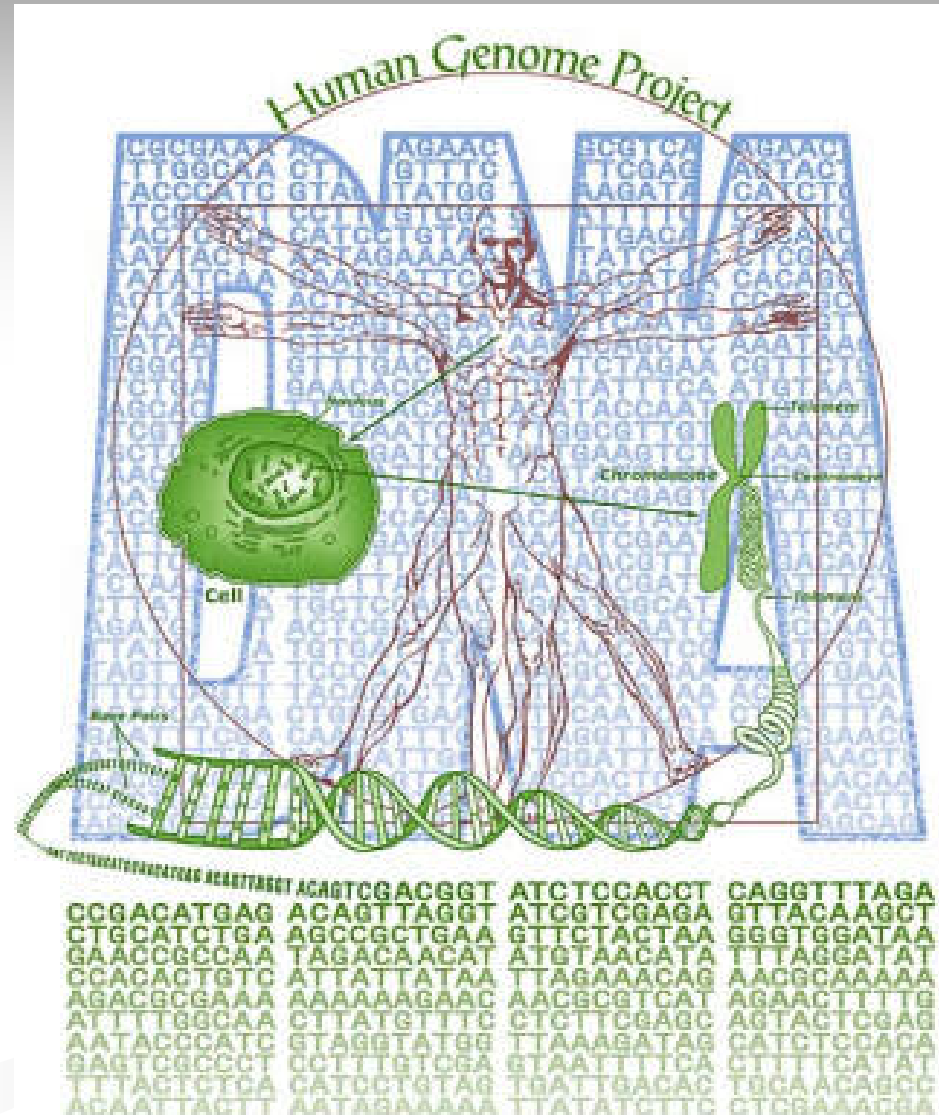


Projekt ľudského genómu

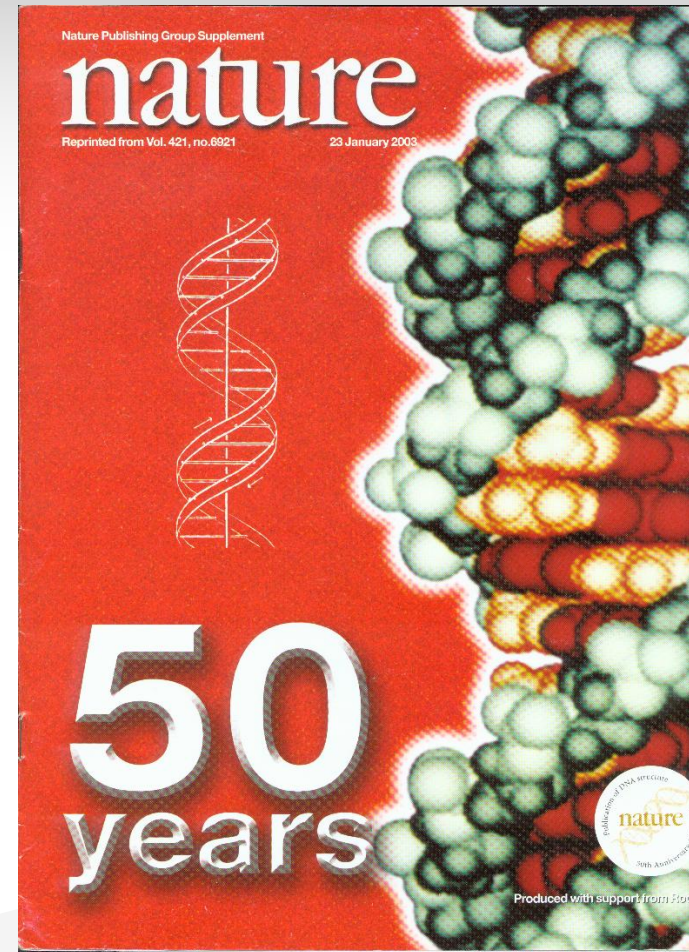
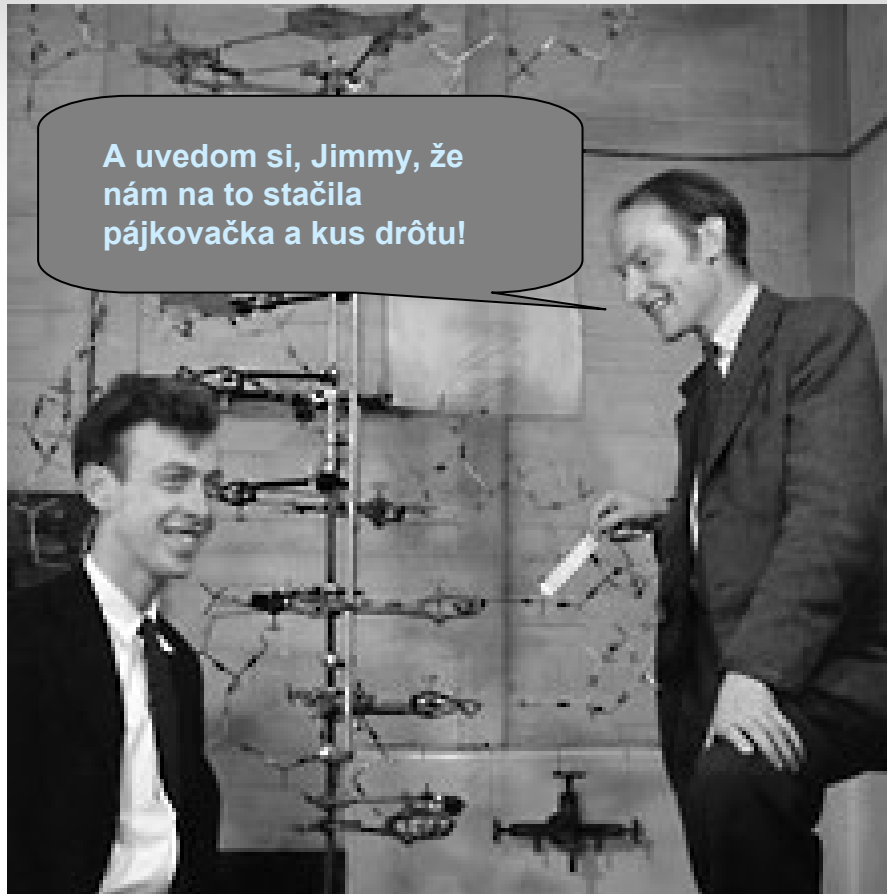
- Prehistória
- História
- Ciele
- Metódy
- Priebeh
- Výsledky
- Využitie
- Perspektívy

základná www stránka:

www.doegenomes.org



Prehistória: Watson a Crick 1953



A STRUCTURE FOR DEOXYRIBOSE NUCLEIC ACID

J.D.Watson

F.H.C.Crick

Nature, April 25, 1953

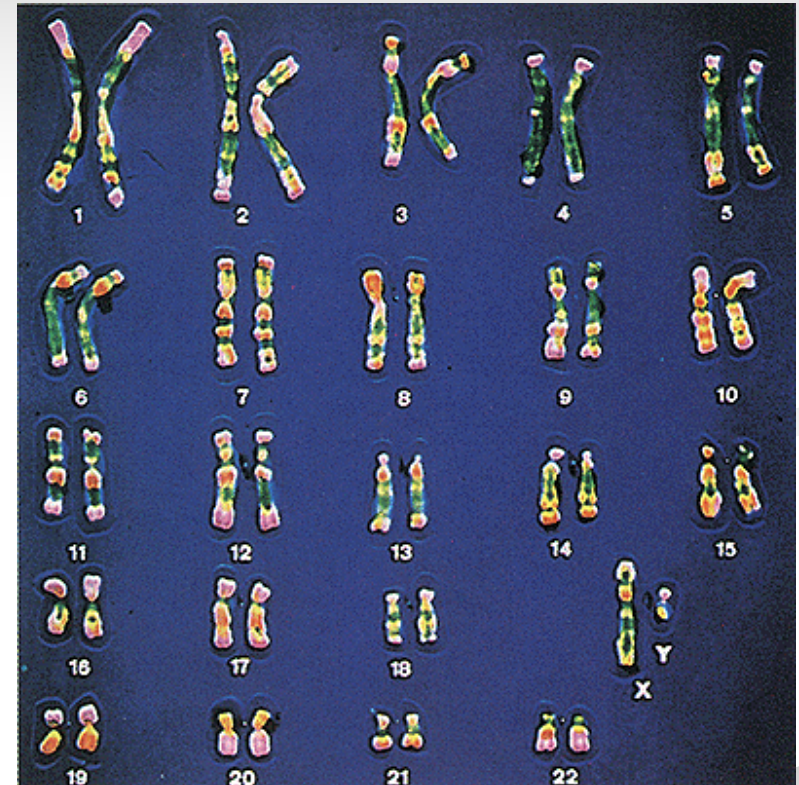
„It has not escaped our notice that
the specific pairing we have
postulated immediately suggests
a possible copying mechanism for
the **genetic material**“

Genetika človeka v r. 1953

- hodnotenie rodokmeňov
- štúdium dvojčiat
- cca 50 monogénných ochorení
- 5 monogénných znakov (ABO, Rh...)
- 48 chromozómov
- 5 miliónov génov
- žiadna DNA (podľa učebníc)

60-te roky: cytogenetika človeka

- 1956: 46 chromozómov
- 1959: chromozómové aberácie a choroby
- 1960: nomenklatúra a metodika
- chromozómy a nádory
- chromozómy: „orgán“ lekárskej genetiky
- ľudských génov jeden milión, o DNA stále ani zmienka...



70-te roky: žeby DNA?

- polymorfizmy (sérové skupiny, izoenzýmy...)
- väzba a chrom. lokalizácia
- bunkové hybridy
- biochemická genetika

Ale už aj:

- **prvé RFLP** v ľudskej DNA (1978)
- **DNA diagnostika** chorôb (1979)
- **100 000 génov**

Human Genome Project: história

- 1986: Santa Fe - DoE Intl. Meeting
- **James Watson**: „vystúpiť na cestu od dvojitej závitnice k 3 miliardám schodov ľudského genómu“
- 1988: Kongres USA schválil 15 ročný projekt a dotáciu 3 mld USD
- **1990: začiatok projektu**
- 2005: predpokladané ukončenie

Koordinácia:

1988: **HUGO** (Human Genome Organization)
HUGO Europe – Americas - Pacific

HGP: rozsah

- $3\,200 \times 10^6$ bp (dĺžka Dunaja v mm !)
- dovtedy najdlhší sekvenovaný úsek: ľudská mtDNA (16 500 bp, 1981)
- vtedajšia rýchlosť: 100 bp / osoba / deň
- cena: 5 - 10 USD/bp

Oficiálny začiatok HGP: 1990

- 20 pracovísk zo 6 štátov
- 3 mld USD
- cca 15 rokov práce (t.j. do r. 2005)

História sekvenovania DNA

bp/osoba/rok	rok	udalost'
1	1965	<i>Holley</i> sekv. alanyl tRNA <i>S. cerev.</i>
15	1970	<i>Wu</i> sekvenoval 12 bp lambda fága
150	1977	<i>Sanger</i> dideoxy, <i>Gilbert</i> chem. sekv.
1 500	1978	<i>Sanger</i> sekv. phi-x174 (5375 bp)
15 000	1980	<i>Messing</i> vyvinul M13 vektor
25 000	1986	<i>Hood</i> – čiastočne automat. sekv.
1,000 000	1995	<i>Venter</i> – autom. fluoresc. sekv.
150, 000 000	1999	Perkin-Elmer 96 kapilárny sekv.
? 000, 000 000	dnes	každý kto má vybavenie

HGP: ciele

- Triangulovať genóm pomocou DNA markerov
- Identifikovať a mapovať gény, určiť ich štruktúru a funkciu v zdraví aj v patológii
- Identifikovať dôležité mimogénové sekvencie
- Určiť úplnú sekvenciu genómu (3,2 Gb)
- Všetky dáta uložiť vo verejne prístupných databázach
- Sekvenovať genómy modelových organizmov (drozofila, myš, *C. elegans*, *A. thaliana* a i.)
- Skúmať etické, legálne a sociálne aspekty

Metódy mapovania

Genetické mapovanie

- rodokmeňová analýza (rekombinácia)
- triangulácia pomocou markerov

Fyzické mapovanie

- **nízke rozlíšenie:**

- hybridizácia somat. buniek
- in situ hybridizácia (FISH)

vysoké rozlíšenie:

restrikčné mapovanie
STS mapovanie
EST mapovanie
sekvenovanie

Triangulácia ľudského genómu

Genetické markery: história

1900	krvné skupiny	cca 20
1960	sérové proteíny a izoenzýmy	10^2
1980	DNA RFLP	10^3
1985	minisatelity	10^4
1990	mikrosatelity	10^5
1995	SNP (single nt polymorphisms)	10^6

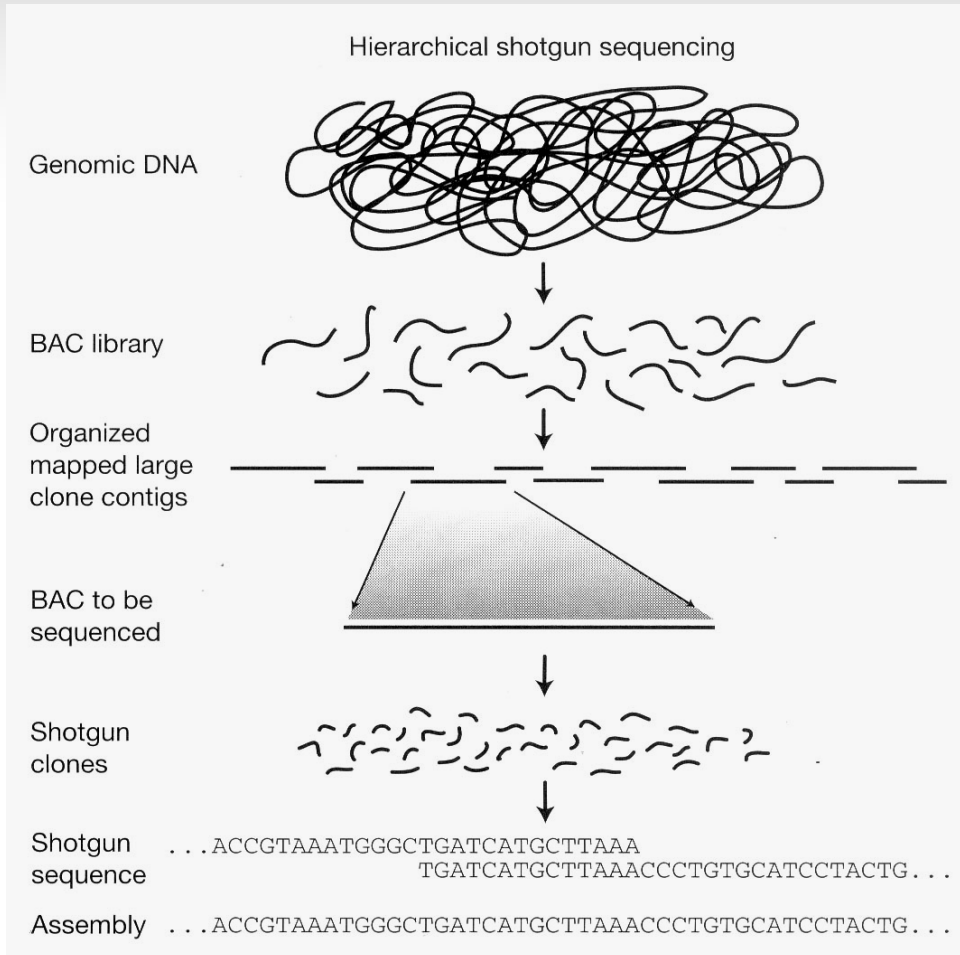
Triangulácia genómu: genetické mapy

1987	RFLP	402	nad 10 cM
1992	STR	813	cca 5 cM
1996	STR	5262	1,6 cM
1997	integrovaná	6800	0,7 cM

podrobnejšia triangulácia nie je potrebná
(v súčasnosti cca 2×10^6 mapovaných SNP)

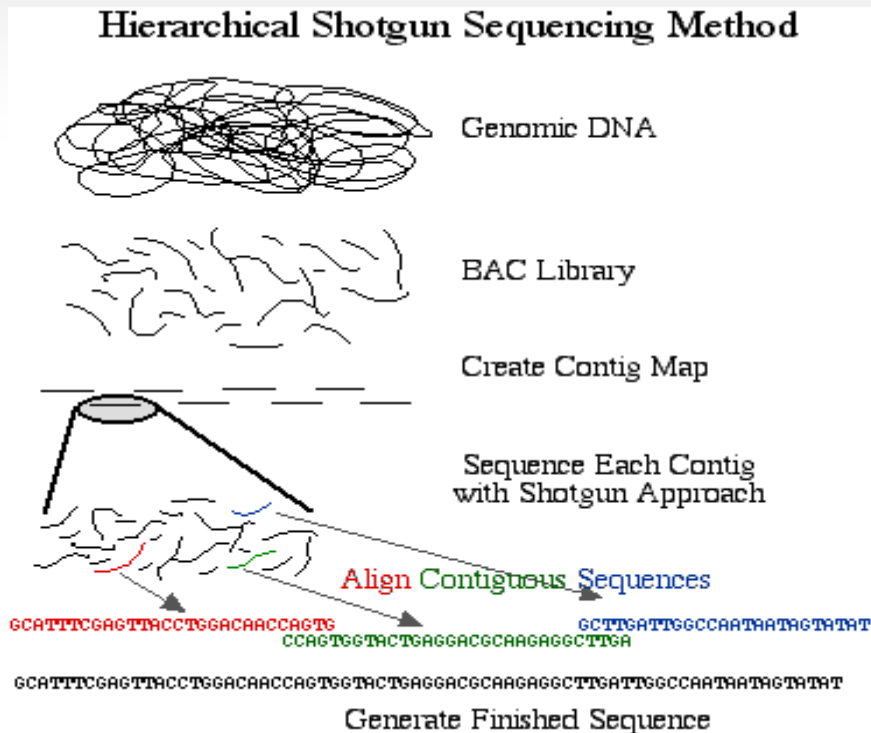
International Human Genome Sequencing Consortium

Hierarchická metóda (clone-by-clone):



- Triangulácia genómu pomocou STR markerov
- 100 až 200 kb dlhé BAC klony, mapovanie týchto klonov do STR mapy
- Zostrojť kontigy
- Štiepiť BACy na náhodné („shotgun“) fragmenty
- Sekvenovať fragmenty
- Zoradiť do sekvencie

HGP: hierarchická stratégia sekvenovania („clone-by-clone“)



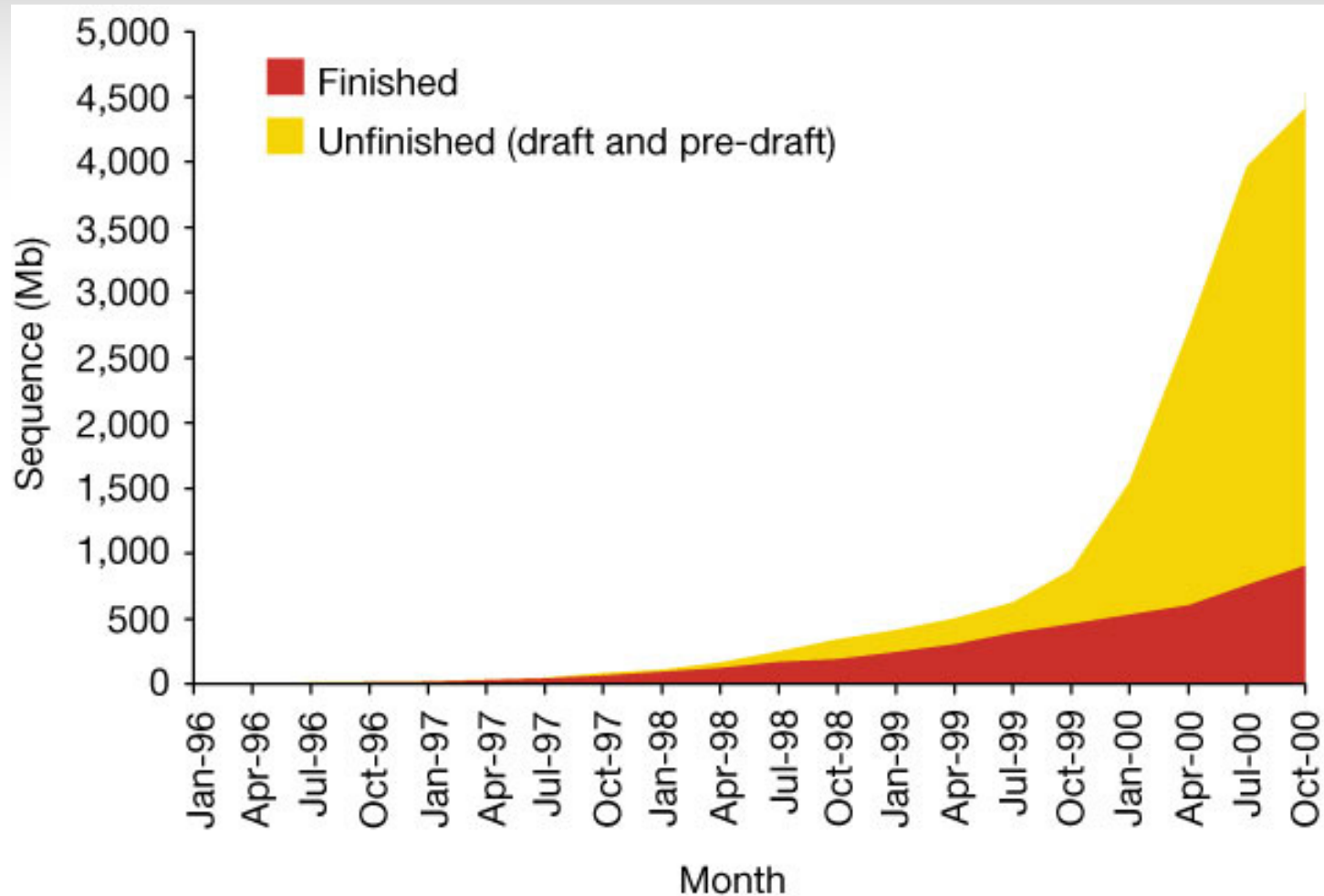
- triangulovať genóm pomocou STR
- zostrojiť BAC-knižnice
- mapovať klony do STR mapy
- utvoriť BAC kontigy
- štiepiť BAC-y náhodne na sekvenovateľné fragmenty
- sekvenovať fragmenty
- zoradiť do konečnej sekvencie klonu
- zoradiť sekvencie klonov do definitívnej sekvencie
- anotovať sekvenciu

HGP: Francis Collins a Eric Lander





Nárast údajov o sekvencii 1996-2001

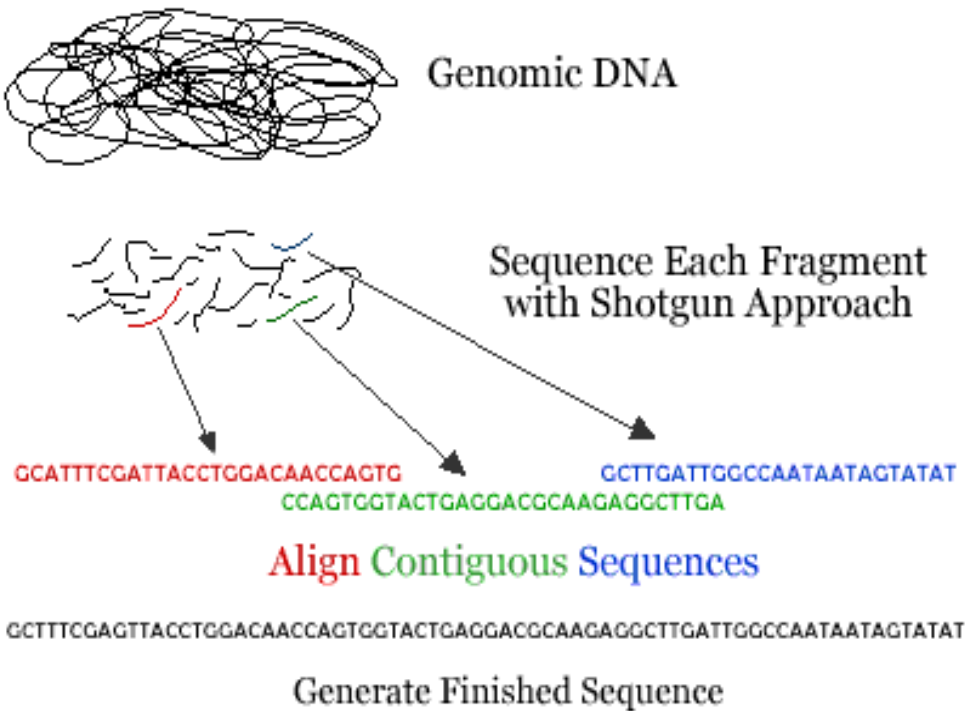


Craig J. Venter a Celera Genomics



Stratégia sekvenovania: Celera Genomics

Whole Genome Shotgun Sequencing Method



- Celogenómová shotgun stratégia sekvenovania
- Skombinovanie s dátami HGP
- Utvorenie knižníc s definovanými dĺžkami (2, 10, 50 kb)

15 February 2001

nature

ISSN 0950-0804 PUBLISHED WEEKLY

www.nature.com

the
human
genome

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

** A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

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,³ Michele Cargill,¹ 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 Fasule
Michael Flanigan,¹ Lilliana Florea,¹ Aaron Halpern,¹ Sridhar Hannenhalli,¹ Saul Kravitz,¹ Samuel Levy,
Clark Mobarry,¹ Knut Reinert,¹ Karin Remington,¹ Jane Abu-Threideh,¹ Ellen Beasley,¹ Kendra Biddick
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
Ping Guan,¹ Thomas J. Heiman,¹ Maureen E. Higgins,¹ Rui-Ru Ji,¹ Zhaoxi Ke,¹ Karen A. Ketchum,¹
Zhongwu Lal,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 Lal 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,
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
Sukye Tse,¹ Claire Vech,¹ Gary Wang,¹ Jeremy Wetter,¹ Sherita Williams,¹ Monica Williams,¹
Sandra Windsor,¹ Emily Winn-Deen,¹ Keriellen Wolfe,¹ Jayshree Zaveri,¹ Karena Zaveri,¹
Josep F. Abril,¹⁴ Roderic Guigó,¹⁴ Michael J. Campbell,¹ Kimmen V. Sjolander,¹ Brian Kartak,¹
Anish Kejariwal,¹ Huaiyu Mi,¹ Betty Lazareva,¹ Thomas Hatton,¹ Apurva Narechania,¹ Karen Diemer,
Anushya Muruganujan,¹ Nan Guo,¹ Shinji Sato,¹ Vineet Bafna,¹ Sorlin 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
Carl Foster,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 Marshall Peterson,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

HGP v Bielom Dome, jún 2001



F. Collins a C. J. Venter 2001



Sekvenované eukaryo genómy

organizmus	rok	Mb	%	% eu	gény	g./Mb
<i>S.cerevisiae</i>	1996	12	93	100	5 800	483
<i>C. elegans</i>	1998	97	99	100	19 099	197
<i>Drozofila</i>	2000	116	64	97	13 601	117
<i>Arabidopsis</i>	2000	115	92	100	25 498	221
H.s. chr. 21	2000	34	75	100	225	7
H.s. chr. 22	1999	34	70	97	545	16
H.s. HGP	2001	2 693	84	90	31 780	12
H.s. Celera	2001	2 654	83	88-93	39 114	15

Prehľad ľudského genómu: Celera

■ Dĺžka sekvencie	2,91 Gb
■ Podiel GC párov	38 %
■ Podiel repetitívnej DNA	35 %
■ Počet anotovaných génov	26 383
■ Celkový počet génov (odhad)	39 114
■ Priemerná dĺžka génu	27 kb
■ Najviac génov v chromozóme	19 (23g./Mb)
■ Najmenej génov v chromozóme	13 a Y (5/Mb)
■ Gény tvoria	25,5-37,8 %
■ Exóny tvoria	1,1 – 1,4 %
■ Výskyt SNP	1 : 1250 bp

Nature 431, 931 - 945 (21 October 2004); doi:10.1038/nature03001

Finishing the euchromatic sequence of the human genome

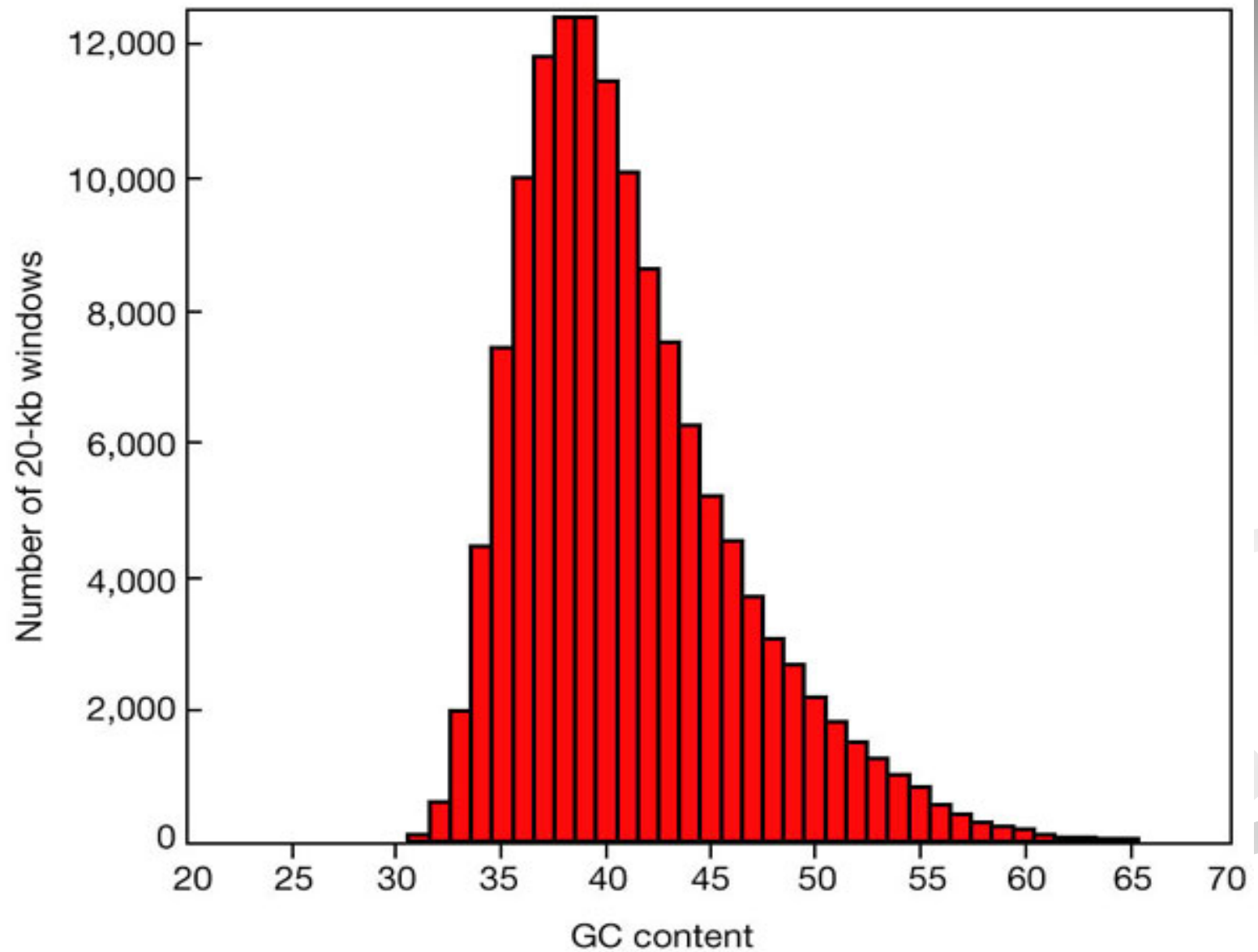
INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM

http://vega.sanger.ac.uk/Homo_sapiens/
www.ensembl.org

Dokončená sekvencia (2004): základné údaje







- pokrýva 2,85 Gb; **celý genóm má 3,08 Gb**
- pokrýva > 99% euchromat. genómu
- sekvenačné chyby zriedkavejšie ako 1 : 100 000, priemerné pokrytie 11-násobné
- len <350 medzier (väčšinou heterochrom. úseky a vysoko repetitívne sekvencie)
- **20 000 – 25 000 proteíny kódujúcich génov** (22 287)
- priemerne 9 génov na Mb; veľká variabilita
- 232 000 exónov; v priemere 10,4 exónu na gén; **exóny tvoria len 1,2% dĺžky sekvencie**
- identifikovaných **cca 20 000 pseudogénov**, asi ich je viac ako génov
- segmentové duplikácie tvoria 5,3% euchromatickej sekvencie („evolutionary gene nurseries“)
- ale pozor: anotácia sekvencie stále ešte nie je dokončená!

Obsah GC párov v ľudskom genóme: 20 Mb okná



Zastúpenie rozličných typov rozptýlených repetícií

Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous		6–8 kb	850,000	21%
	Non-autonomous		100–300 bp		
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Non-autonomous		1.5–3 kb		
DNA transposon fossils	Autonomous		2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

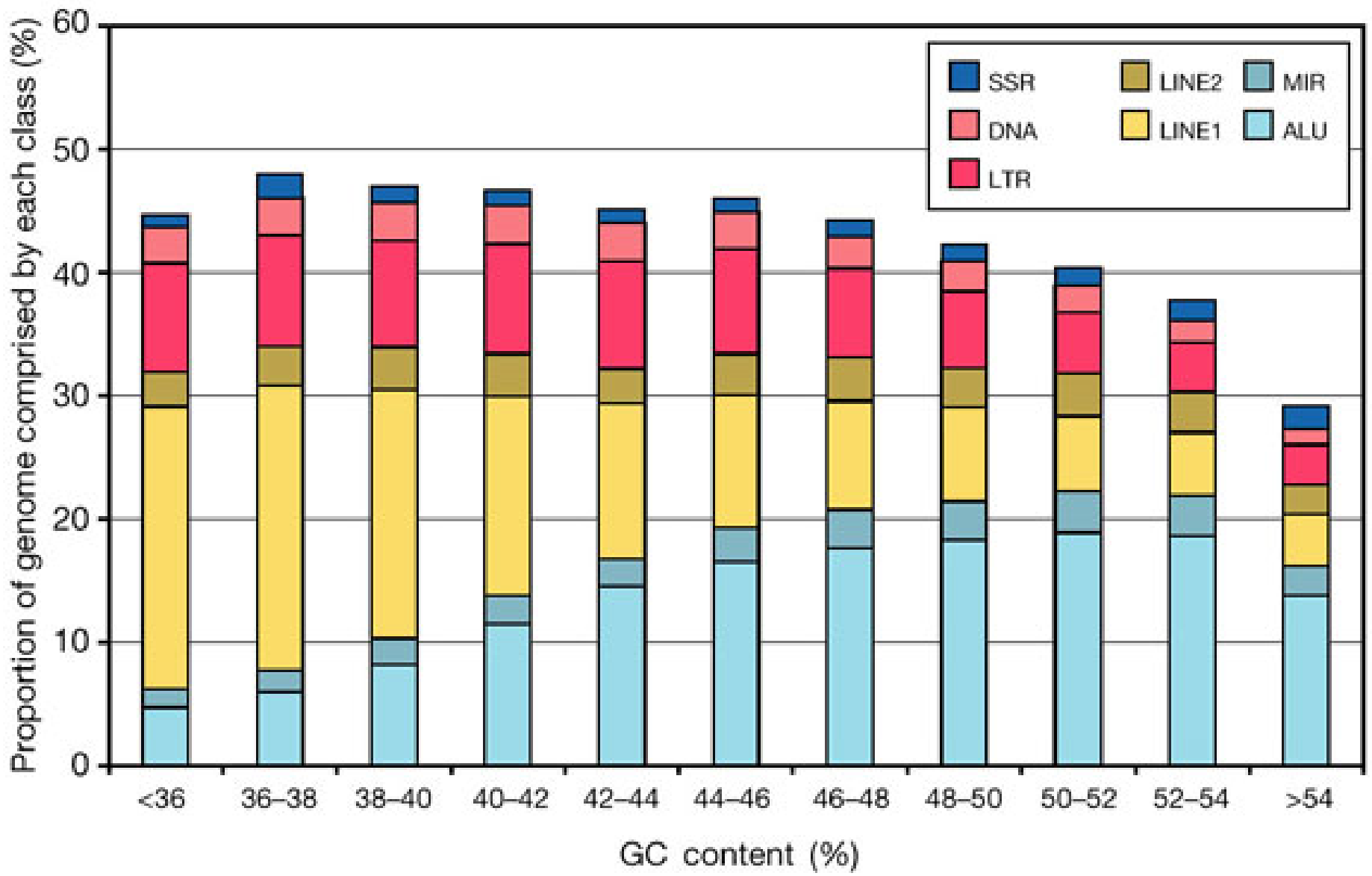
Triedy rozptýlených repetícií

Table 11 Number of copies and fraction of genome for classes of interspersed repeat

	Number of copies (× 1,000)	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)	Number of families (subfamilies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1 (~20)
MIR	393	60.1	2.20	1 (1)
MIR3	75	9.3	0.34	1 (1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	1 (~55)
LINE2	315	88.2	3.22	1 (2)
LINE3	37	8.4	0.31	1 (2)
LTR elements	443	227.0	8.29	
ERV-class I	112	79.2	2.89	72 (132)
ERV(K)-class II	8	8.5	0.31	10 (20)
ERV (L)-class III	83	39.5	1.44	21 (42)
MaLR	240	99.8	3.65	1 (31)
DNA elements	294	77.6	2.84	
hAT group				
MER1-Charlie	182	38.1	1.39	25 (50)
Zaphod	13	4.3	0.16	4 (10)
Tc-1 group				
MER2-Tigger	57	28.0	1.02	12 (28)
Tc2	4	0.9	0.03	1 (5)
Mariner	14	2.6	0.10	4 (5)
PiggyBac-like	2	0.5	0.02	10 (20)
Unclassified	22	3.2	0.12	7 (7)
Unclassified	3	3.8	0.14	3 (4)
Total interspersed repeats		1,226.8	44.83	

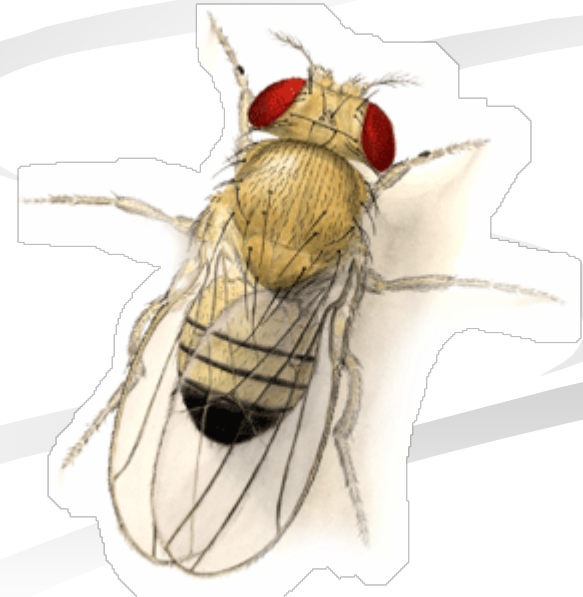
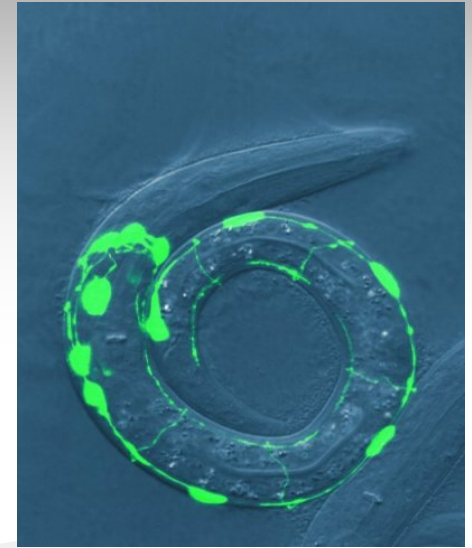
.....
 The number of copies and base pair contributions of the major classes and subclasses of

Hustota hlavných tried repetícií ako funkcia obsahu GC párov



Počet génov u rozličných organizmov

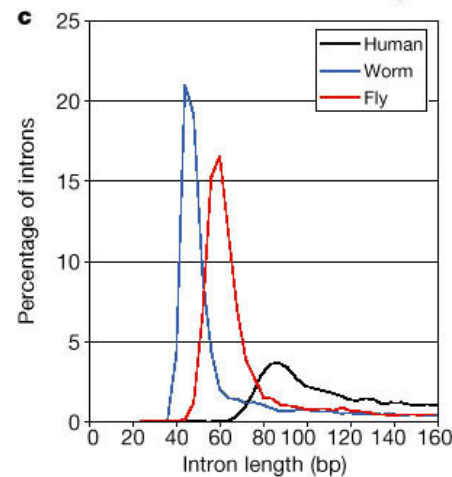
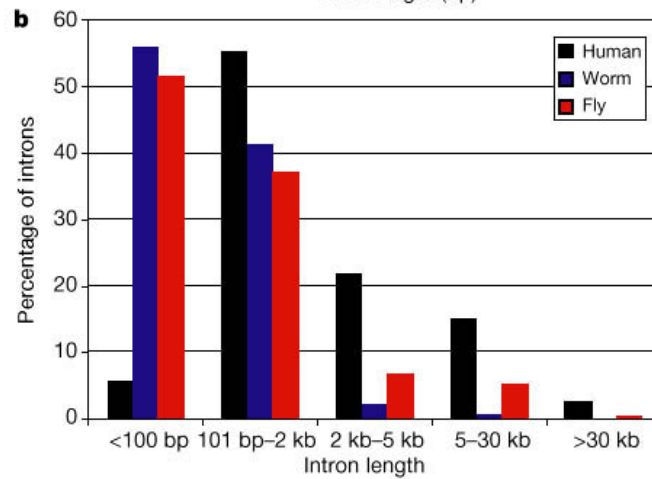
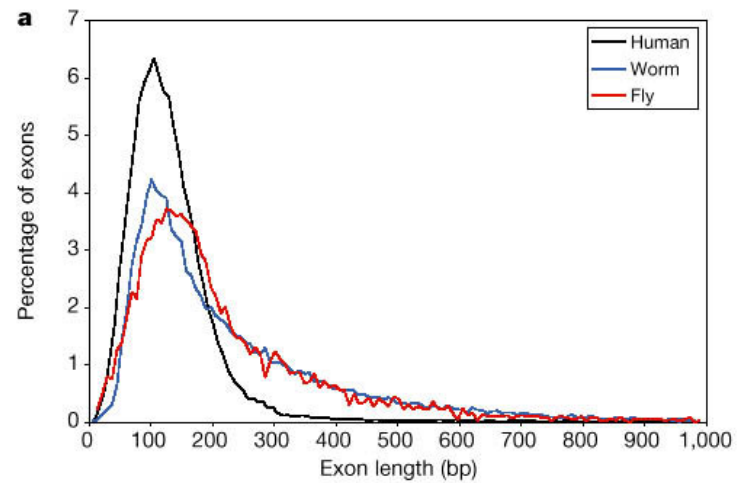
- **Človek**
okolo 25 000
- ***Caenorhabditis elegans***
okolo 20 000
- ***Drosophila melanogaster***
14 000
- ***E. coli***
okolo 6000



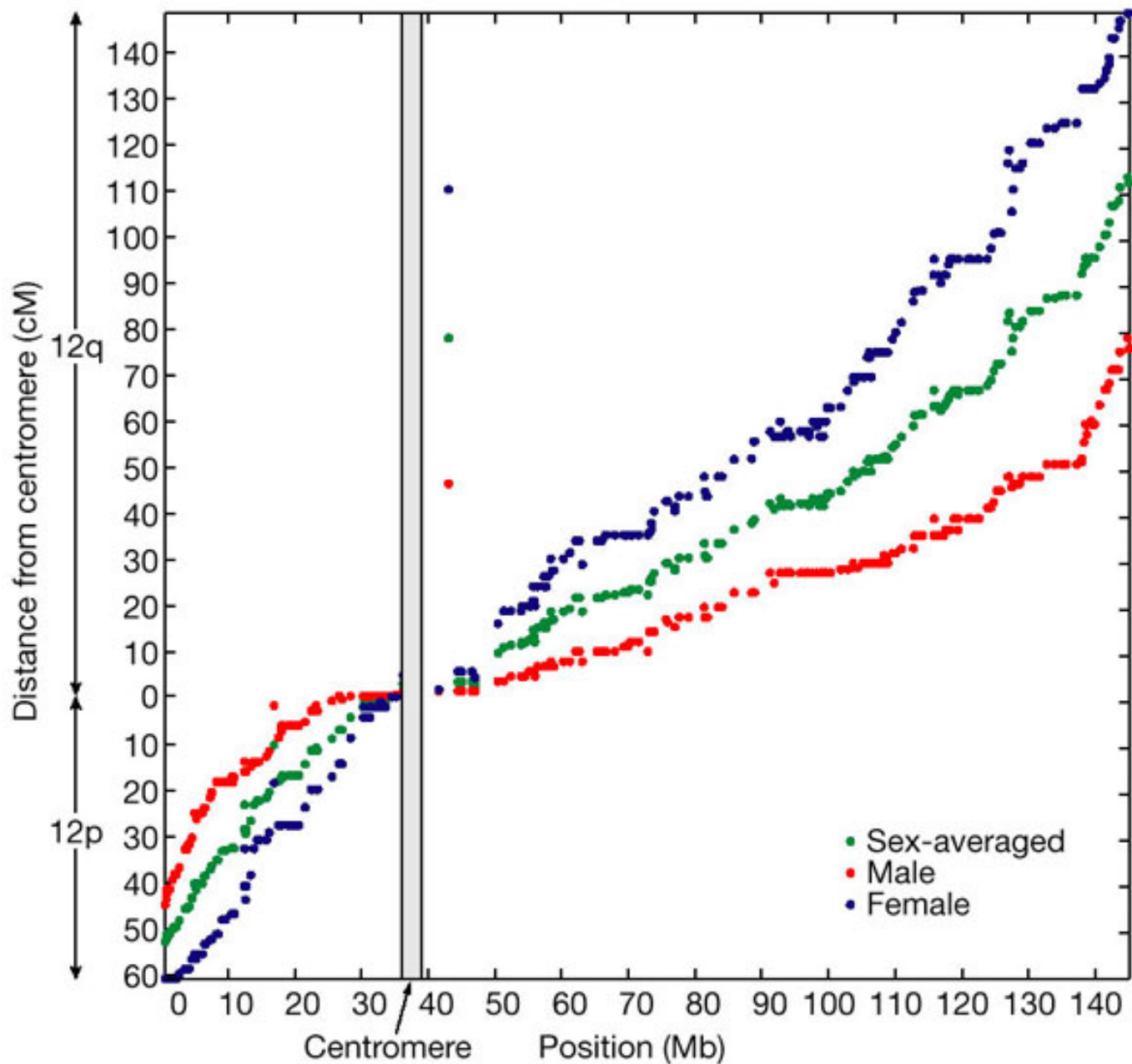
Dížková distribúcia (človek, drozofila, C. e.):

exónov:

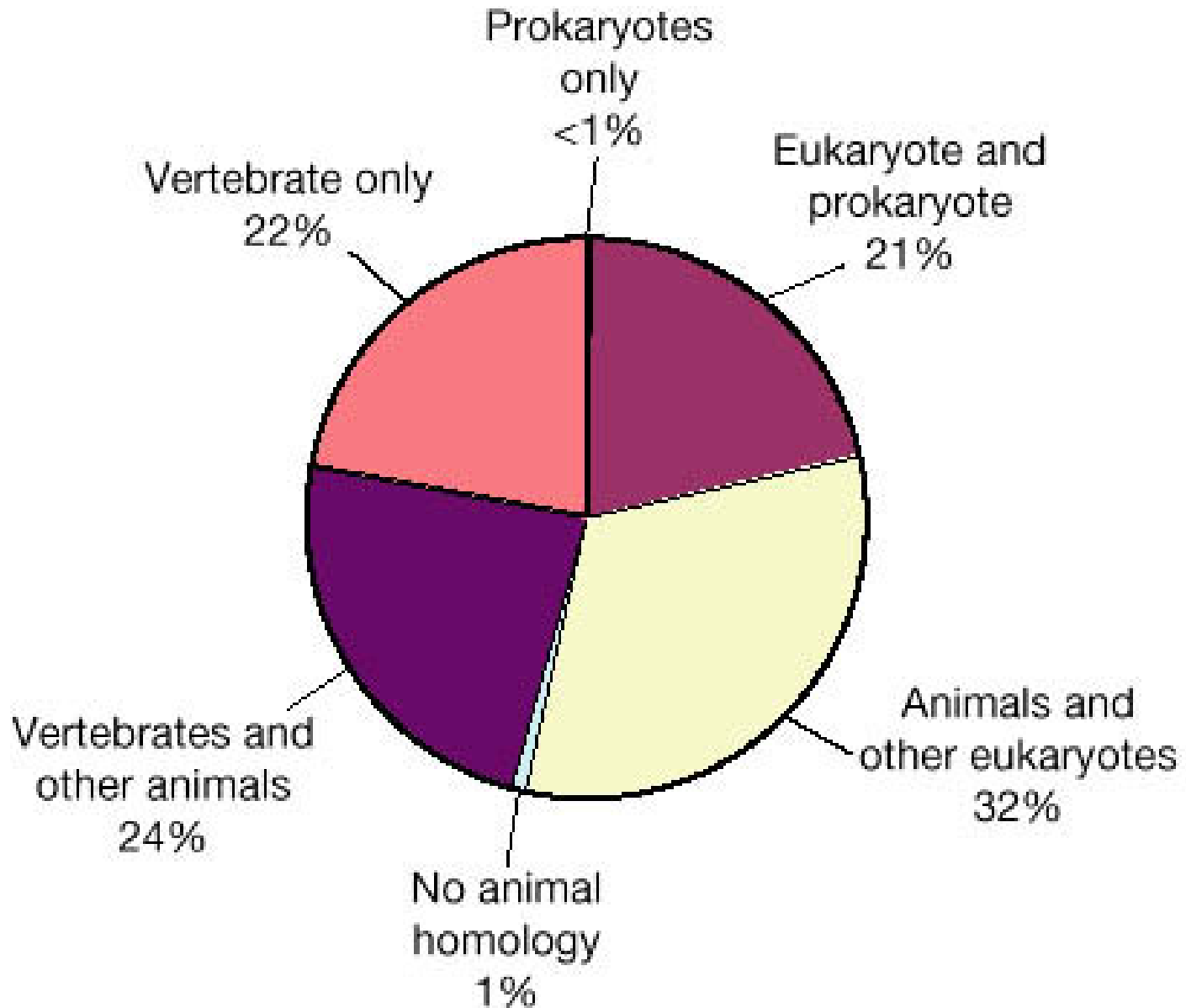
intrónov:



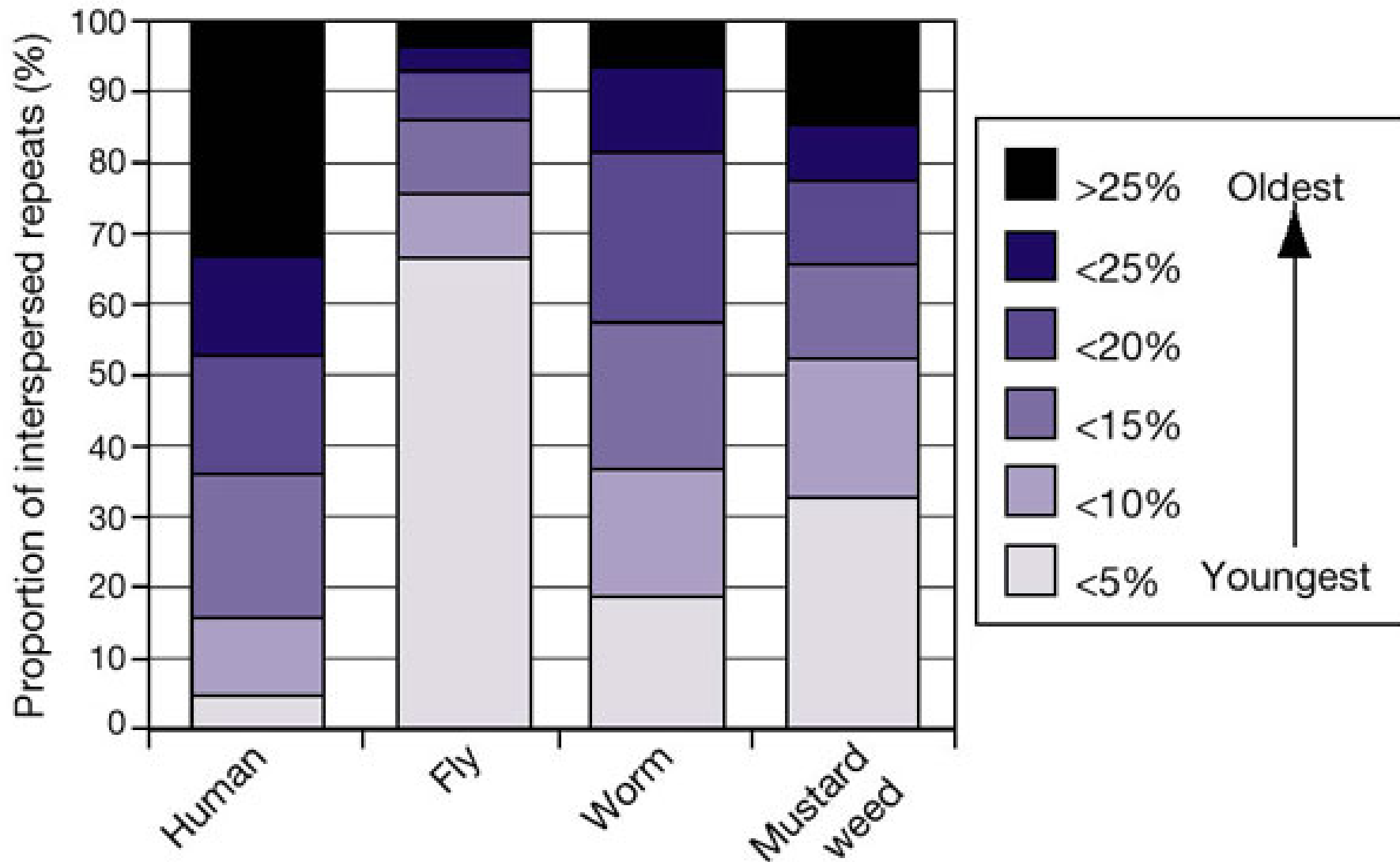
Chromozóm 12: Mb / cM



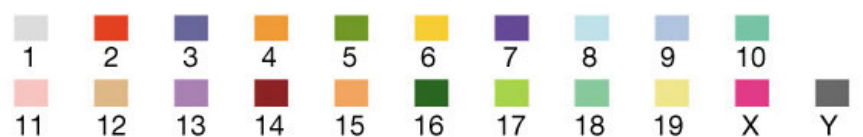
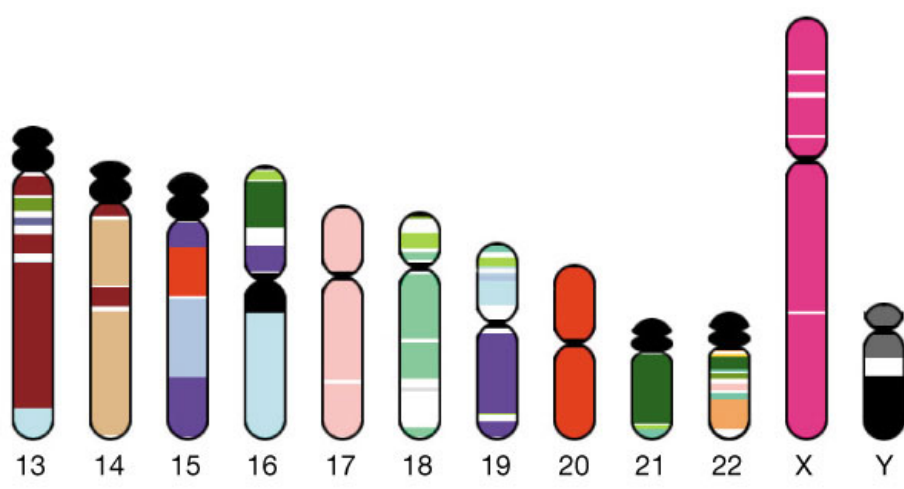
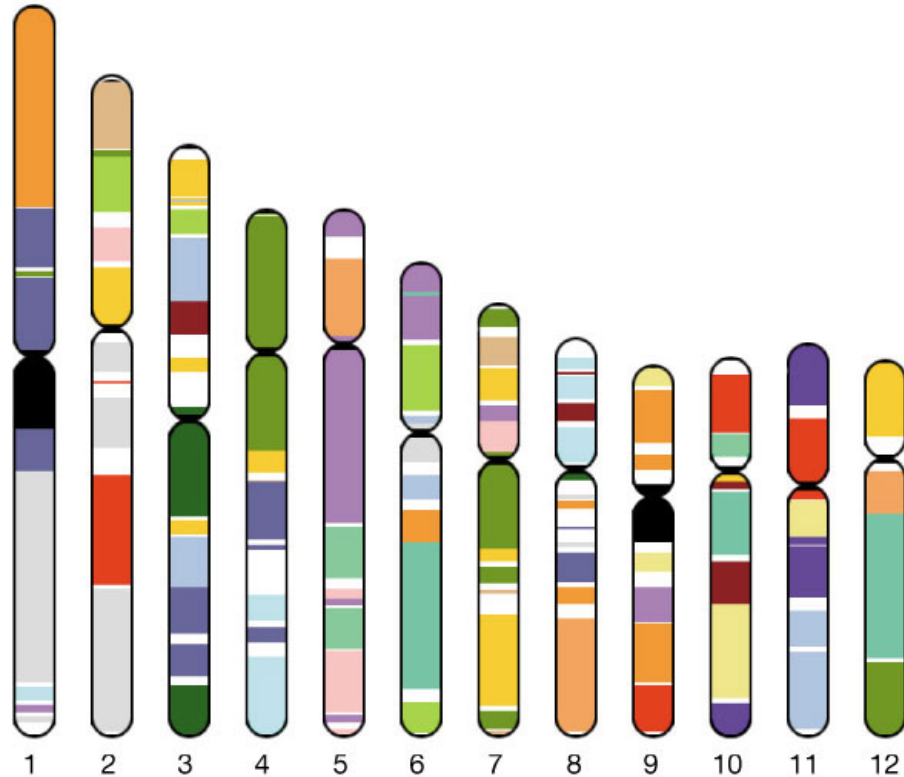
Distribúcia homológov ľudských proteínov



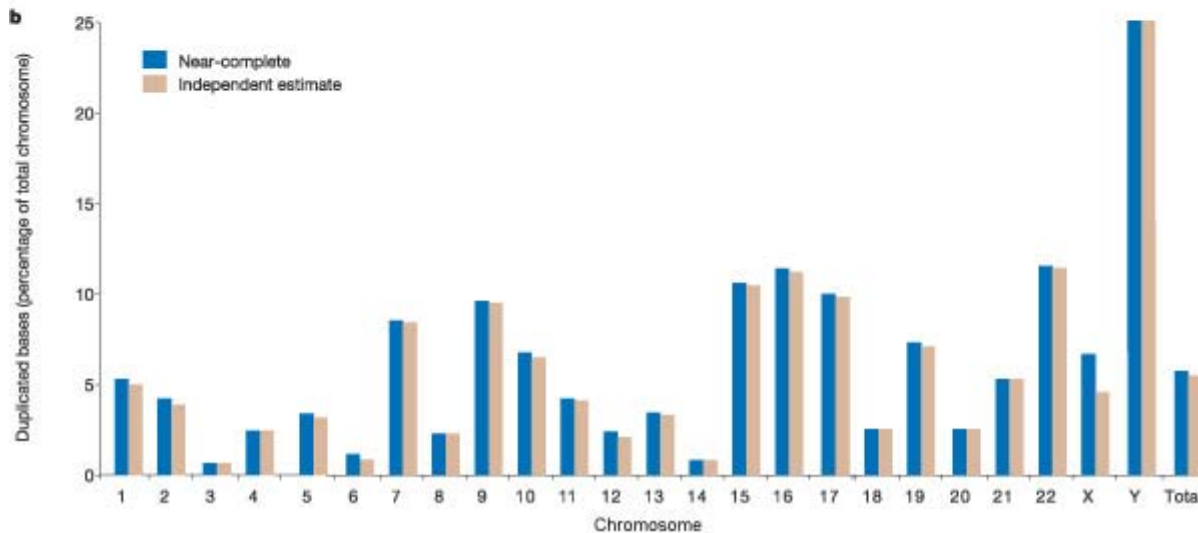
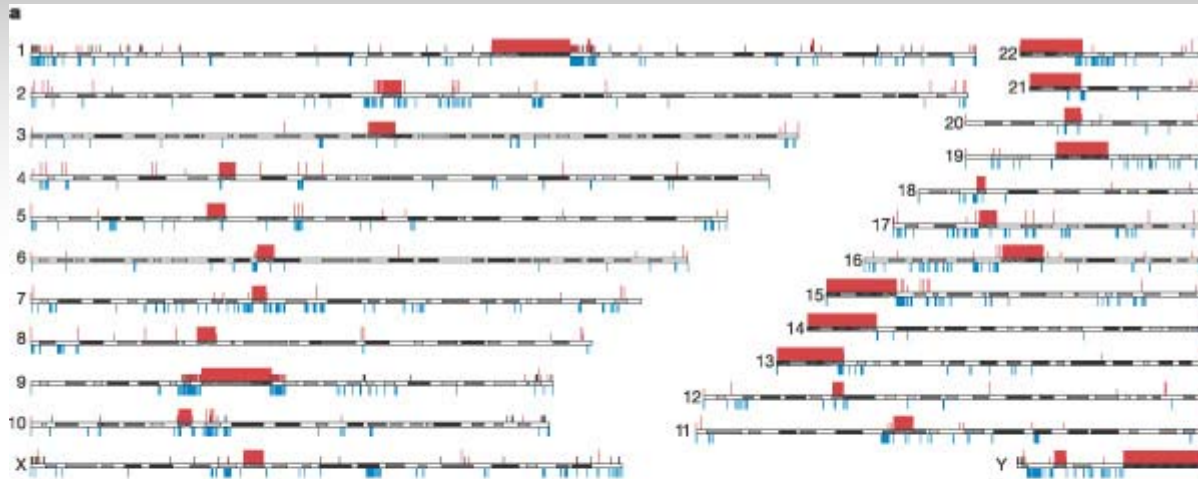
Evolučný vek rozptýlených repetícií v eukaryo- genómoch



Konzervované segmenty ľudského a myšieho genómu



Segmentové duplikácie v ľudskom genóme



Pozične
klonované
gény s
využitím HGP
sekvencie
(kandidátne
gény)

Table 26 Disease genes positionally cloned using the draft genome sequence

Locus	Disorder	Reference(s)
<i>BRCA2</i>	Breast cancer susceptibility	55
<i>AIRE</i>	Autoimmune polyglandular syndrome type 1 (APS1 or APECED)	389
<i>PEX1</i>	Peroxisome biogenesis disorder	390, 391
<i>PDS</i>	Pendred syndrome	392
<i>XLP</i>	X-linked lymphoproliferative disease	393
<i>DFNA5</i>	Nonsyndromic deafness	394
<i>ATP2A2</i>	Darier's disease	395
<i>SEDL</i>	X-linked spondyloepiphyseal dysplasia tarda	396
<i>WISP3</i>	Progressive pseudorheumatoid dysplasia	397
<i>CCM1</i>	Cerebral cavernous malformations	398, 399
<i>COL11A2/DFNA13</i>	Nonsyndromic deafness	400
<i>LGMD 2G</i>	Limb-girdle muscular dystrophy	401
<i>EVC</i>	Ellis-Van Creveld syndrome, Weyer's acrodistal dysostosis	402
<i>ACTN4</i>	Familial focal segmental glomerulosclerosis	403
<i>SCN1A</i>	Generalized epilepsy with febrile seizures plus type 2	404
<i>AASS</i>	Familial hyperlysinaemia	405
<i>NDRG1</i>	Hereditary motor and sensory neuropathy-Lom	406
<i>CNGB3</i>	Total colour-blindness	407, 408
<i>MUL</i>	Mulibrey nanism	409
<i>USH1C</i>	Usher type 1C	410, 411
<i>MYH9</i>	May-Hegglin anomaly	412, 413
<i>PRKAR1A</i>	Carney's complex	414
<i>MYH9</i>	Nonsyndromic hereditary deafness DFNA17	415
<i>SCA10</i>	Spinocerebellar ataxia type 10	416
<i>OPA1</i>	Optic atrophy	417
<i>XLCSNB</i>	X-linked congenital stationary night blindness	418
<i>FGF23</i>	Hypophosphataemic rickets	419
<i>GAN</i>	Giant axonal neuropathy	420
<i>AAAS</i>	Triple-A syndrome	421
<i>HSPG2</i>	Schwartz-Jampel syndrome	422

HGP: možnosti využitia

- Etiológia genetických a nádorových ochorení
- Diagnostika monogénnych a nádorových ochorení
- Diagnostika multifaktoriálnych ochorení
- Génová terapia (genetických a nádorových ochorení)
- Cielená liečba („molekulárna farmakológia“)
- Štúdium biologických funkcií na molekulárnej úrovni (diferenciácia, starnutie ...)
- Evolúcia človeka a pôvod ľudských populácií
- **Nosce te ipsum**: podstatný krok k poznaniu biológie nášho druhu
- Podstatná časť práce je ešte len pred nami