

BIO-INFORMATICA

COURSE 3B VERGELIJKENDE GENOOMANALYSE

WEEK 2: GENEN EN GC PERCENTAGE

OVERZICHT ONDERZOEKSPLAN 3B

Week	Onderwerp	Activiteiten	M&M
1	Oriëntatie op onderwerp Sequenties verzamelen	Inlezen Stap 1	NCBI database
2	Genen en GC percentage	Stap 2	Python script
3	Codon gebruik	Stap 3	Python script
4	Eiwit karakterisering / aminozuren	Stap 4	Python script
5	Kenmerken oppervlakte proteïnen	Stap 5	Python script
6	Fylogenetisch onderzoek	Stap 6	Bioinf tools
7	Onderzoeksverslag		

ONDERZOEKSVRAGEN GENOOMONDERZOEK

HIV kan bij de mens AIDS veroorzaken.

AIDS → Acquired Immune Deficiency Syndrome

Vooraf HIV-1 erg gevaarlijk (pandemie)

I. Waarom is HIV-1 hoog virulent?

II. Wat is de fylogenetische oorsprong van HIV-1?

DEZE WEEK

Onderwerpen

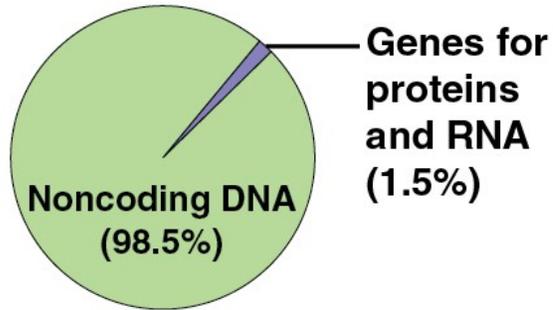
- Humane genoom
- Genoom sequencing
- Onderzoeksvelden Bio-IT
 - Genomics
 - Transcriptomics
 - Metagenomics
- Verschillen genoom prokaryoten en eukaryoten
- Alternative splicing
- GC%

Materialen

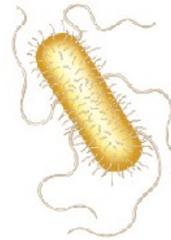
- Campbell Hfst 20.3, 20.4
- Pevsner Hfst 8, blz 308-314
- Pevsner Hfst 17 fig 17.1,17.2, table 17.5,
 - 17.6
- Pevsner Hfst 17 817-818, 969 GC content
- Pevnsner Hfst 20, 969 CpG islands

What are some questions that can be explored by sequencing and comparing genomes?

What are the **functions** of the human genome?

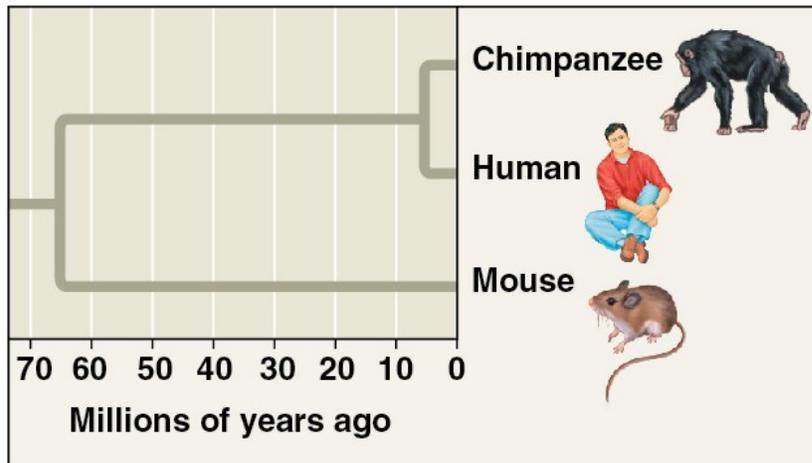


How do genomes differ in **number of genes**?



Escherichia coli 4,400 genes *Homo sapiens* 21,300 genes *Zea mays* (corn) 32,000 genes

What do gene sequences tell us about **evolutionary relationships** between species?



How do genomes **evolve** over time?



Elephant shark: slower-evolving genome

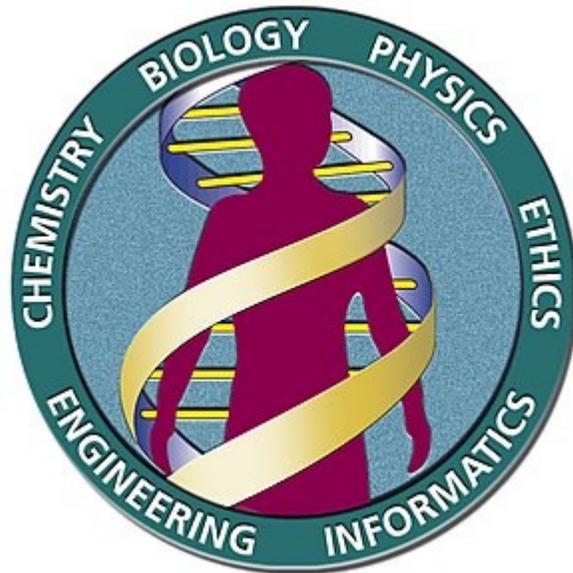


Tiger tail sea horse: faster-evolving genome

Fig. 20.1B, Cambell Biology, 2021 Global edition

HET MENSELIJK GENOOM PROJECT

Duur: 1990 - 2006



[Human Genome Project - Wikipedia](#)

The Sequence of the Human Genome

J. Craig Venter,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹ Granger G. Sutton,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹ Jeannine D. Gocayne,¹ Peter Amanatides,¹ Richard M. Ballew,¹ Daniel H. Huson,¹ Jennifer Russo Wortman,¹ Qing Zhang,¹ Chinnappa D. Kodira,¹ Xiangqun H. Zheng,¹ Lin Chen,¹ Marian Skupski,¹ Gangadharan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹ George L. Gabor Miklos,² Catherine Nelson,² Samuel Broder,¹ Andrew G. Clark,⁴ Joe Nadeau,⁵ Victor A. McKusick,⁶ Norton Zinder,⁷ Arnold J. Levine,⁷ Richard J. Roberts,⁸ Mel Simon,⁹ Carolyn Slayman,¹⁰ Michael Hunkapiller,¹¹ Randall Bolanos,¹ Arthur Delcher,¹ Ian Dew,¹ Daniel Fasulo,¹ Michael Flanigan,¹ Liliana Florea,¹ Aaron Halpern,¹ Sridhar Hannenhalli,¹ Saul Kravitz,¹ Samuel Levy,¹ Clark Mobarry,¹ Knut Reinert,¹ Karin Remington,¹ Jane Abu-Threideh,¹ Ellen Beasley,¹ Kendra Biddick,¹ Vivien Bonazzi,¹ Rhonda Brandon,¹ Michele Cargill,¹ Ishwar Chandramouliswaran,¹ Rosane Charlab,¹ Kabir Chaturvedi,¹ Zuoming Deng,¹ Valentina Di Francesco,¹ Patrick Dunn,¹ Karen Eilbeck,¹ Carlos Evangelista,¹ Andrei E. Gabrielian,¹ Weiniu Gan,¹ Wangmao Ge,¹ Fangcheng Gong,¹ Zhiping Gu,¹ Ping Guan,¹ Thomas J. Heiman,¹ Maureen E. Higgins,¹ Rui-Ru Ji,¹ Zhaoxi Ke,¹ Karen A. Ketchum,¹ Zhongwu Lai,¹ Yiding Lei,¹ Zhenya Li,¹ Jiayin Li,¹ Yong Liang,¹ Xiaoying Lin,¹ Fu Lu,¹ Gennady V. Merkulov,¹ Natalia Milshina,¹ Helen M. Moore,¹ Ashwinikumar K Naik,¹ Vaibhav A. Narayan,¹ Beena Neelam,¹ Deborah Nusskern,¹ Douglas B. Rusch,¹ Steven Salzberg,¹² Wei Shao,¹ Bixiong Shue,¹ Jingtao Sun,¹ Zhen Yuan Wang,¹ Aihui Wang,¹ Xin Wang,¹ Jian Wang,¹ Ming-Hui Wei,¹ Ron Wides,¹³ Chunlin Xiao,¹ Chunhua Yan,¹ Alison Yao,¹ Jane Ye,¹ Ming Zhan,¹ Weiqing Zhang,¹ Hongyu Zhang,¹ Qi Zhao,¹ Liansheng Zheng,¹ Fei Zhong,¹ Wenyan Zhong,¹ Shiaoping C. Zhu,¹ Shaying Zhao,¹² Dennis Gilbert,¹ Suzanna Baumhueter,¹ Gene Spier,¹ Christine Carter,¹ Anibal Cravchik,¹ Trevor Woodage,¹ Feroze Ali,¹ Huijin An,¹ Aderonke Awe,¹ Danita Baldwin,¹ Holly Baden,¹ Mary Barnstead,¹ Ian Barrow,¹ Karen Beeson,¹ Dana Busam,¹ Amy Carver,¹ Angela Center,¹ Ming Lai Cheng,¹ Liz Curry,¹ Steve Danaher,¹ Lionel Davenport,¹ Raymond Desilets,¹ Susanne Dietz,¹ Kristina Dodson,¹ Lisa Doup,¹ Steven Ferriera,¹ Neha Garg,¹ Andres Gluecksmann,¹ Brit Hart,¹ Jason Haynes,¹ Charles Haynes,¹ Cheryl Heiner,¹ Suzanne Hladun,¹ Damon Hostin,¹ Jarrett Houck,¹ Timothy Howland,¹ Chinyere Ibegwam,¹ Jeffery Johnson,¹ Francis Kalush,¹ Lesley Kline,¹ Shashi Koduru,¹ Amy Love,¹ Felecia Mann,¹ David May,¹ Steven McCawley,¹ Tina McIntosh,¹ Ivy McMullen,¹ Mee Moy,¹ Linda Moy,¹ Brian Murphy,¹ Keith Nelson,¹ Cynthia Pfannkoch,¹ Eric Pratts,¹ Vinita Puri,¹ Hina Qureshi,¹ Matthew Reardon,¹ Robert Rodriguez,¹ Yu-Hui Rogers,¹ Deanna Romblad,¹ Bob Ruhfel,¹ Richard Scott,¹ Cynthia Sitter,¹ Michelle Smallwood,¹ Erin Stewart,¹ Renee Strong,¹ Ellen Suh,¹ Reginald Thomas,¹ Ni Ni Tint,¹ Sukyee Tse,¹ Claire Vech,¹ Gary Wang,¹ Jeremy Wetter,¹ Sherita Williams,¹ Monica Williams,¹ Sandra Windsor,¹ Emily Winn-Deen,¹ Kerriellen Wolfe,¹ Jayshree Zaveri,¹ Karena Zaveri,¹ Josep F. Abril,¹⁴ Roderic Guigó,¹⁴ Michael J. Campbell,¹ Kimmen V. Sjolander,¹ Brian Karlak,¹ Anish Kejarival,¹ Huaiyu Mi,¹ Betty Lazareva,¹ Thomas Hatton,¹ Apurva Narechania,¹ Karen Diemer,¹ Anushya Muruganujan,¹ Nan Guo,¹ Shinji Sato,¹ Vineet Bafna,¹ Sorin Istrail,¹ Ross Lippert,¹ Russell Schwartz,¹ Brian Walenz,¹ Shibu Yooseph,¹ David Allen,¹ Anand Basu,¹ James Baxendale,¹ Louis Blick,¹ Marcelo Caminha,¹ John Carnes-Stine,¹ Parris Caulk,¹ Yen-Hui Chiang,¹ My Coyne,¹ Carl Dahlke,¹ Anne Deslattes Mays,¹ Maria Dombroski,¹ Michael Donnelly,¹ Dale Ely,¹ Shiva Esparham,¹ Carl Fosler,¹ Harold Gire,¹ Stephen Glanowski,¹ Kenneth Glasser,¹ Anna Glodek,¹ Mark Gorokhov,¹ Ken Graham,¹ Barry Gropman,¹ Michael Harris,¹ Jeremy Heil,¹ Scott Henderson,¹ Jeffrey Hoover,¹ Donald Jennings,¹ Catherine Jordan,¹ James Jordan,¹ John Kasha,¹ Leonid Kagan,¹ Cheryl Kraft,¹ Alexander Levitsky,¹ Mark Lewis,¹ Xiangjun Liu,¹ John Lopez,¹ Daniel Ma,¹ William Majoros,¹ Joe McDaniel,¹ Sean Murphy,¹ Matthew Newman,¹ Trung Nguyen,¹ Ngoc Nguyen,¹ Marc Nodell,¹ Sue Pan,¹ Jim Peck,¹ William Rowe,¹ Robert Sanders,¹ John Scott,¹ Michael Simpson,¹ Thomas Smith,¹ Arlan Sprague,¹ Timothy Stockwell,¹ Russell Turner,¹ Eli Venter,¹ Mei Wang,¹ Meiyuan Wen,¹ David Wu,¹ Mitchell Wu,¹ Ashley Xia,¹ Ali Zandieh,¹ Xiaohong Zhu¹

HET MENSELIJK GENOOM PROJECT

- DNA kwam van een klein aantal mensen
- Deze resultaten werden bekeken, en wetenschappers poolde de data en waren het eens over eens over een **referentie genoom**.
- Het doel was en vinden van de complete nucleotide sequentie van elk chromosoom

HET MENSELIJKE GENOOM PROJECT EN VERDER

- Een grote boost in het onderzoek met next generation sequencing → sneller, meer data.
- Met deze grote data nog onderzoek doen waren en nieuwe technieken nodig → bio-informatica, databases, en software.
- Alles kwam ook op het internet beschikbaar.
- Over de jaren heen o.a.:
 - NCBI – National Center for Biotechnology Information
 - EMBL – European Molecular Biology Laboratory
 - DDBJ – DNA Bank of Japan
 - BGI – Beijing Genomics Institute China
 - BLAST – Basic Local Alignment Search Tool
 - PDB – Protein Data Bank
 - ...

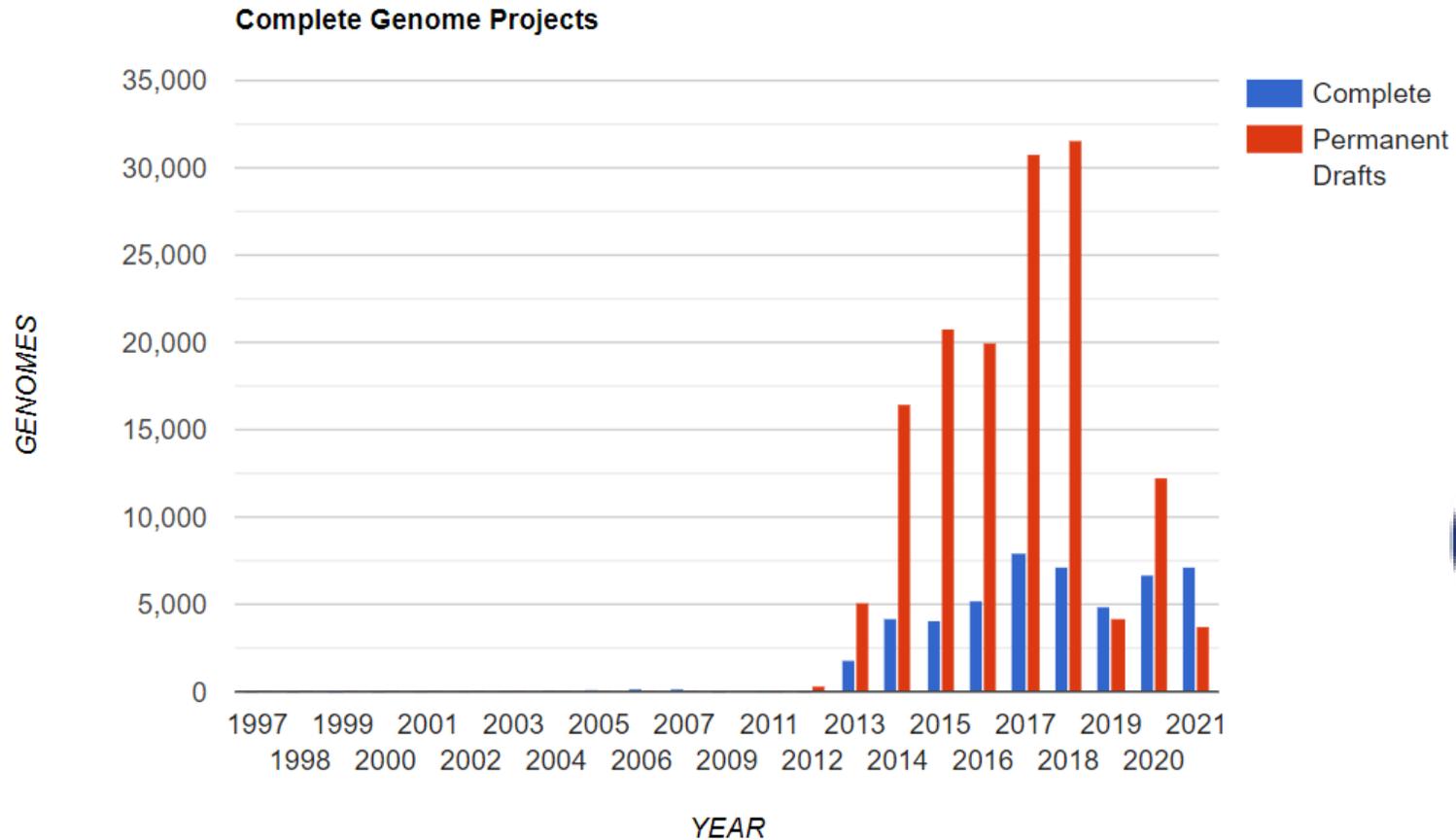
GENOME SEQUENCING

- “**GOLD**: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.”



Genomes online database
[JGI GOLD | Home \(doe.gov\)](#)

GENOME SEQUENCING

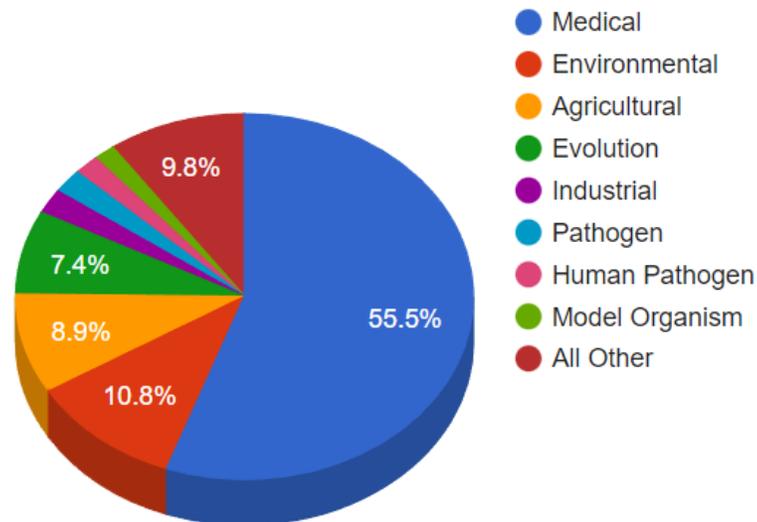


Genomes online database
[JGI GOLD | Home \(doe.gov\)](http://JGI_GOLD.Home.doe.gov)

GENOME SEQUENCING

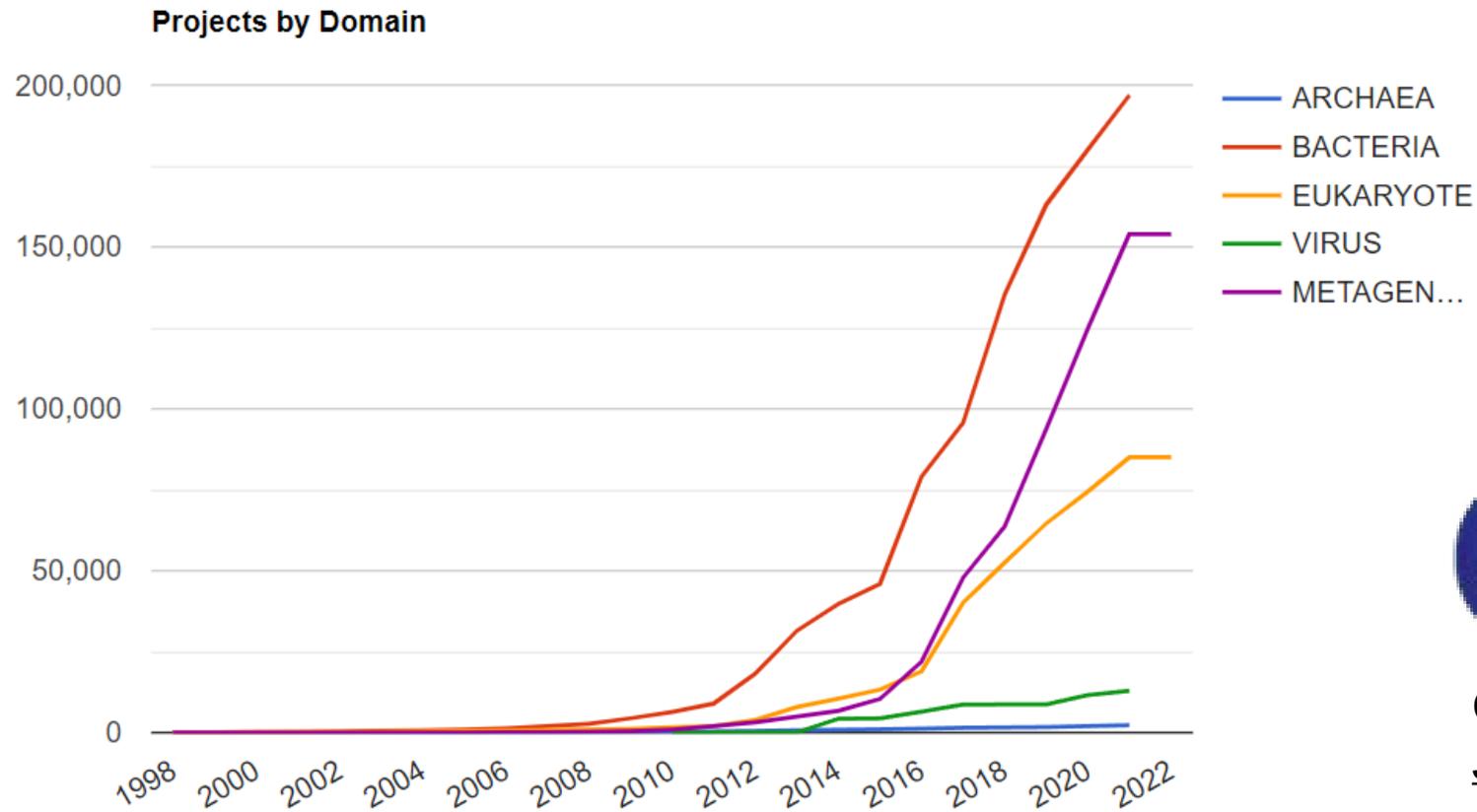
- 1995 - 2000: Veel pathogene organismen & model organismen (gist, *Escherichia coli*, *Bacillus subtilis*)
- 2000 - nu: Genomen van veel industriële micro-organismen (schimmels, melkzuur bacteriën, verschillende productie gastheren)

Project Relevance of Bacterial Projects



Genomes online database
[JGI GOLD | Home \(doe.gov\)](http://JGI_GOLD.Home.doe.gov)

GENOME SEQUENCING



Genomes online database
[JGI GOLD | Home \(doe.gov\)](http://JGI_GOLD.Home.doe.gov)

HOW VERKRIJGEN WE GENOMEN

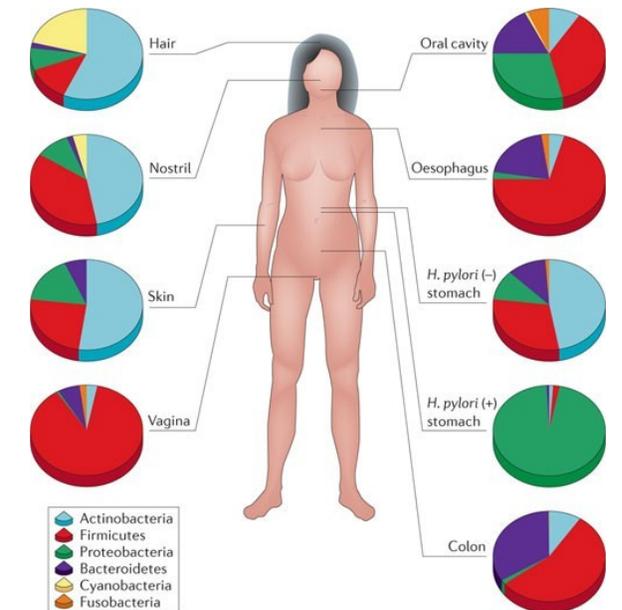
- Genome sequencing (direct)
- Metagenomics (direct maar in populatie)
- Transcriptomics (indirect)

METAGENOMICS

- De studie van een collectie genomen van een gemixte gemeenschap.
- Dit zijn meestal microbiële gemeenschappen.
- Monsters uit een “omgeving”:
 - Milieu
 - Organisme



JGI GOLD | Biosample Map (doe.gov)



The human microbiome: at the interface of health and disease | Nature Reviews Genetics

METAGENOMICS (ALS BIO-INFORMATICS OP EXPEDITIE)



JGI GOLD | Biosample Map (doe.gov)

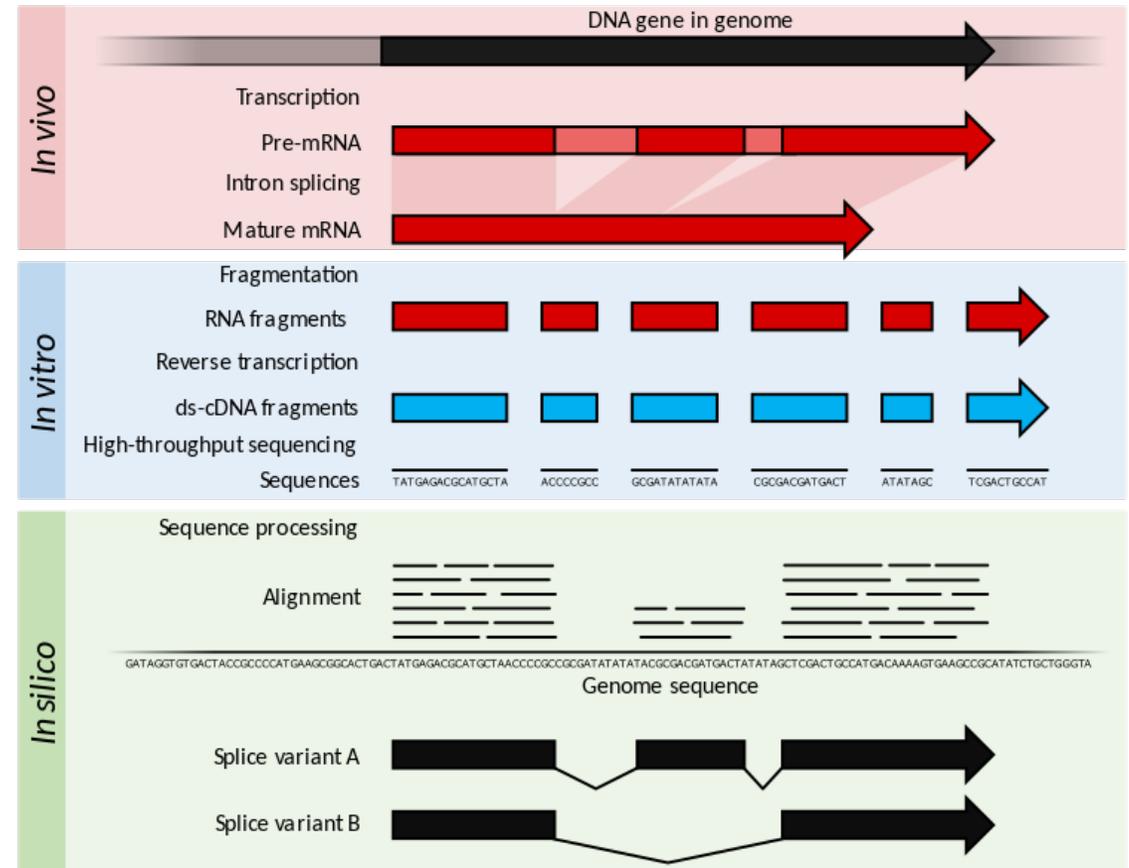
The RV Pelagia (NIOZ) in the North Sea on its way to Iceland.



Expedition STRATIPHYT-17 - NIOZ

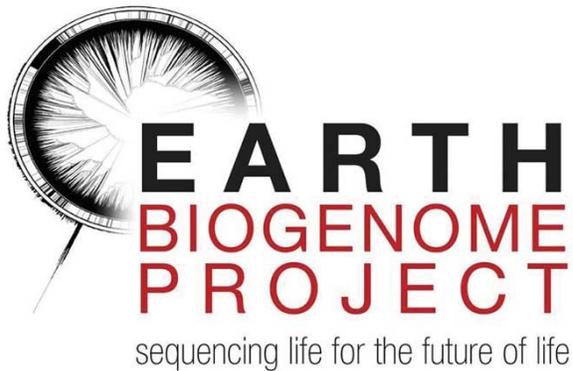
TRANSCRIPTOMICS

- RNA-Seq: sequencing van RNA moleculen
 - Vaak zijn dit mRNA molecule
 - Transcriptomen
 - Dan weet je welke genen een functioneel product (eiwit) worden



RNA-Seq – Wikipedia

ANDERE GENOOM SEQUENTIE PROJECTEN (2022)



[Earth BioGenome Project](#)

The Earth BioGenome Project (EBP), a *moonshot* for biology, aims to sequence, catalog and characterize the genomes of all of Earth's eukaryotic biodiversity over a period of ten years.



[Home](#) | [10KP](#) | [CNGB](#)

The 10,000 plants (tenKP or 10KP) aims to sequence over 10,000 genomes representing every major clade of plants and eukaryotic microbes. This project would generate large-scale plant genome data within the next five years (2017-2022), addressing fundamental questions about plant evolution.



[Home - Genome10K \(ucsc.edu\)](#)

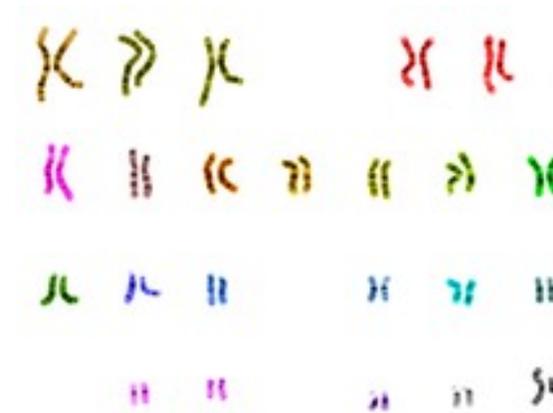
Genome 10K is a project to sequence the genome of at least one individual from each vertebrate genus, approximately 10,000 genomes, at least one individual from each of the approximately 66,000 vertebrate species.

ANDERE GENOOM SEQUENTIE PROJECTEN

- Sequenties van duizenden genomen nu beschikbaar.
- Heel veel zijn nog in staat van analyse (draft) of blijven zo.
- Rond 22.000 zijn metagenomen.

HET GENOOM

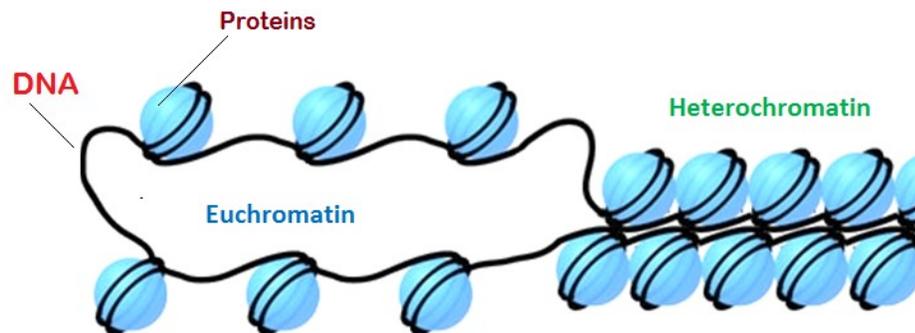
- Wat is een genoom?
 - “In modern molecular biology, the genome is the entirety of an organism's hereditary information. It is encoded either in DNA or, for many types of virus, in RNA.”
 - Betreft coderende (genen) en niet coderende onderdelen.
 - Inclusief mitochondriaal DNA (alle eukaryoten) en chloroplast DNA (in planten en algen).
- Waar bestaat het humane genoom uit?
 - 46 Chromosomen
 - Mitochondriaal DNA
 - ± 25000 genen
 - ± 3.200.000.000 bp



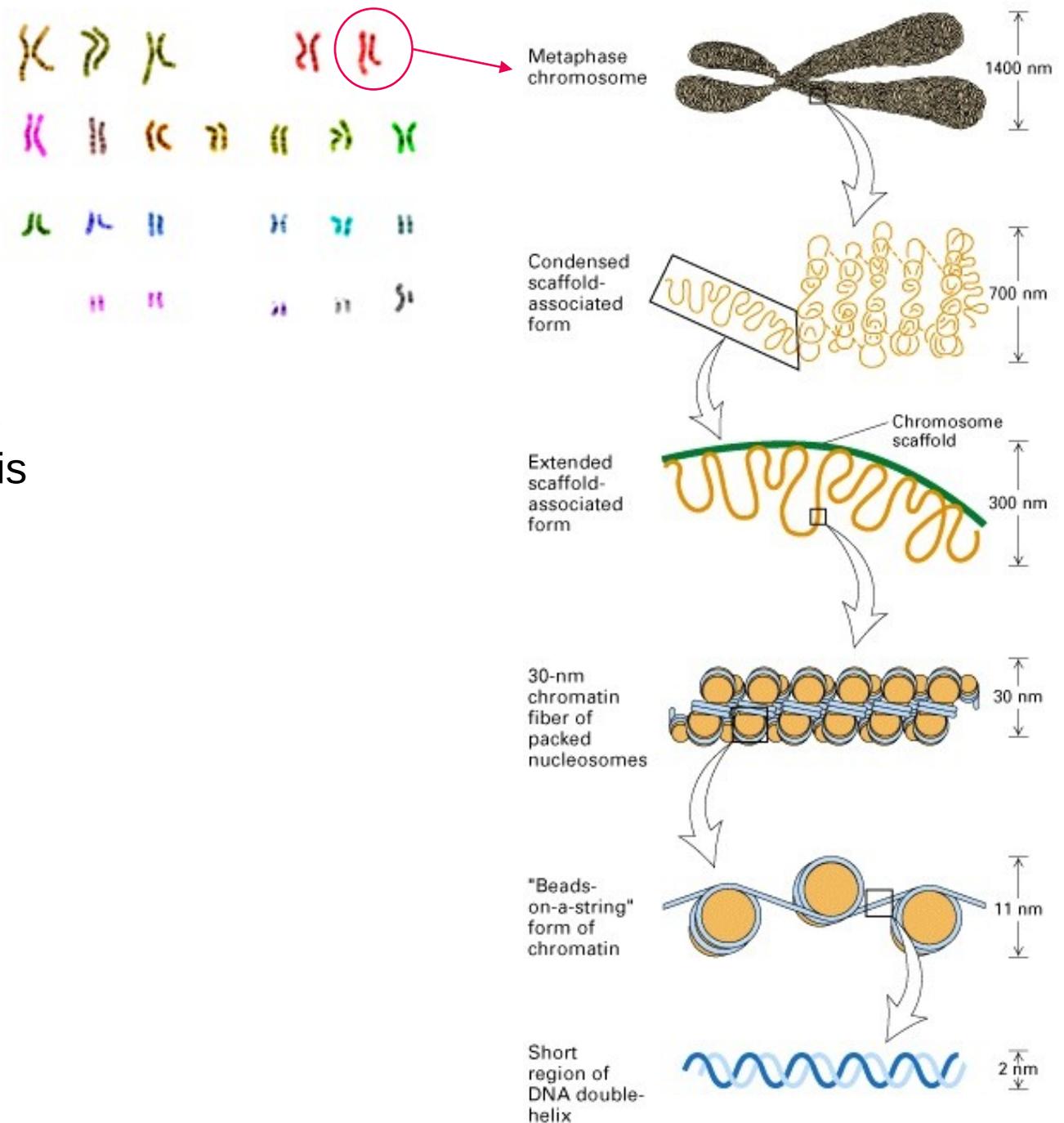
Karyotype van het menselijk genoom (excl. mitochondriaal).

HET GENOOM EN OPVOUWEN VAN DNA

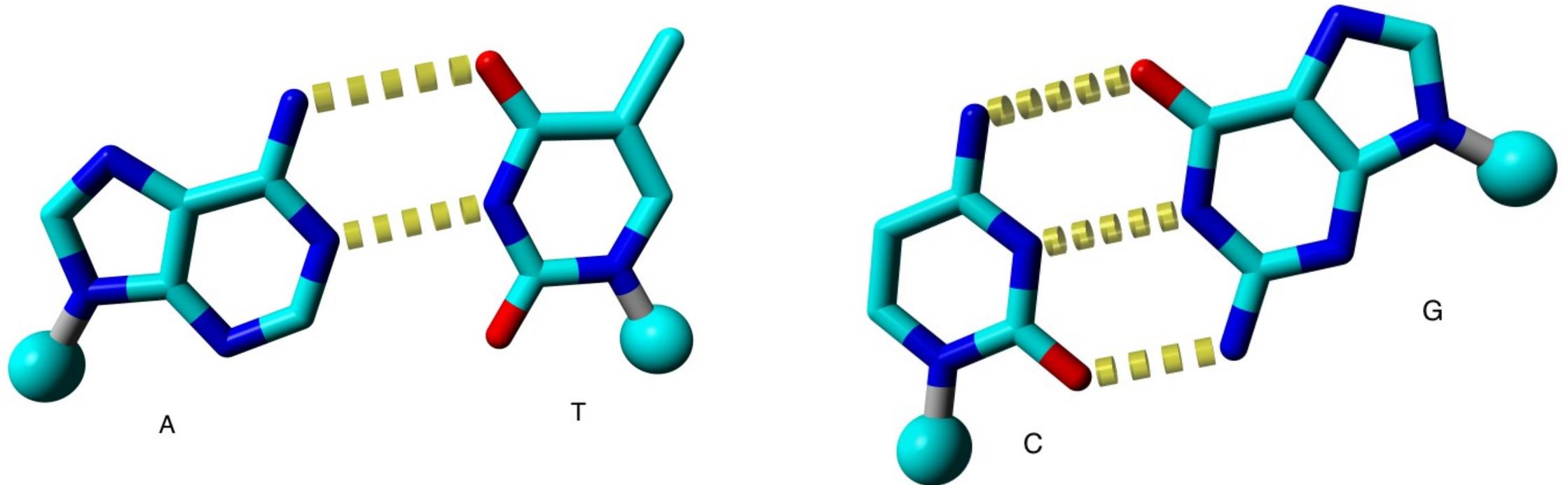
- Euchromatine is minder compact. Hier bevinden zich de meeste actieve genen.
- Heterochromatine is veel compacter en is voornamelijk inactief.



New Cure for Cancer? By: Cameron Bourgeois – Biology 441 (wordpress.com)

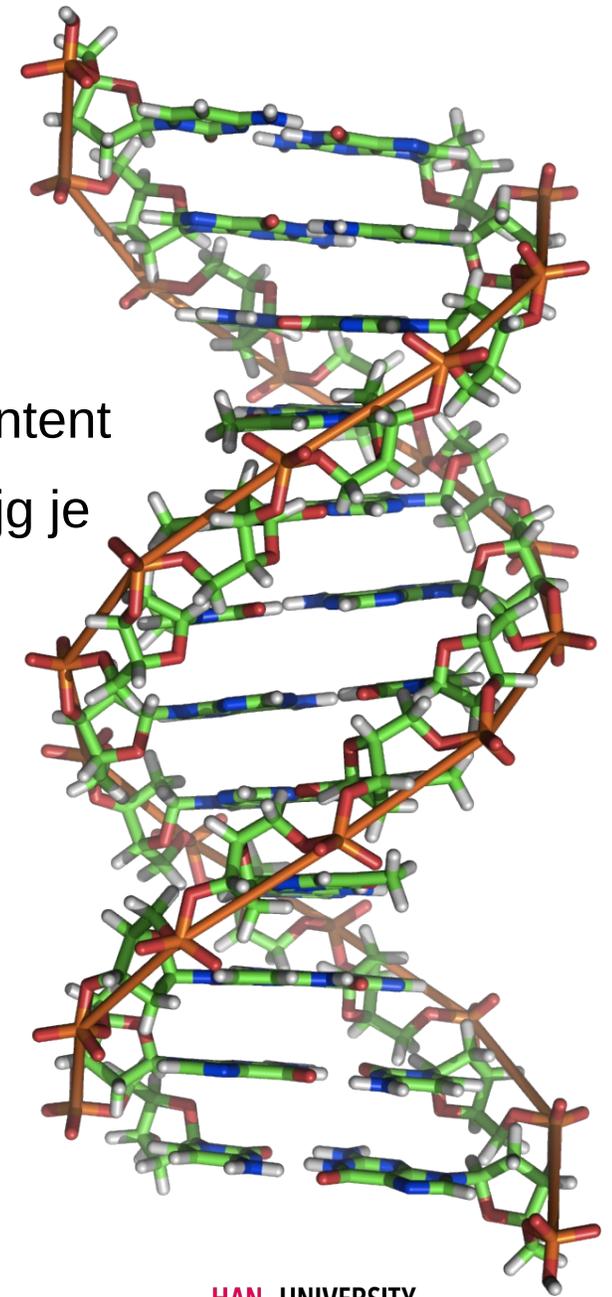


BOUWSTENEN DNA: NUCLEOTIDEN



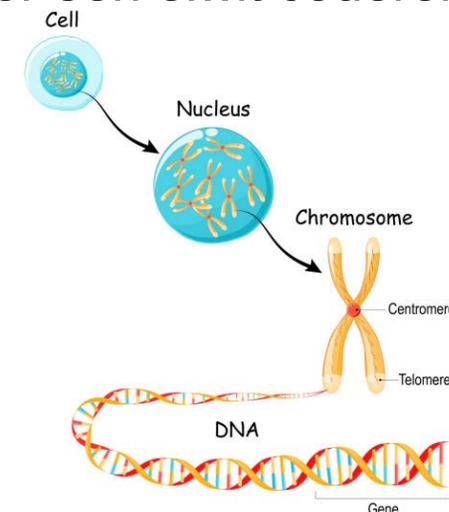
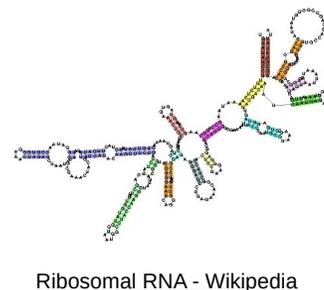
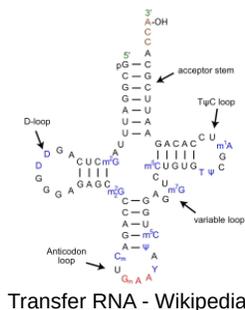
DNA EN GC CONTENT

- DNA met een hoge GC content is stabielere dan DNA met lage GC content
- Doordat de nucleotiden op bepaalde manieren op elkaar stapelen krijg je hele stukken stabiel en andere minder stabiel DNA, dit komt voornamelijk door de gunstige stapeling is tussen GC paren.



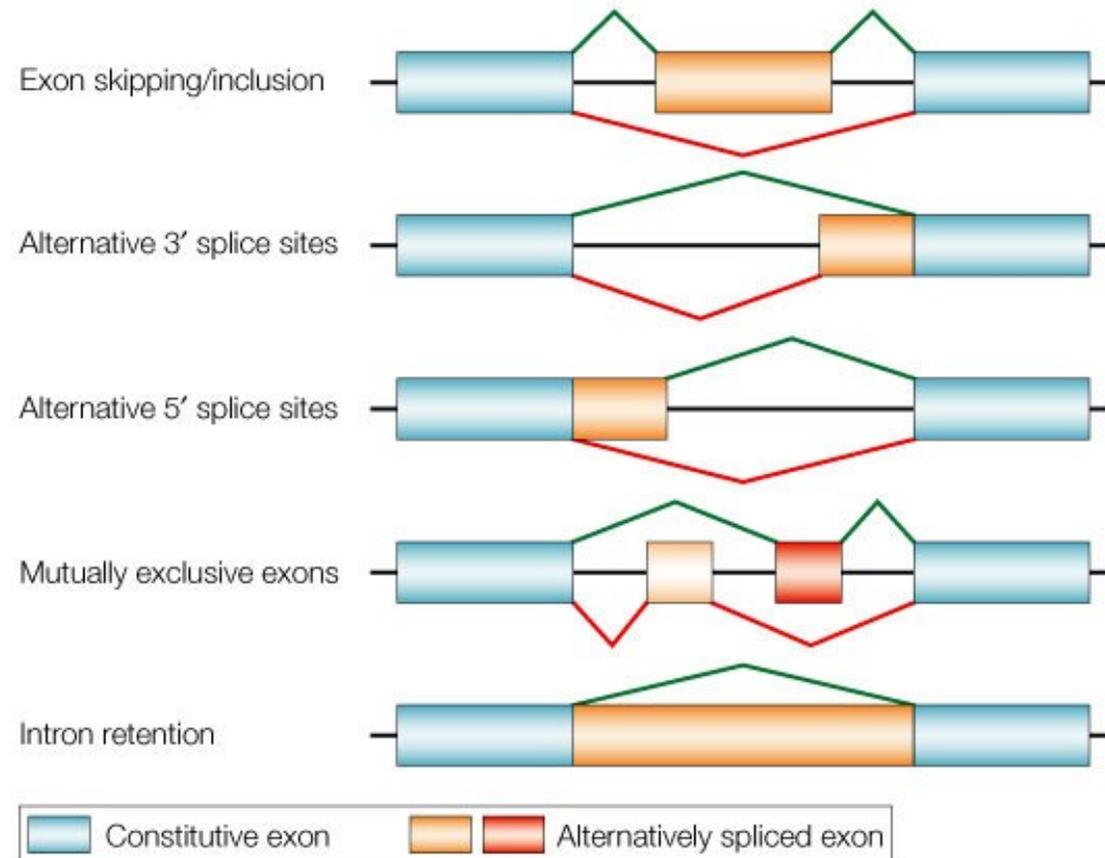
VAN DNA NAAR GENEN: WAT IS EEN GEN?

- Een gen is de basis eenheid van erfelijkheid en een sequentie van nucleotiden in het DNA die coderen voor (functionele) producten, zoals eiwitten en RNA.
- Door alternatieve splicing kan een gen voor meerdere producten coderen.
- Een gen kan:
 - Coding zijn: codeert voor een eiwit.
 - Non-coding zijn: voor bijvoorbeeld RNAs die niet voor een eiwit coderen of DNA met andere regulatoire functies.
 - tRNAs, rRNAs, regulatoire RNAs.
 - Centromeren, telomeren, origin of replication.



Centromere afbeeldingen, beelden en stockfoto's - iStock (istockphoto.com)

GENEN, EXONEN EN ALTERNATIVE SPLICING



<http://www.nature.com/scitable/content/a-schematic-representation-of-alternative-splicing-14263384>

Drosophila *Dscam* gene: 38.016 splice variants

Down syndrome cell adhesion molecule (Dscam): axon guidance receptor

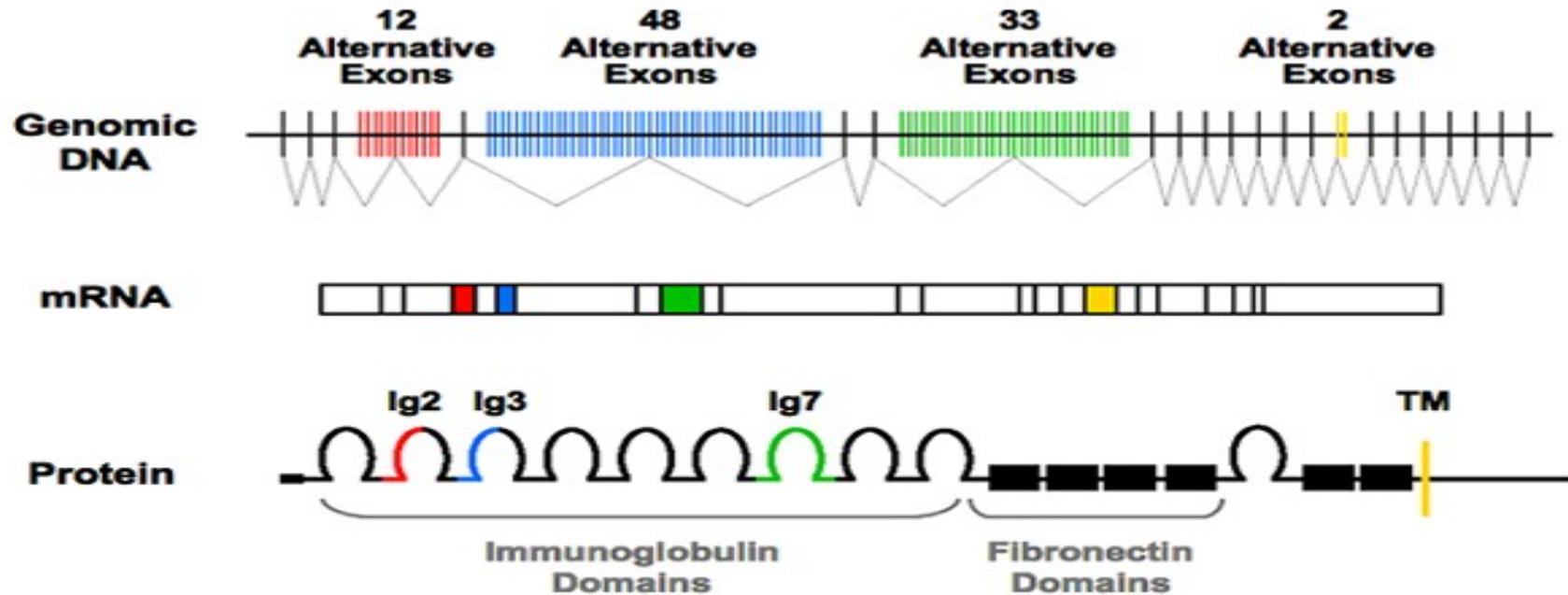


Figure 6: *Dscam*1 is alternatively spliced. There are four blocks of alternative exons (red, blue, green and yellow). Each mRNA contains one and only one alternative. The proteins generated from the locus all have the same domain structure with 10 immunoglobulin domains, 6 fibronectin type III repeats, and transmembrane domain and a cytoplasmic domain. As splicing within each block is independent of the locus has the potential to generate 19,008 ectodomains each with a unique combination of Ig2, Ig3 and Ig7 domains. See Schmucker et al. 2000; Wojtowicz et al 2004, 2007.

<http://www.biocchem.ucla.edu/Labs/Zipursky/research.html>

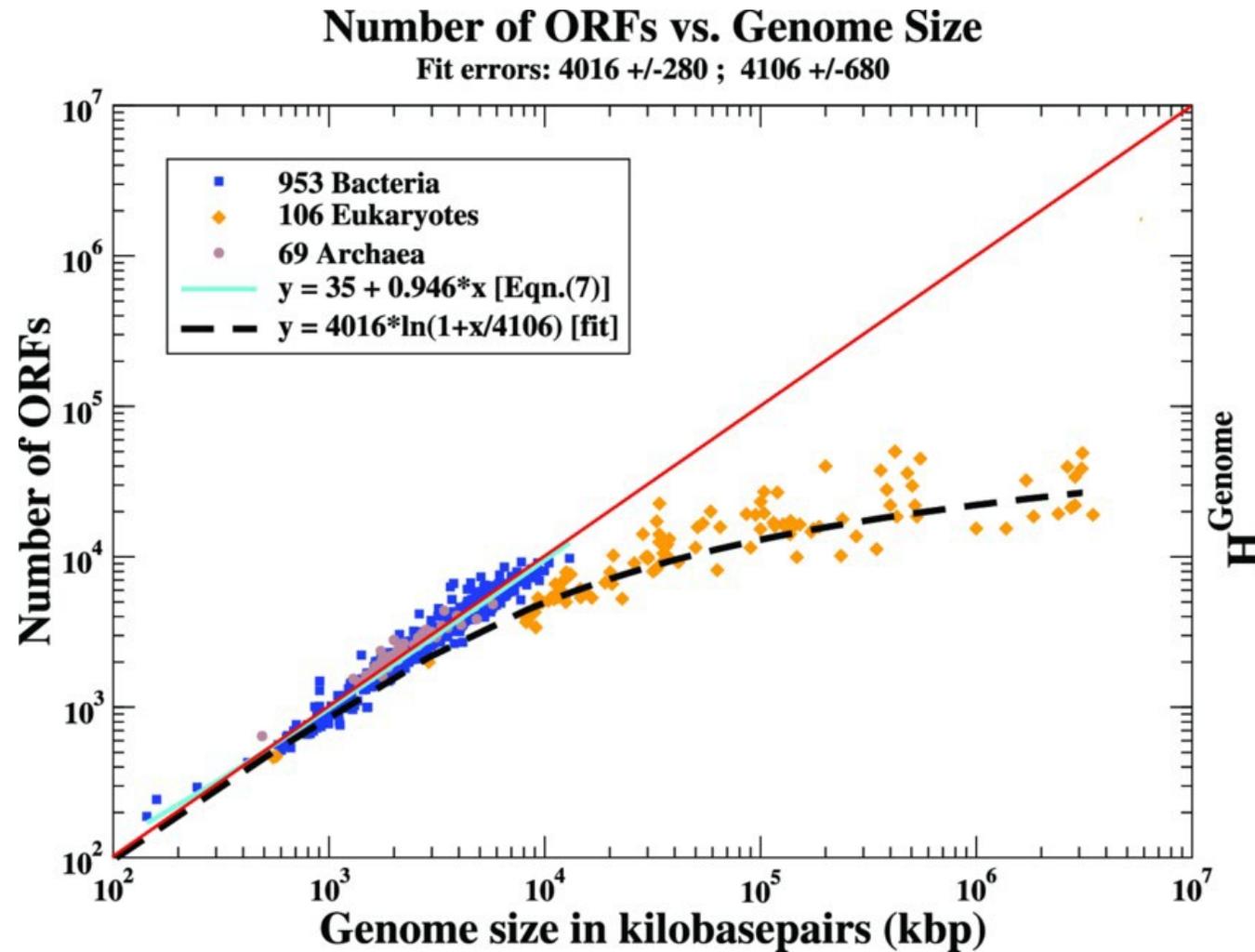
GENOOM GROOTTE

- Genomen van de meest bacteriën en archaea tussen 1 tot 12 miljoen base paren (Mb).
- Eukaryoten genomen groter tussen 10 tot > 100.000 Mb.
- Virussen 3 baseparen tot 1 Mb.
- Er lijkt niet echt een relatie te zijn tussen genoom grootte en fenotype.

AANTAL GENEN

- Vrij levende (free-living) bacteriën en archaea hebben rond de 1.500 tot 7.500 genen.
- Uni cellulaire schimmels (eukaryoot) rond de 5000 en multi cellulaire eukaryoten rond de 40.000.

GENOOM GROOTTE EN AANTAL GENEN



Genome Sizes and the Benford Distribution (plos.org)

GENEN NIET ALTIJD GECORRELEERD AAN GENOOM GROOTTE

- Wetenschappers hadden voorspeld dat de mens rond de 50.000 tot 100.000 genen zou hebben. Maar we hebben maar rond de 21.300.
- Vertebraten produceren meer producten per gen door alternative splicing.

Table 20.1 Genome Sizes and Estimated Numbers of Genes*

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
<i>Haemophilus influenzae</i>	1.8	1,700	940
<i>Escherichia coli</i>	4.6	4,400	950
Archaea			
<i>Archaeoglobus fulgidus</i>	2.2	2,500	1,130
<i>Methanosarcina barkeri</i>	4.8	3,600	750
Eukaryotes			
<i>Saccharomyces cerevisiae</i> (yeast, a fungus)	12	6,300	525
<i>Utricularia gibba</i> (floating bladderwort)	82	28,500	348
<i>Caenorhabditis elegans</i> (nematode)	100	20,100	200
<i>Arabidopsis thaliana</i> (mustard family plant)	120	27,000	225
<i>Drosophila melanogaster</i> (fruit fly)	165	14,000	85
<i>Daphnia pulex</i> (water flea)	200	31,000	155
<i>Zea mays</i> (corn)	2,300	32,000	14
<i>Ailuropoda melanoleuca</i> (giant panda)	2,400	21,000	9
<i>Homo sapiens</i> (human)	3,000	21,300	7
<i>Paris japonica</i> (Japanese canopy plant)	149,000	ND	ND

*Some values given here are likely to be revised as genome analysis continues. Mb = million base pairs; the haploid number is used because it represents a complete set of genetic information. ND = not determined.

GENEN NIET ALTIJD GECORRELEERD AAN GENOOM GROOTTE

- Mens en andere zoogdieren hebben de laagste gen dichtheid (aantal genen per lengte DNA).
- Multi cellulaire eukaryoten hebben veel intronen in de genen en grote hoeveelheden noncoding DNA tussen genen.

Table 20.1 Genome Sizes and Estimated Numbers of Genes*

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
<i>Haemophilus influenzae</i>	1.8	1,700	940
<i>Escherichia coli</i>	4.6	4,400	950
Archaea			
<i>Archaeoglobus fulgidus</i>	2.2	2,500	1,130
<i>Methanosarcina barkeri</i>	4.8	3,600	750
Eukaryotes			
<i>Saccharomyces cerevisiae</i> (yeast, a fungus)	12	6,300	525
<i>Utricularia gibba</i> (floating bladderwort)	82	28,500	348
<i>Caenorhabditis elegans</i> (nematode)	100	20,100	200
<i>Arabidopsis thaliana</i> (mustard family plant)	120	27,000	225
<i>Drosophila melanogaster</i> (fruit fly)	165	14,000	85
<i>Daphnia pulex</i> (water flea)	200	31,000	155
<i>Zea mays</i> (corn)	2,300	32,000	14
<i>Ailuropoda melanoleuca</i> (giant panda)	2,400	21,000	9
<i>Homo sapiens</i> (human)	3,000	21,300	7
<i>Paris japonica</i> (Japanese canopy plant)	149,000	ND	ND

*Some values given here are likely to be revised as genome analysis continues. Mb = million base pairs; the haploid number is used because it represents a complete set of genetic information. ND = not determined.

EUKARYOTEN VS. PROKARYOTEN

- Wat hebben eukaryoten van prokaryoten niet hebben?
- Noncoding DNA tussen genen:
 - Pseudogenen: genen die (teveel) mutaties hebben gekregen en niet meer functioneel zijn
 - Repetierend DNA: meerdere kopieën.
 - Hoge conservatie van deze noncoding sequenties suggereert dat ze wel een belangrijke functie hebben.

EUKARYOTEN VS. PROKARYOTEN

- Sequenzen van menselijk genoom liet zien dat 98.5% niet voor eiwitten, rRNA or tRNA codeert.
- Gen regulatie sequenties (5%) en intronen (20%).

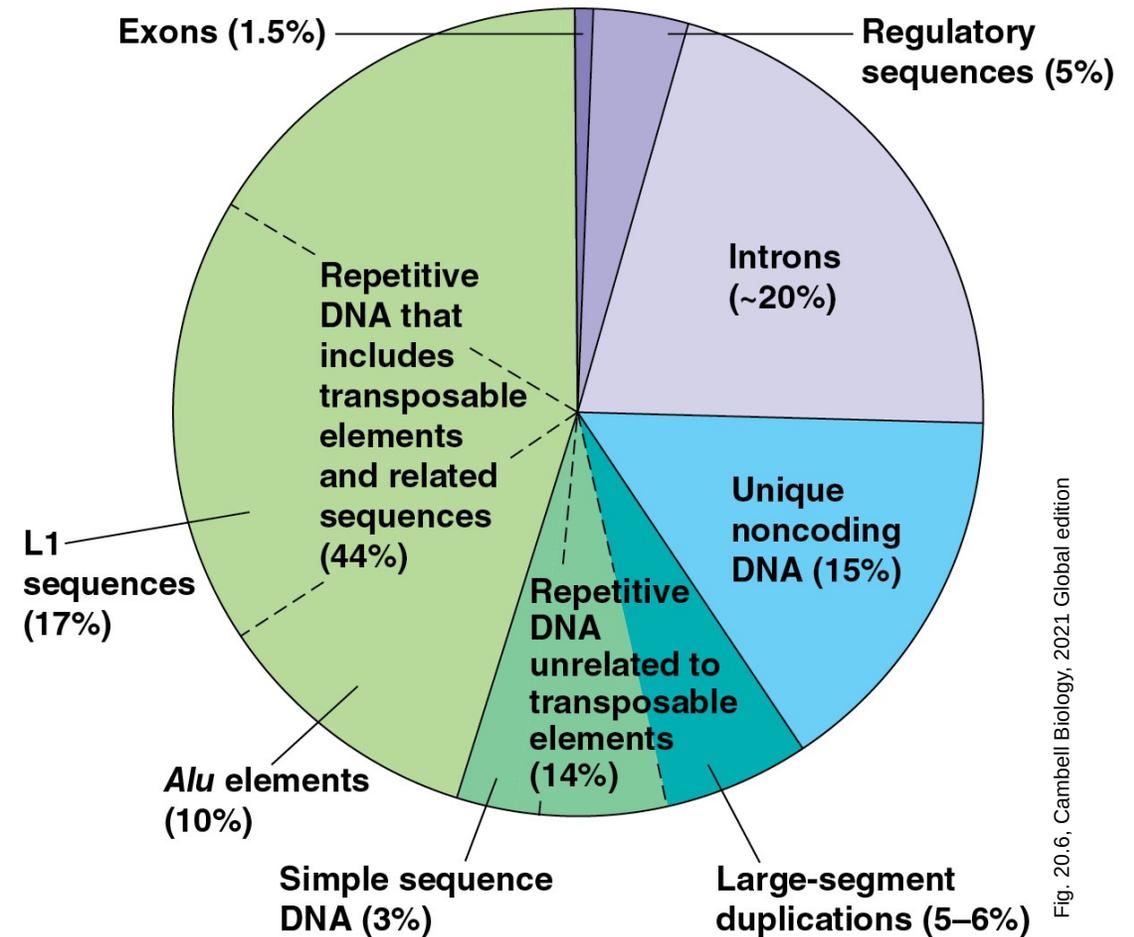


Fig. 20.6, Cambell Biology, 2021 Global edition

EUKARYOTEN VS. PROKARYOTEN

- Heel groot deel bestaat uit transposable elementen en gerelateerd sequenties → ook springende genen genoemd.
Transposonen en retrotransposonen.

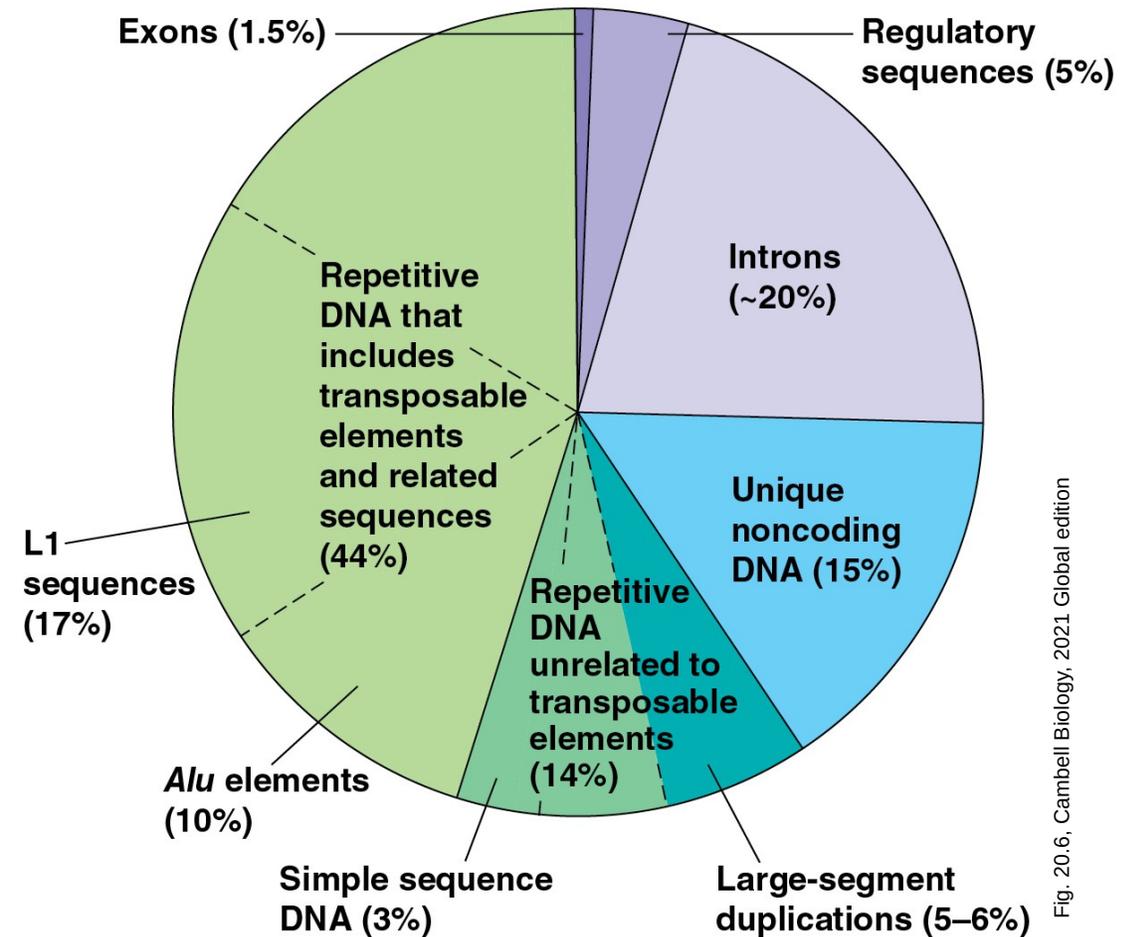
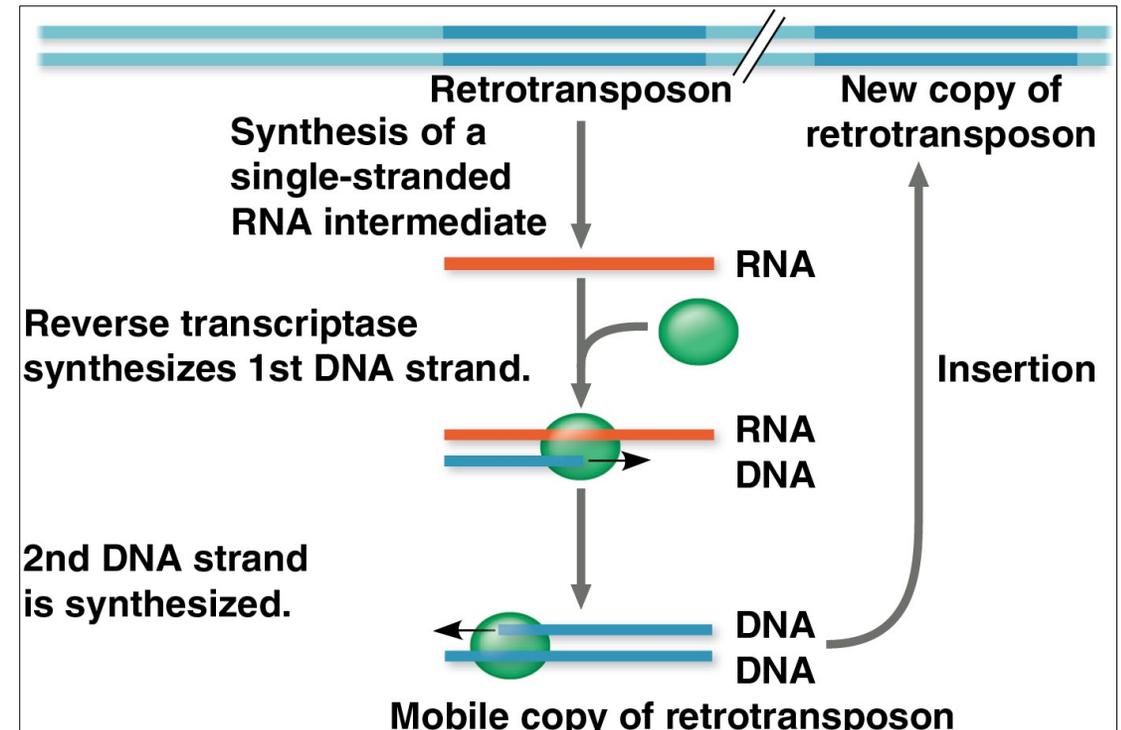
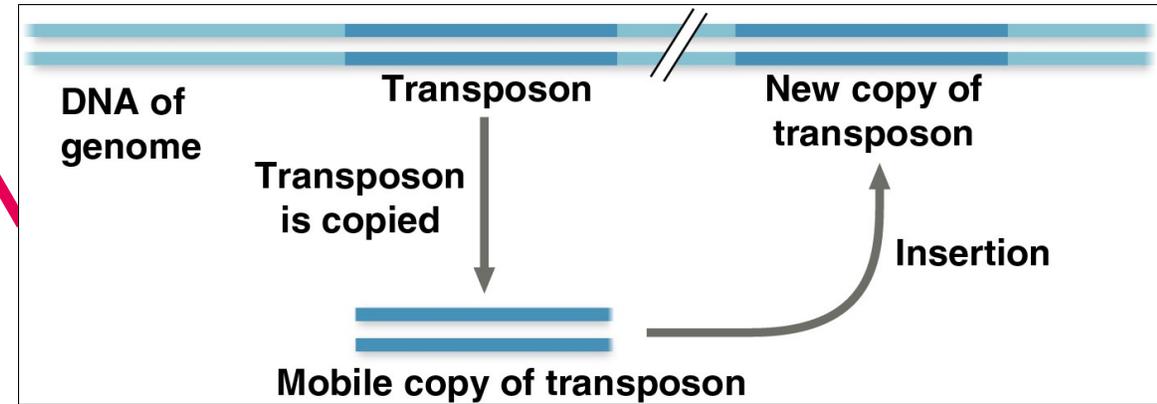


Fig. 20.6, Cambell Biology, 2021 Global edition

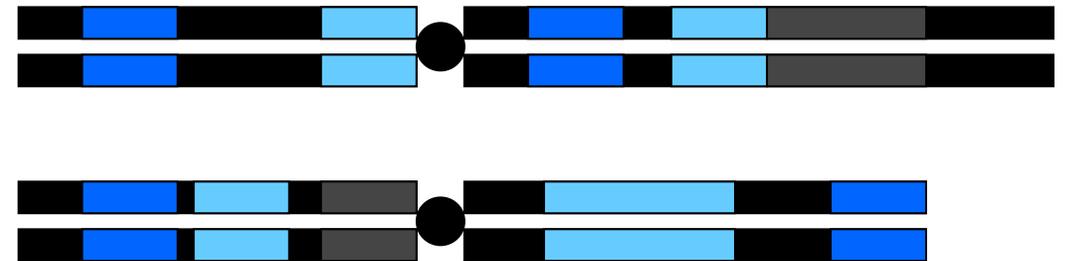
TRANSPOSABLE ELEMENTEN

- Transposon
Hebben een cut-and-paste of copy-and-paste mechanisme
Bewegen door een DNA intermediaire en hebben een enzym transposase nodig
- Retrotransposon (meest dominante vorm in eukaryoten)
Bewegen door een RNA intermediaire en gebruiken hierbij een reverse transcriptase.
Retrovirussen geëvolueerd uit dit systeem of andersom!



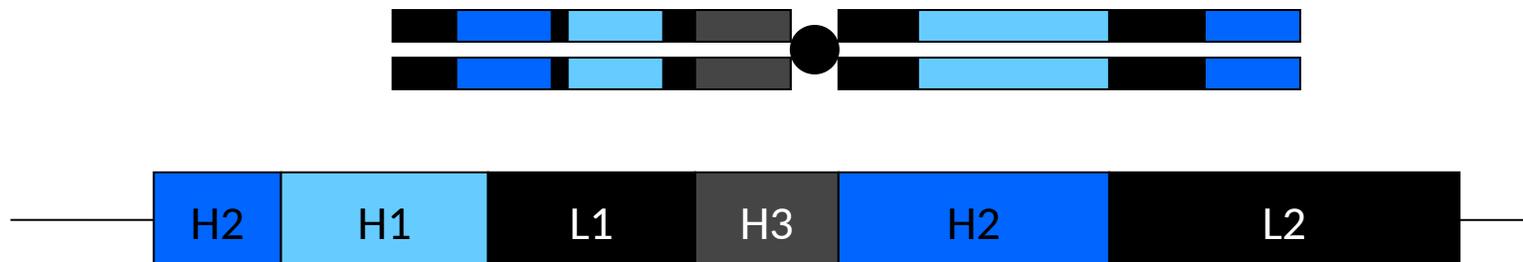
WAT WE OBSERVEREN

- GC% van de DNA sequentie varieert:
 - Tussen genomen (organismen)
 - Coderende sequenties
- Genen hogere GC ratio
- Op het genoom
- Mozaïek achtige formatie: regio's op het genoom van GC-rijke en niet rijke regio's die we isochores noemen

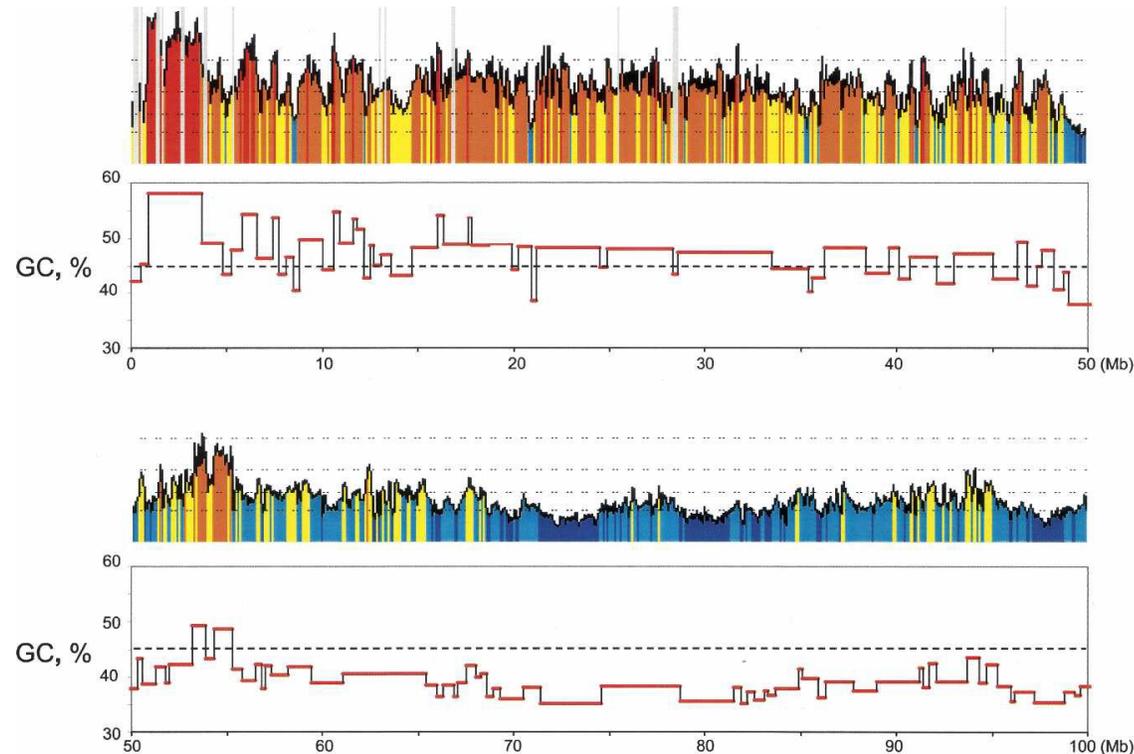


ISOCHORES = LONG (>300 KB) REGIONS OF HOMOGENEOUS BASE COMPOSITION

Genomic core	H1, H2 and H3 (47%, 52% >52%)
“Empty” space	L1 and L2 (<40%)



WAT ZIJN "ISOCHOORES"?



>300 kb regionen in
het genoom met
een stabiel GC%

Discovered chemically, (before
sequencing) using CsCl
ultracentrifugation! (Bernardi
1976)

Figure 4. Overview of isochores on 100 Mb of chromosome 1. The isochores identified on the telomeric 100-Mb region of the short arm of chromosome 1 (as a representative region of human chromosomes) are shown. Broken horizontal guidelines in the *top frames represent GC levels*. Horizontal red stretches in the *bottom frames represent isochores*. Our strategy was to identify boundaries on the basis of GC jumps between adjacent isochores (GC). This strategy may occasionally lead to border misassignments, for instance, in the small number of cases (<3%) when GC is lower than 1% (see Fig. 3). Several of the latter borders could be assigned, however, based on differences in the standard deviations of GC.

Constantini M., Clay O. and Auletta F. (2006) An isochore map of human chromosomes. *Genome Research* 16: 536-541.

ISOCHORES EN GC-RATIO'S

- GC-rijke isochores hebben vaak veel eiwit coderende genen.
- GC ratios van deze isochores berekenen helpt bij het vinden/mappen van gen-rijke regio's op het genoom.

GC PERCENTAGE EN RATIO'S (BEREKENEN)

GC content percentage: $\frac{G + C}{A + T + G + C} \times 100$

GC ratio: $\frac{A + T}{G + C}$

HUMAN ISOCHORES

L1 en L2 arm in GC (~40%):

gen arm

85% van weefsel specifieke genen

H1, H2 en H3 rijk in GC (~50%):

Gen rijk

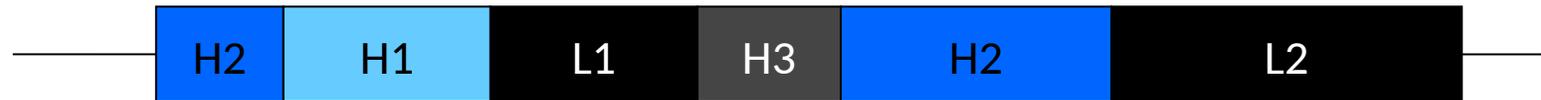
80% huishoudgenen

Promoter dichtbij transcriptie

Kortere intronen en genen

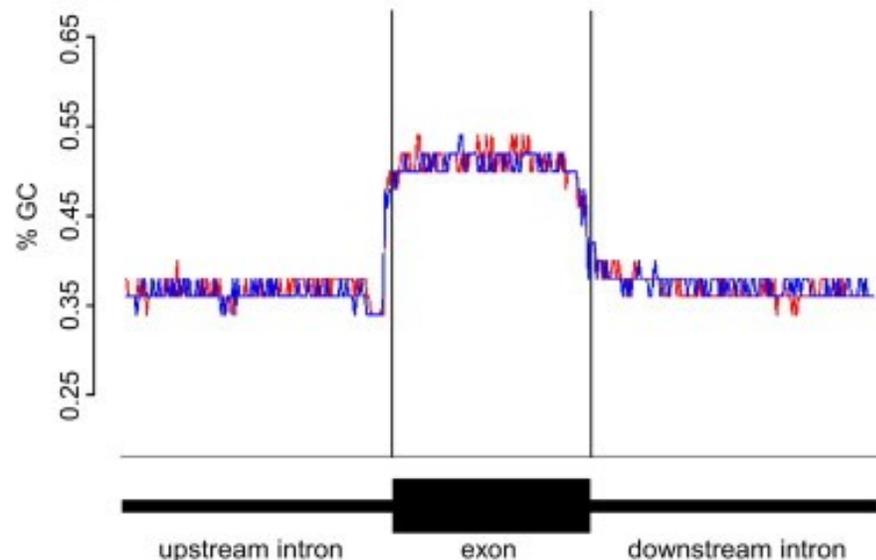
Andere codon bias (More G+C)

Less long repeated sequences (LINEs)



VRAAG

- Waarom zullen eiwit coderende genen GC-rijk zijn?
- Waarom krijg je codon bias met een andere GC-content?
- Leg het plaatje hieronder uit:



[Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition - ScienceDirect](#)

SAMENVATTING

- We kunnen genoom/gen data krijgen via direct genoom sequencing, metagenomen of transcriptomen.
- GC percentages/ratios/content kan ons informatie geven over organismen en eiwit coderende regio's.
- Genoom grootte en aantal genen en fenotype niet altijd gerelateerd.

BRONVERMELDING

Campbell biology, 8th edition, Pearson, 2008 ISBN: 321-53616-9

Constantini M., Clay O. and Auletta F. (2006) An isochore map of human chromosomes. *Genome Research* 16: 536-541.

Antonarakis, S. E. (2010) Human Genome Sequence and Variation. Chapter 2 from M.R. Speicher et al. (eds.), Vogel and Motulsky's Human Genetics: Problems and Approaches, DOI 10.1007/978-3-540-37654-5_2, © Springer-Verlag Berlin Heidelberg 2010