge: 1400 bp

Exons pro Transkript: 8.7 (Maus 8.4)

1,5 % des Genoms: kodierende Sequenz

Durschnittl. Genlänge: 30 kb

=> 1/3 des Genoms



Band, GC, SINE, LINE, SNP, ncRNA, CpG, **Exofish**, EST, Genes, OMIM



### **Exofish (Exon Finding by Sequences Homology)**

is a genomic tool based on the assumption that coding regions are more conserved than non-coding regions through evolution. Therefore, if we compare the genome of two species, regions of homology should reveal coding regions. Exofish is performed with sequences of *Tetraodon nigroviridis*, and calibrated to annotate human genome.

Band, GC, SINE, LINE, SNP, ncRNA, CpG, Exofish, EST, Genes, **OMIM**

OMIM - MUSCULAR DYSTROPHY, DUCHENNE TYPE; DMD - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=310200> Search Print

Home Bookmarks The Mozilla Organizat... Latest Builds Mozilla deutsch

NCBI

MIM #310200

- Text
- Description
- Clinical Features
- Other Features
- Inheritance
- Cytogenetics
- Mapping
- Diagnosis
- Clinical Management
- Population Genetics
- Animal Model
- See Also
- References
- Contributors
- Creation Date
- Edit History

MINI-MIM

Clinical Synopsis

Gene map

LinkOut

- CCR
- HGMD

OMIM Online Mendelian Inheritance in Man Johns Hopkins University

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM

Search OMIM for Go Clear

Limits Preview/Index History Clipboard Details

Display Detailed Show: 20 Send to Text

**#310200** Links

**MUSCULAR DYSTROPHY, DUCHENNE TYPE; DMD**

*Alternative titles; symbols*

DUCHENNE MUSCULAR DYSTROPHY  
MUSCULAR DYSTROPHY, PSEUDOHYPERTROPHIC PROGRESSIVE, DUCHENNE TYPE

Gene map locus [Xp21.2, 12q21](#)

**TEXT**

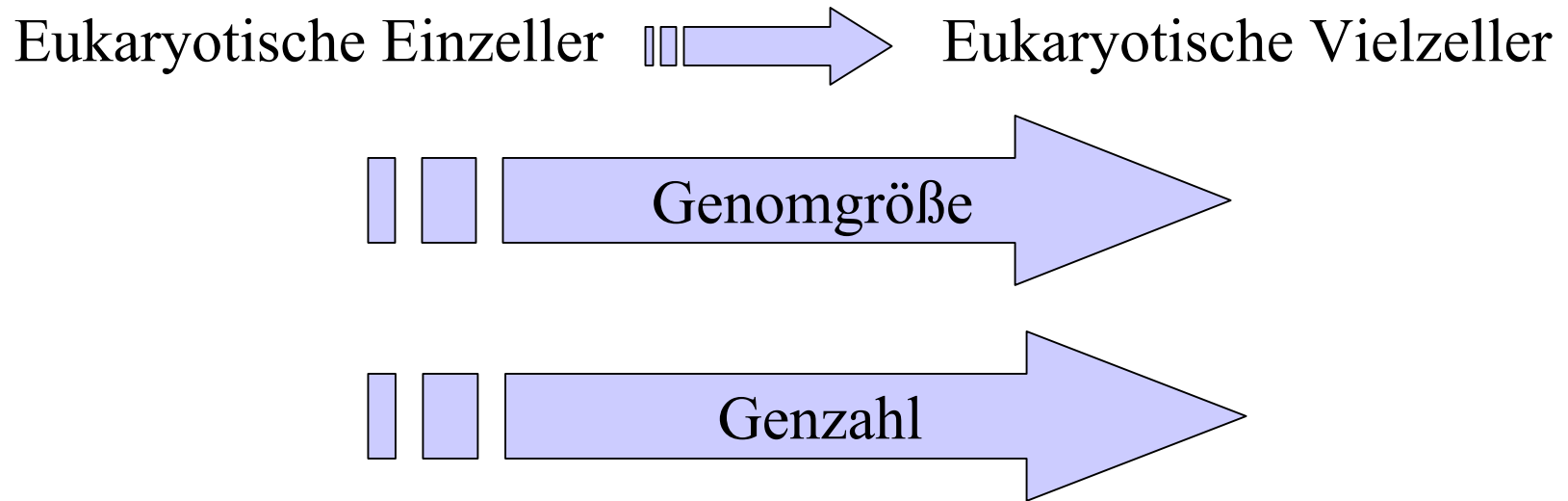
A number sign (#) is used with this entry because Duchenne muscular dystrophy is caused by mutation in the gene encoding dystrophin (DMD; [300377](#)).

**DESCRIPTION**

Dystrophin-associated muscular dystrophies range from the severe Duchenne muscular dystrophy (DMD) to the milder Becker muscular dystrophy (BMD; [300376](#)). Mapping and molecular genetic studies indicate that both are the result of mutations in the huge gene that encodes dystrophin, also symbolized DMD. Approximately two-thirds of the mutations in both forms are deletions of one or many exons in the dystrophin gene. Although there is no clear correlation found between the extent of the deletion and the severity of the disorder, DMD

Start Microsoft PowerPoint... Exploring - D:\Eigene... Nature Index by Year... Download Manager Find: Files named jou... OMIM - MUSCUL... 12:22 PM

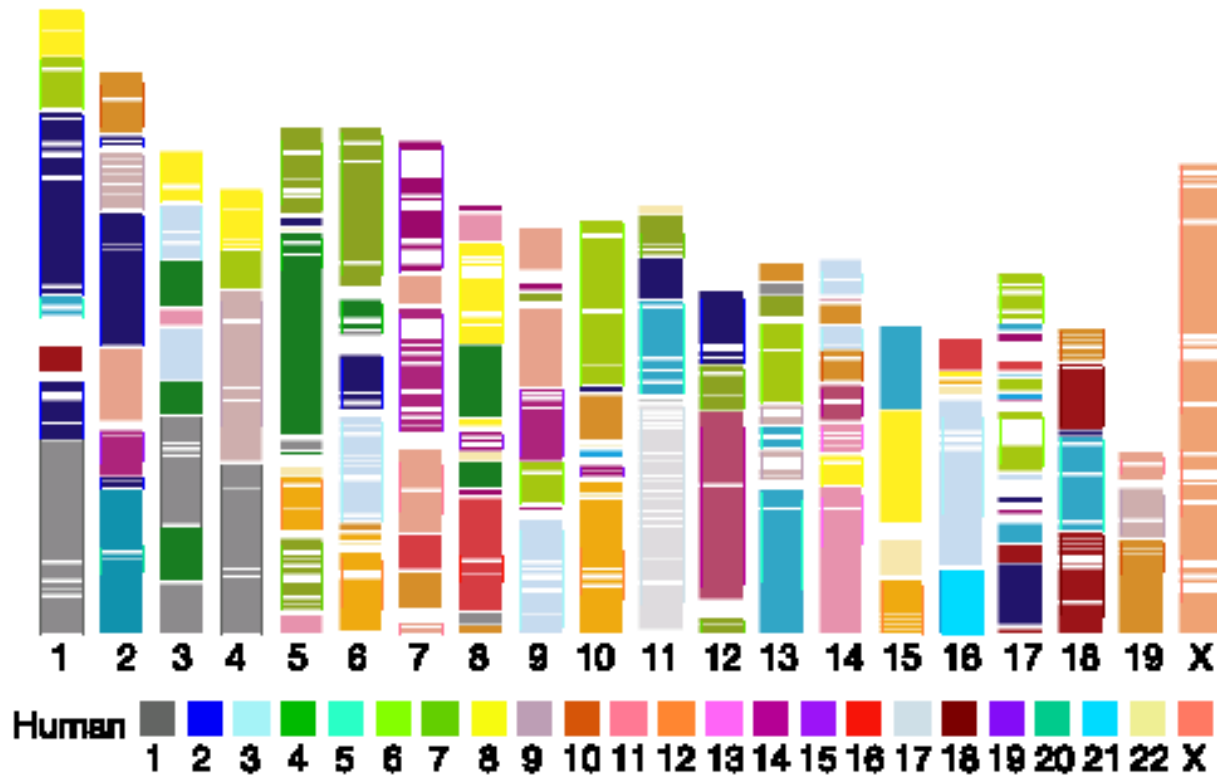
## A2. Evolution des nukleären Genoms



- Duplikationen
- Entstehung von Mosaikgenen aus Teilen anderer Gene
- De novo Genentstehung aus nichtkodierender DNA
- Lateraler Gentransfer zwischen Spezies
- Chromosomale Rearrangements

## A2. Evolution des nukleären Genoms

### Chromosomale Umbauten Vergleich Maus/Mensch



## A2. Evolution des nukleären Genoms – wichtige Begriffe !

**Paraloge:** ähnliche nicht-allelische chromosomale oder DNA-Segmente innerhalb einer Spezies

weisen auf engen evolutionären Zusammenhang hin

**Orthologe:** Ähnlichkeit zwischen chromosomalen oder DNA-Segmenten verschiedener Spezies

**Homologie:** ursprünglich Beziehung allelischer chromosomaler Segmente im weiteren Sinne jede Art von Verwandtschaft, welche auf gemeinsamen evolutionären Ursprung hinweist, innerhalb und zwischen Spezies