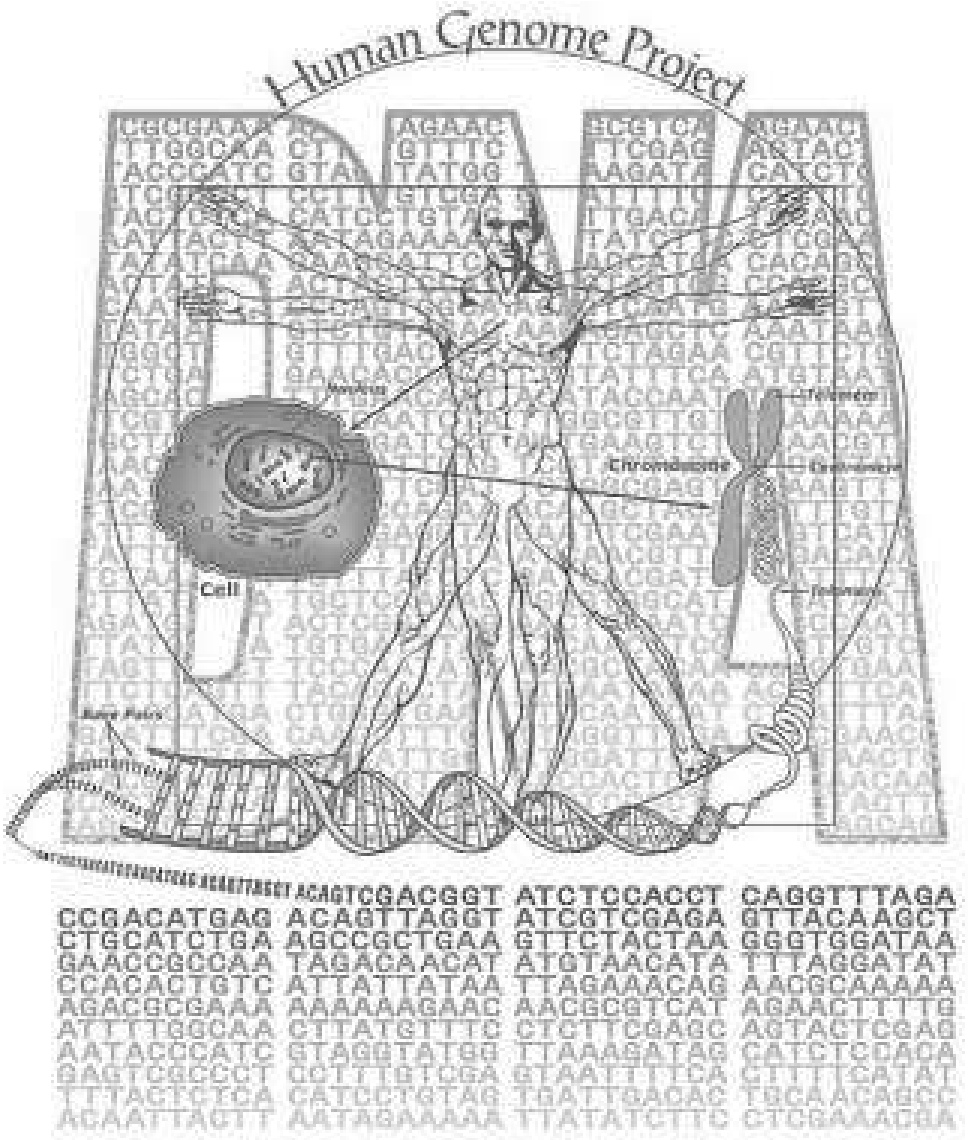
Projekt ľudského genómu

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

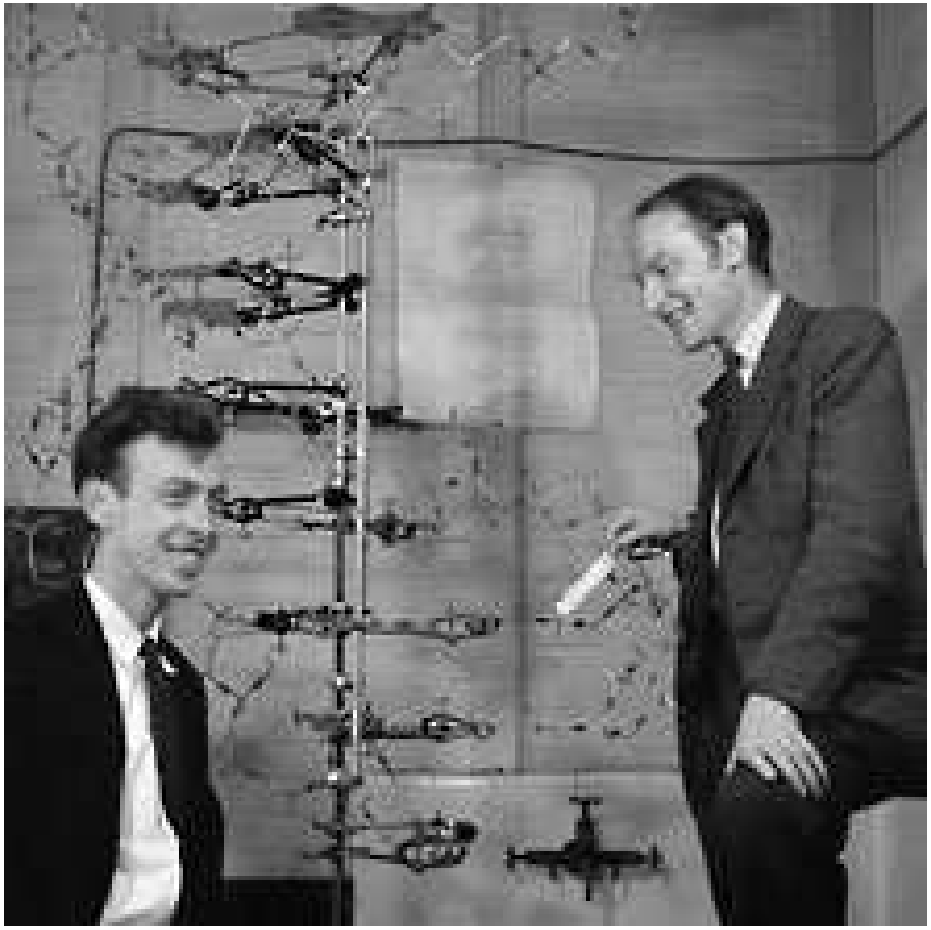
---

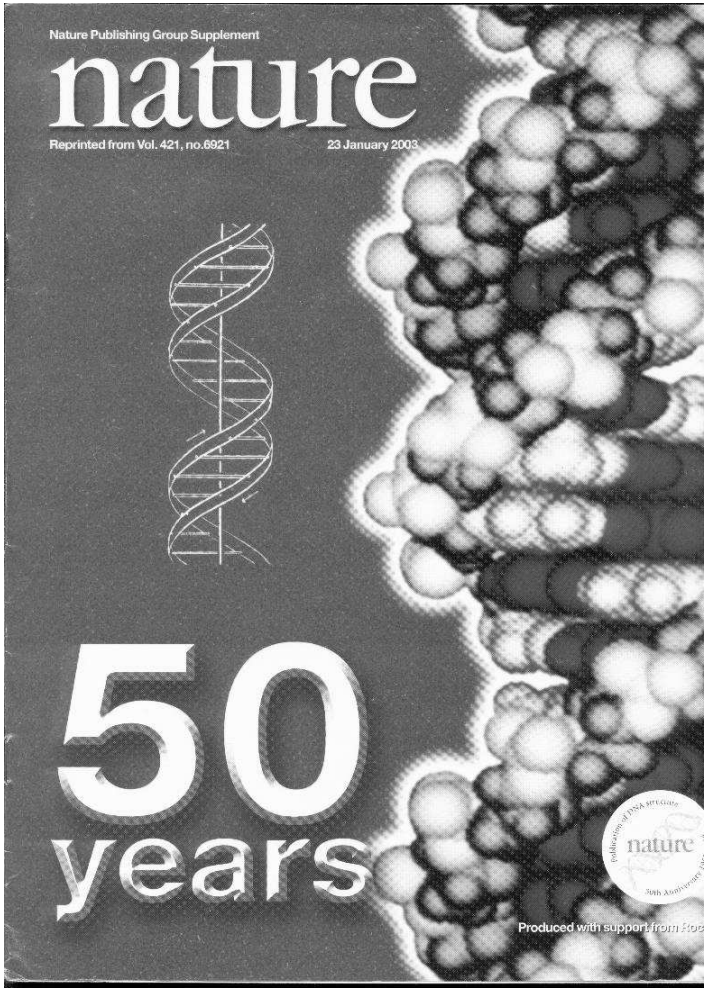
základná www stránka:

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

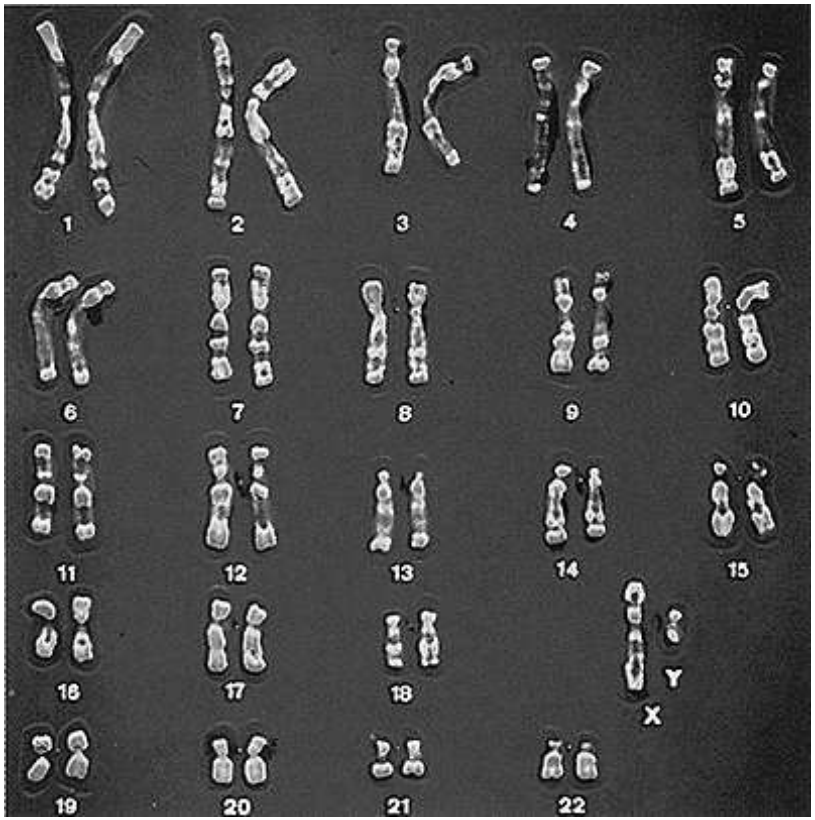
„It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material“

Genetika človeka v r. 1953

- hodnotenie rodokmeňov
- štúdium dvojčiat
- cca 50 monogénnych ochorení
- 5 monogénnych 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 x 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	udalosť
1	1965	*Holley* sekv. alanyl tRNA *S. cerev.*
15	1970	*Wu* sekvenoval 12 bp lambda fága
150	1977	*Sanger* dideoxy, *Gilbert* chem. sekv.
1 500	1978	*Sanger* sekv. phi-x174 (5375 bp)
15 000	1980	*Messing* vyvinul M13 vektor
25 000	1986	*Hood* – čiastočne automat. sekv.
1,000 000	1995	*Venter* – autom. fluoresc. sekv.
150, 000 000	1999	Perkin-Elmer 96 kapilárny sekv.
? 000, 000 000	2004	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

Snímek 12

d11 dekanat1; 3.2.2002

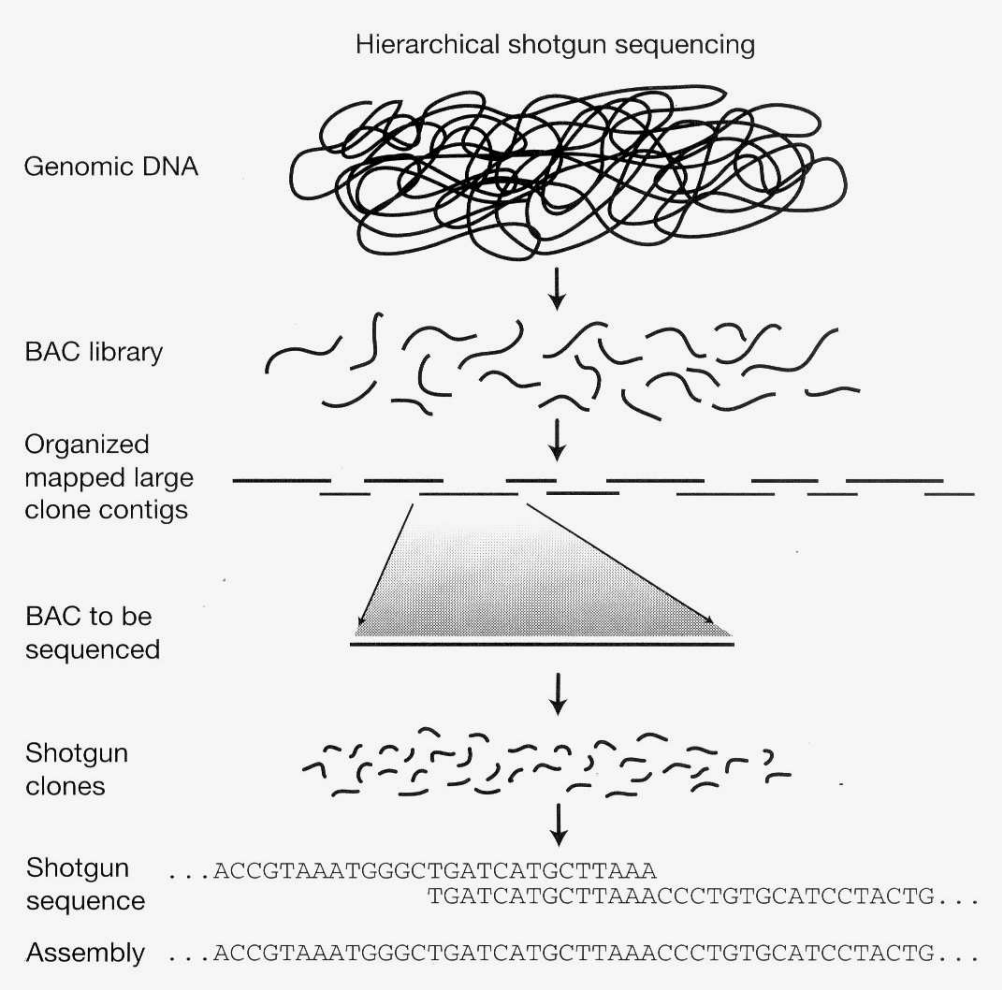
Triangulácia genómu: genetické mapy

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 x 10^6 mapovaných SNP)

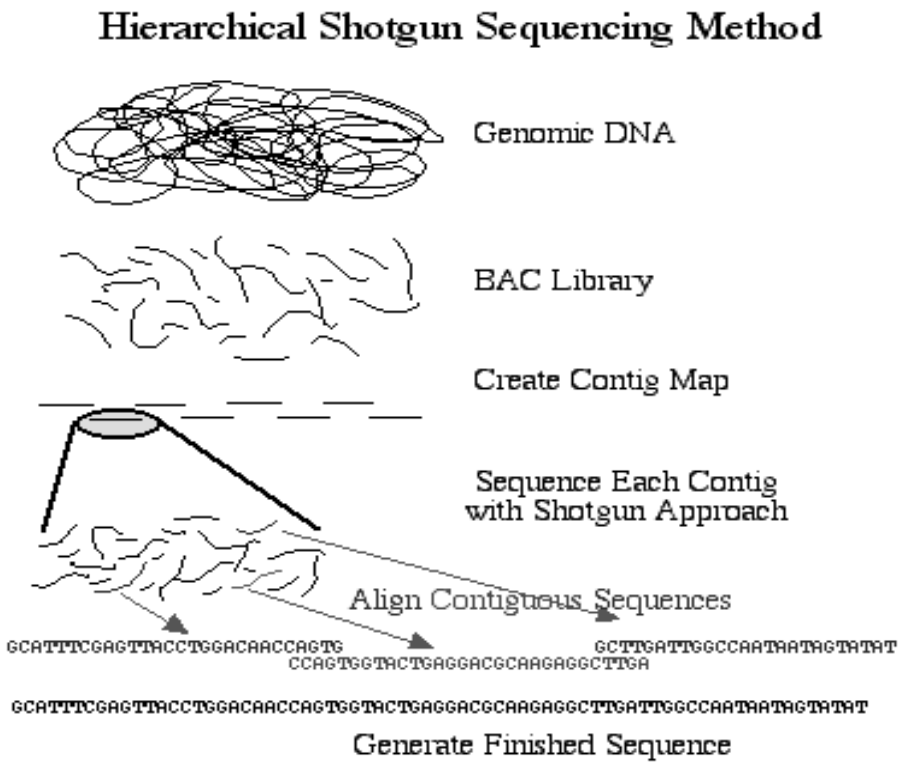
International Human Genome Sequencing Consortium



Hierarchická metóda (clone-by-clone):

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

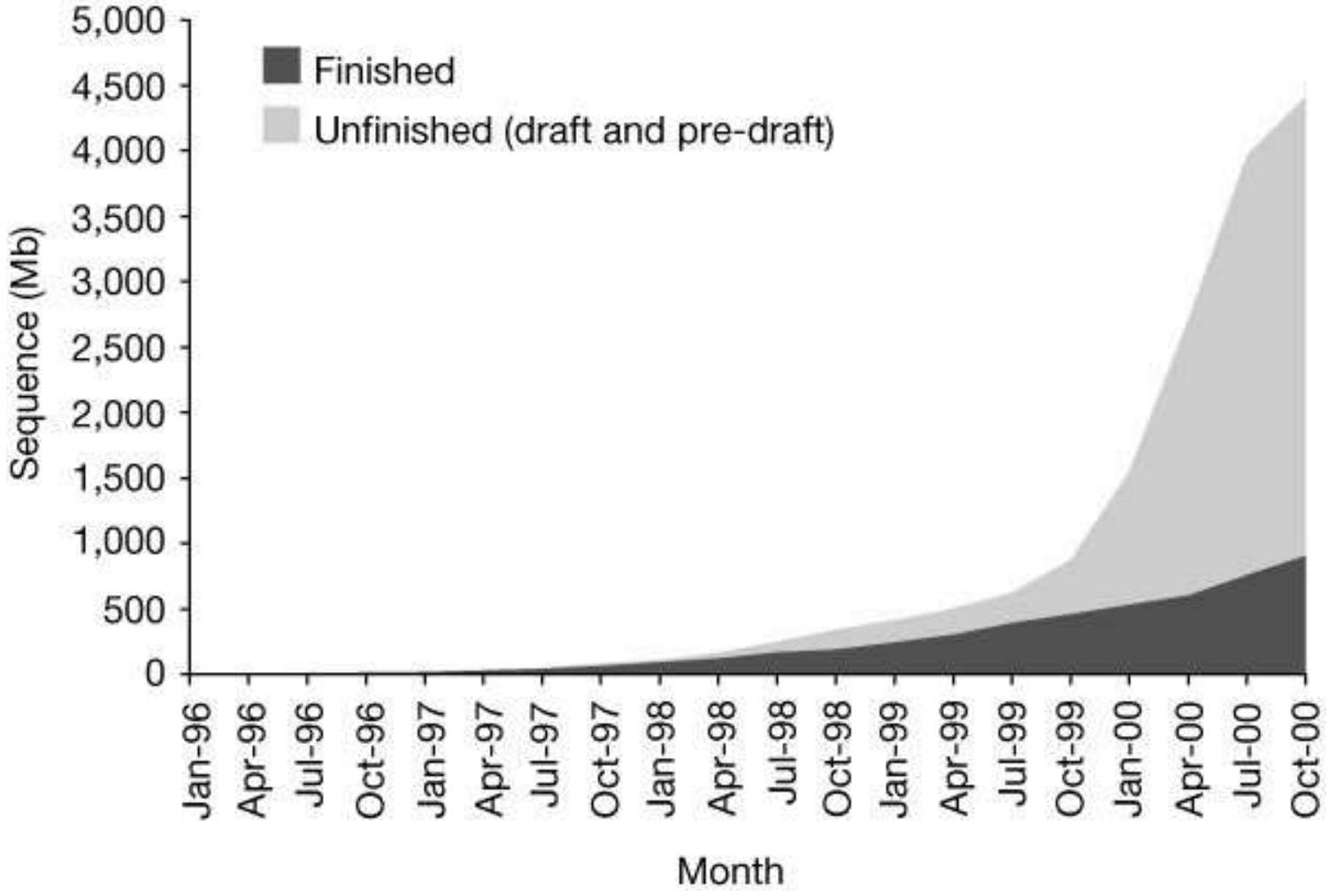
HGP: hierarchická stratégia sekvenovania („clone-by-clone“)



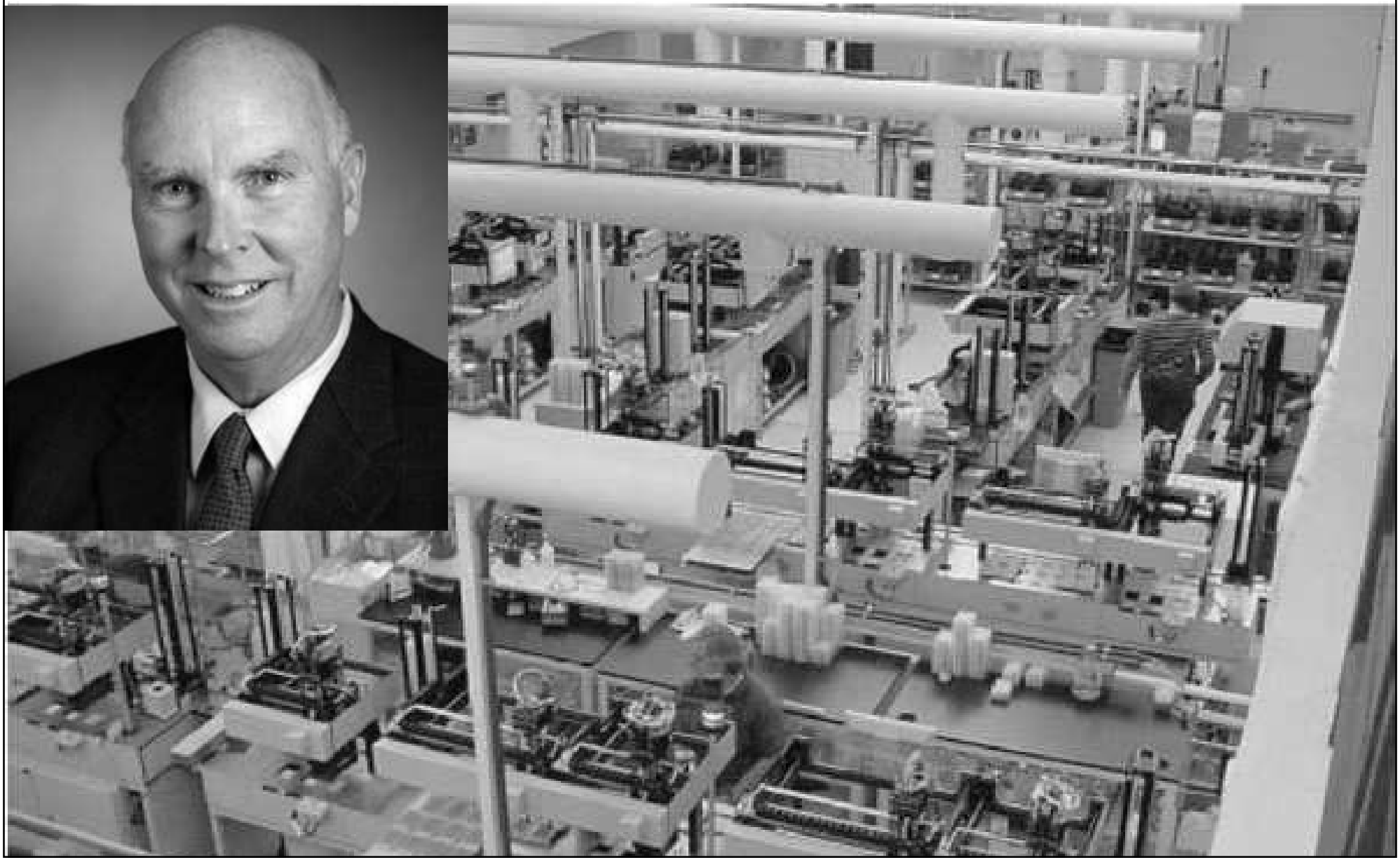
- triangulovať genóm pomocou STR
- zostrojiť BAC-knižnice
- mapovať klony do STR mapy
- utvoriť BAC kontigy
- štiepiť BAC-y náhodne na sekvenovateľné fragmenty
- sekvenovať fragmenty
- zoradiť do konečnej sekvencie klonu
- zoradiť sekvencie klonov do definitívnej sekvencie
- anotovať sekvenciu

HGP: Francis Collins a Eric Lander

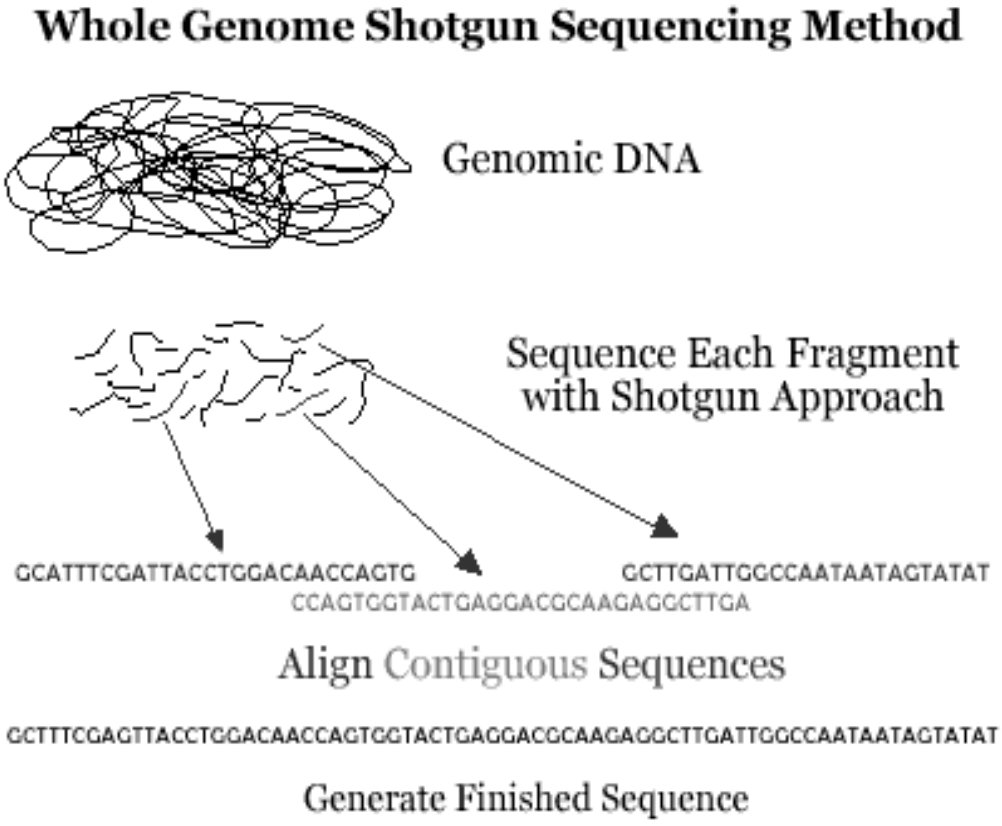
Nárast údajov o sekvencii 1996-2001



Craig J. Venter a Celera Genomics



Stratégia sekvenovania: Celera Genomics



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


15 February 2001
nature
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 evol
Here we report the results of an international collaboration to produce and make freely available a draft sequence of the h
genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequ**

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] 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 Fasul
Michael Flanigan,[1] Liliana 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 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,
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] 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 Karlak,[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] 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
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] 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 Zhu[1]

HGP v Bielom Dome, jún 2001

F. Collins a C. J. Venter 2001

Sekvenované eukaryo genómy

organizmus	rok	Mb	%	% eu	gény	g./Mb
S.cerevisiae	1996	12	93	100	5 800	483
C. elegans	1998	97	99	100	19 099	197
Drozofila	2000	116	64	97	13 601	117
Arabidopsis	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 repetitivnej 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

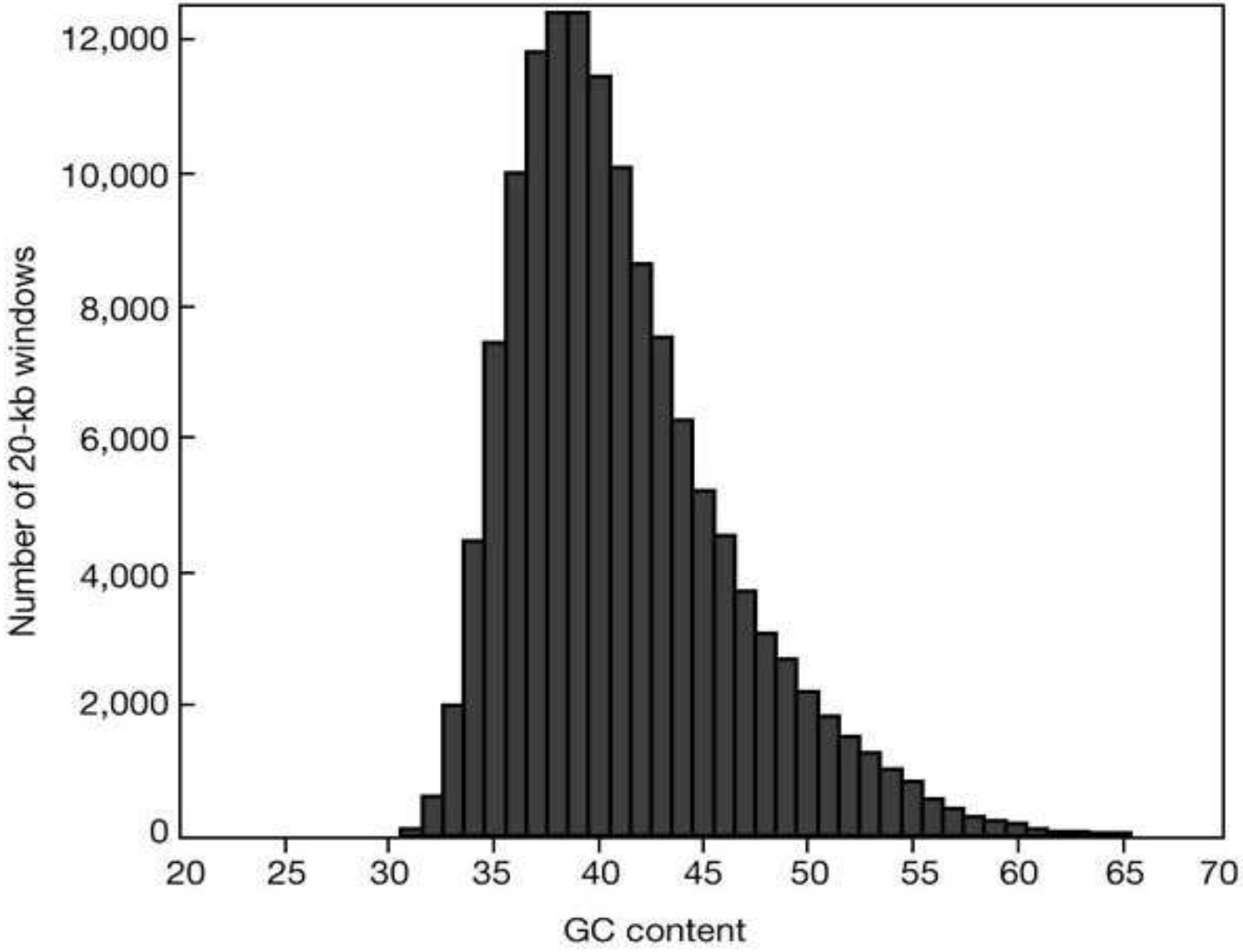
INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM

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

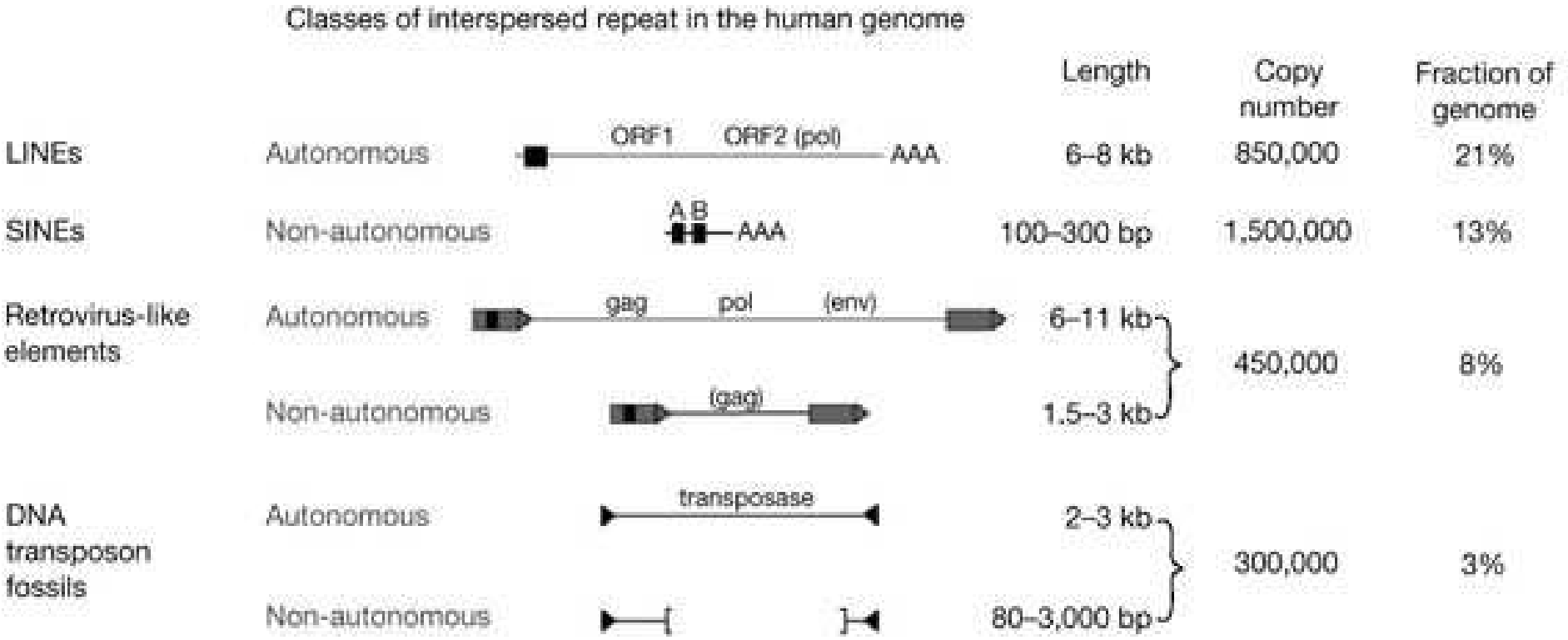
Dokončená sekvencia (2004): základné údaje

- 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



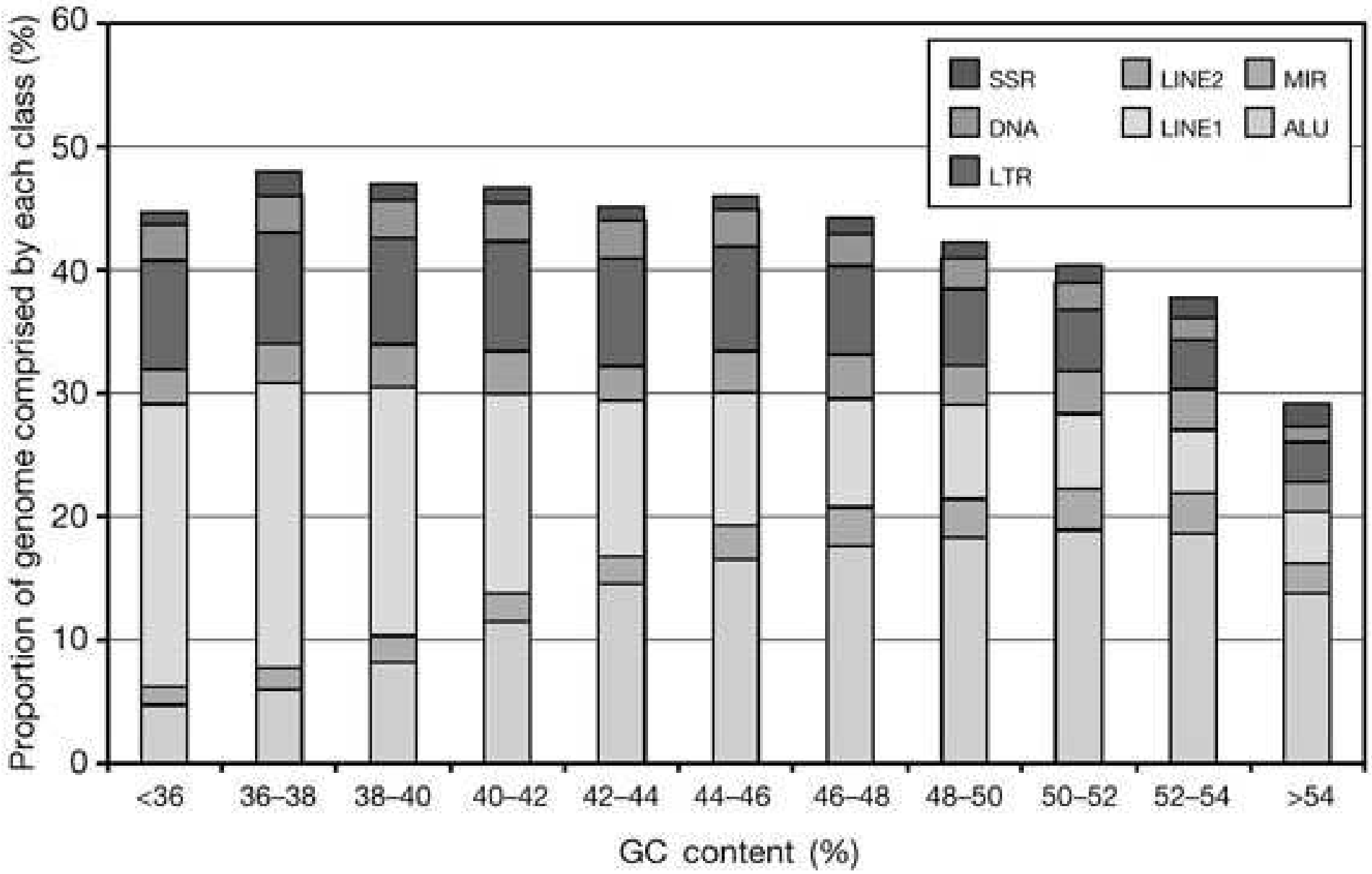
Triedy rozptýlených repetícií

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

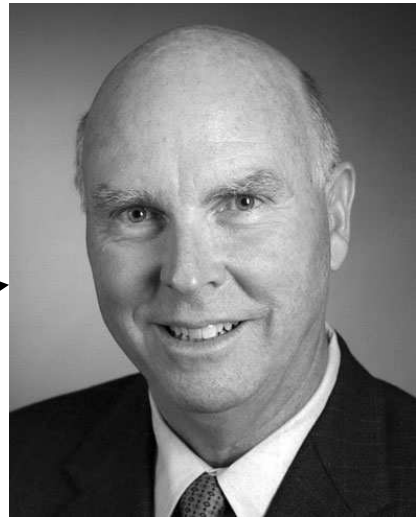
	Number of copies (× 1,000)	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)	Number of families (subfamilies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1 (~20)
MIR	393	60.1	2.20	1 (1)
MIR3	75	9.3	0.34	1 (1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	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

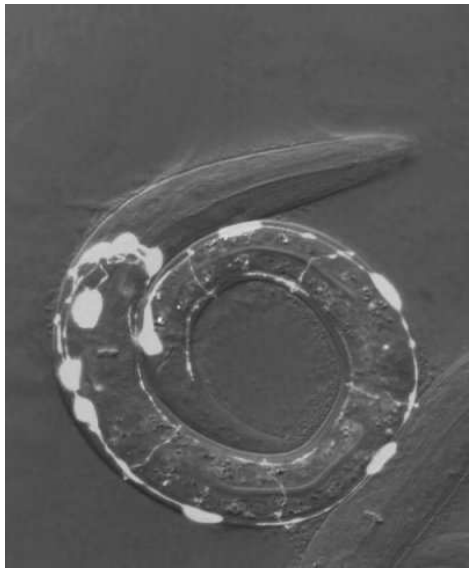


Počet génov u rozličných organizmov



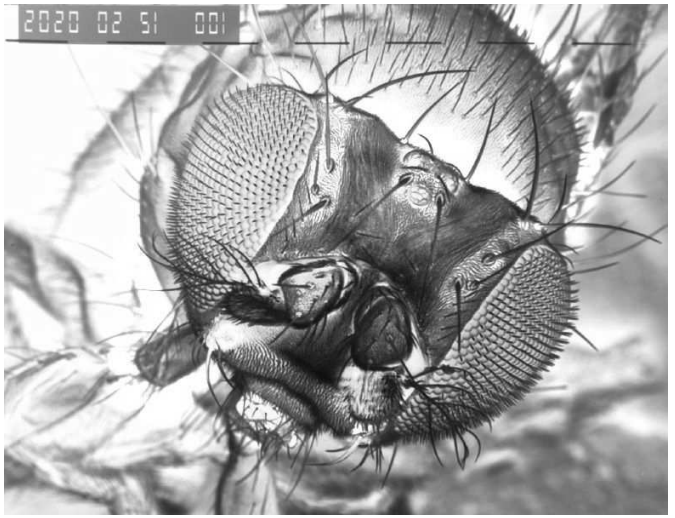
- **Človek**

  okolo 25 000



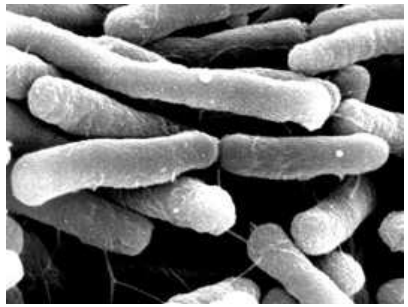
- ***Caenorhabditis elegans***

  okolo 20 000



- ***Drosophila melanogaster***

  14 000

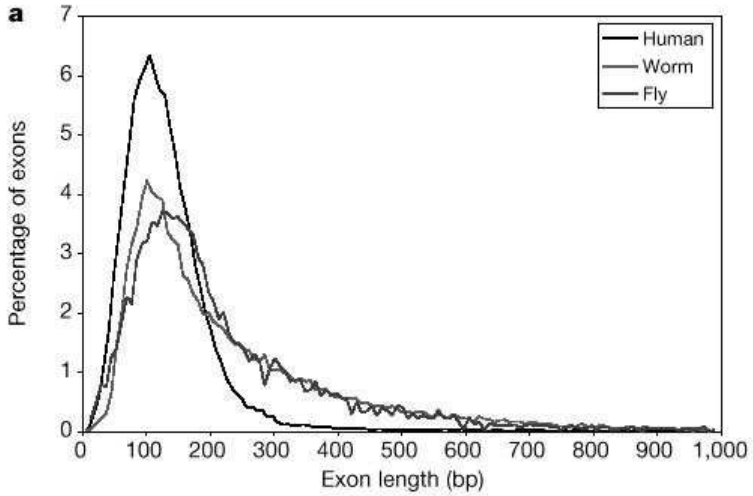


- ***E. coli***

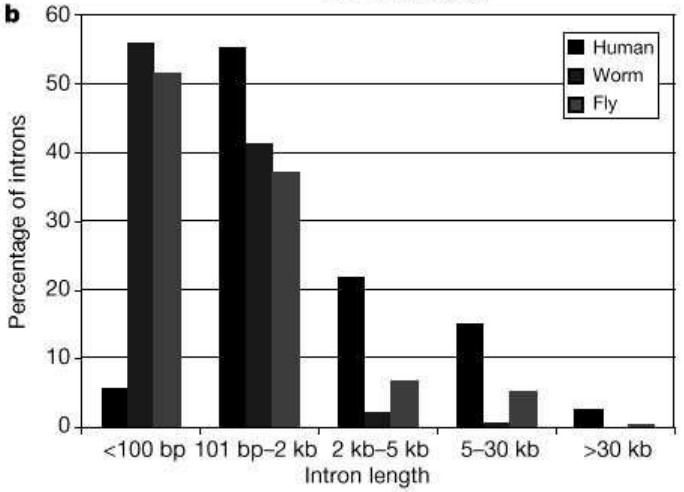
  okolo 6000

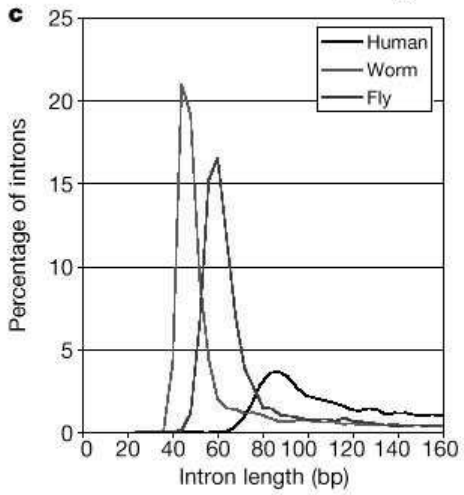
Dĺžková distribúcia
(človek, drozofila, C. e.):

exónov:

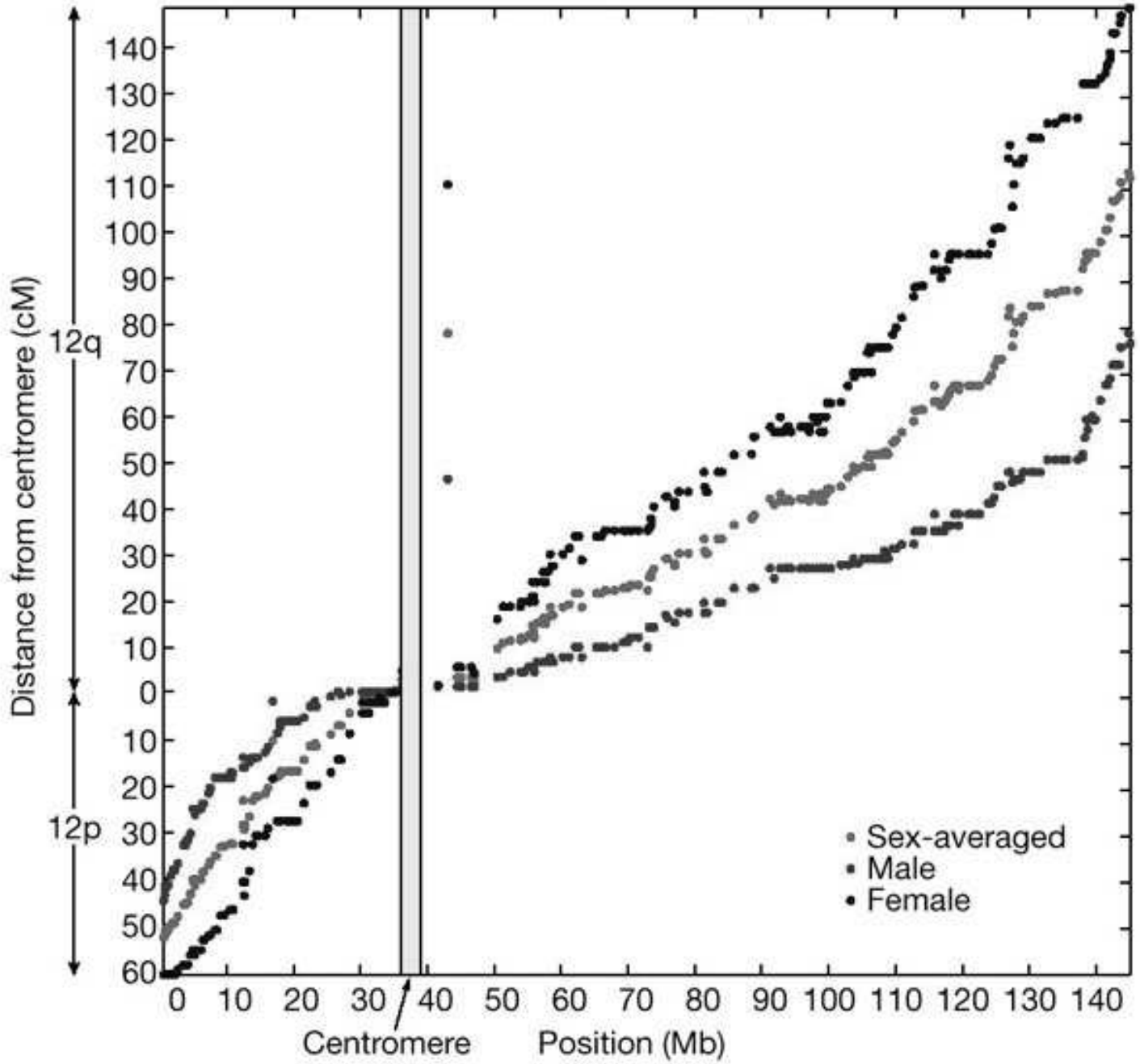


intrónov:

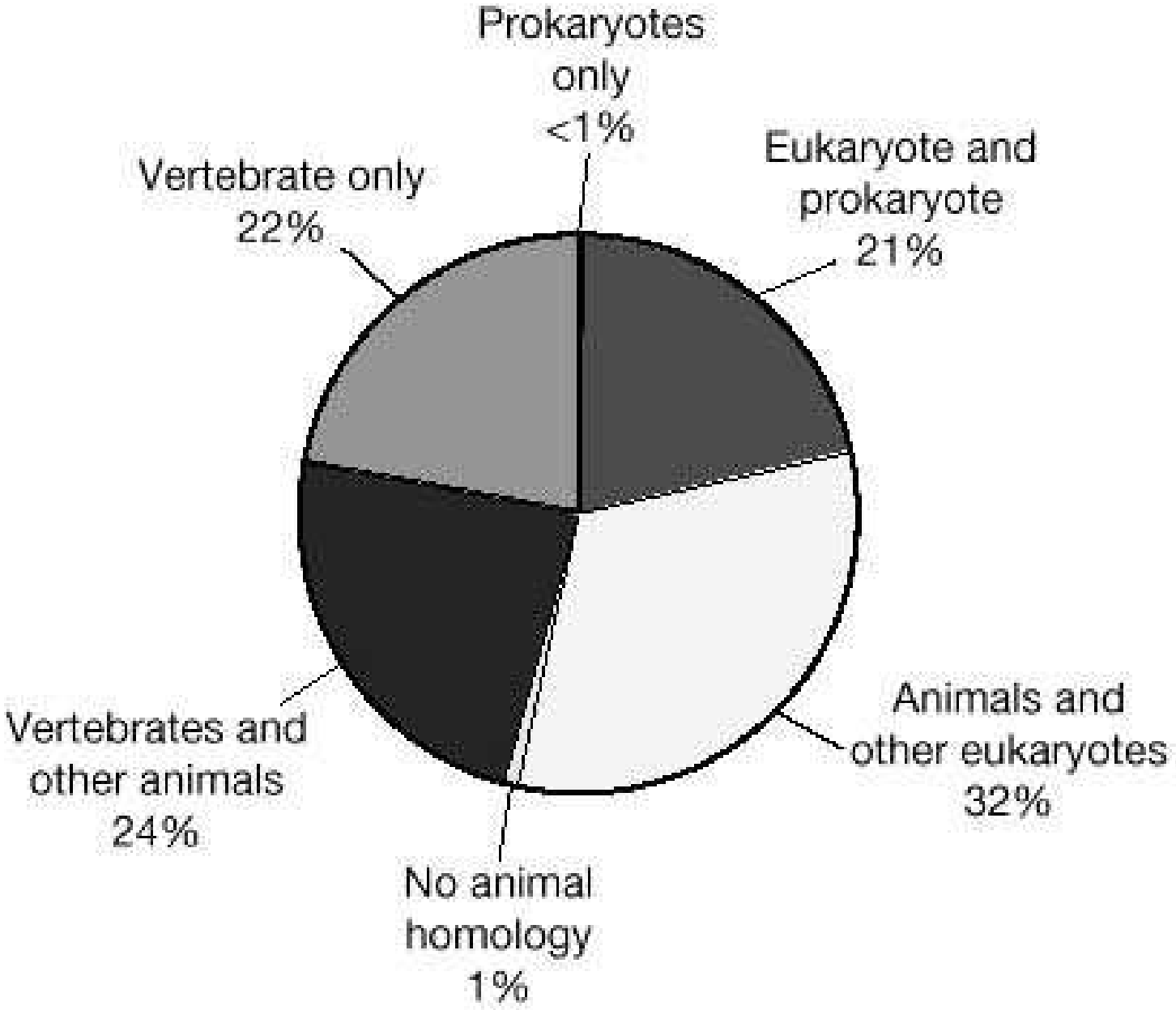




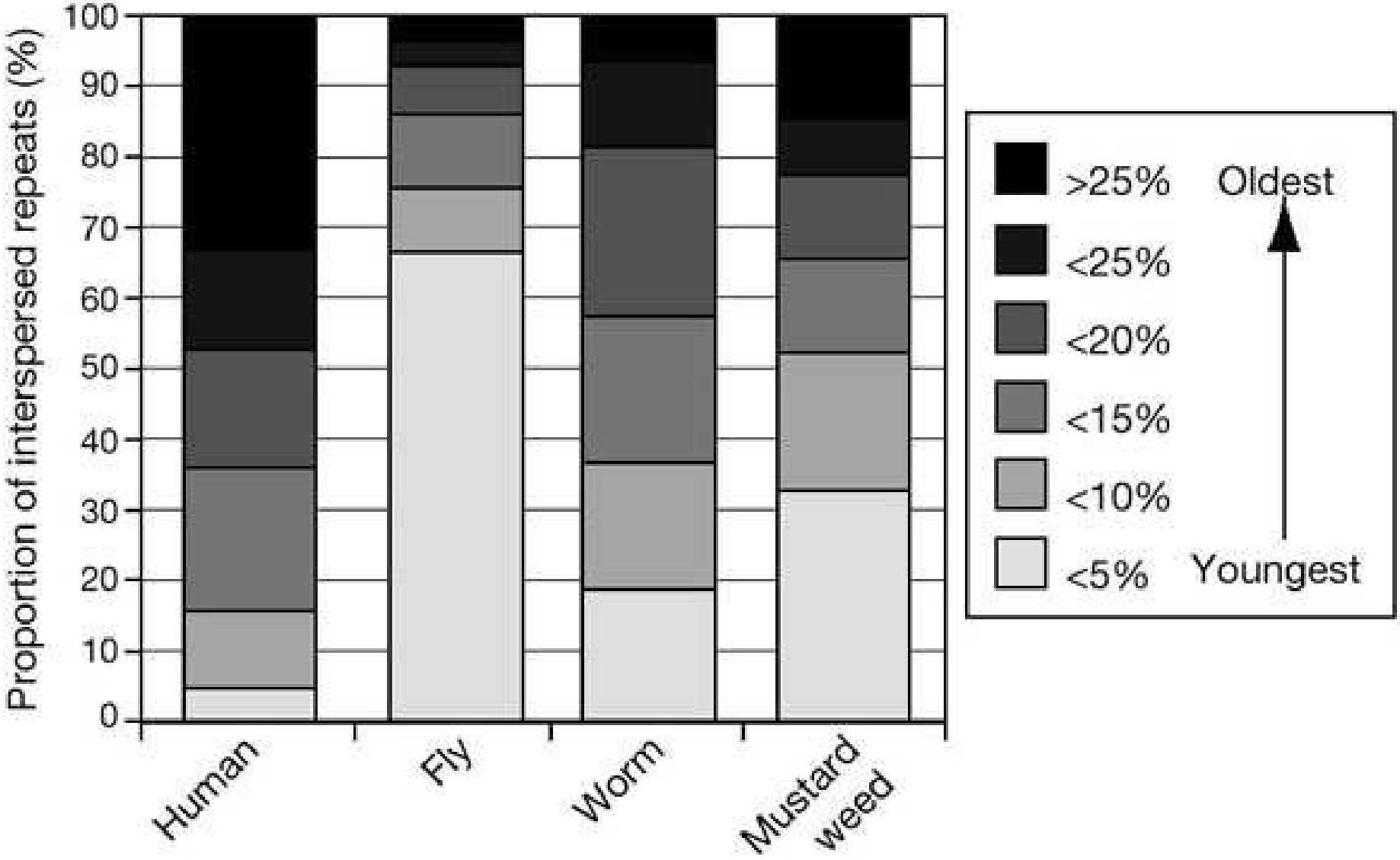
Chromozóm 12: Mb / cM



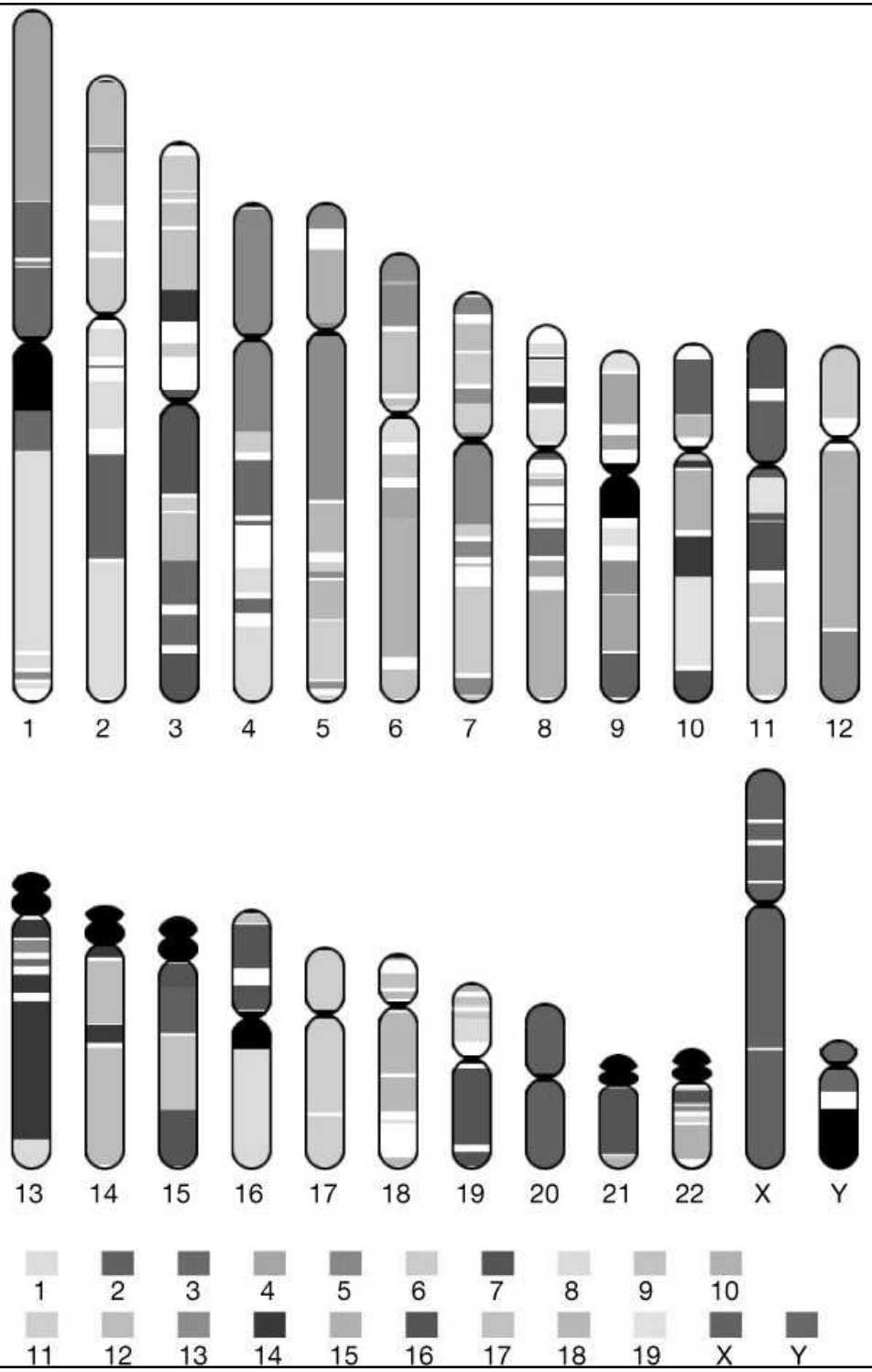
Distribúcia homológov ľudských proteínov



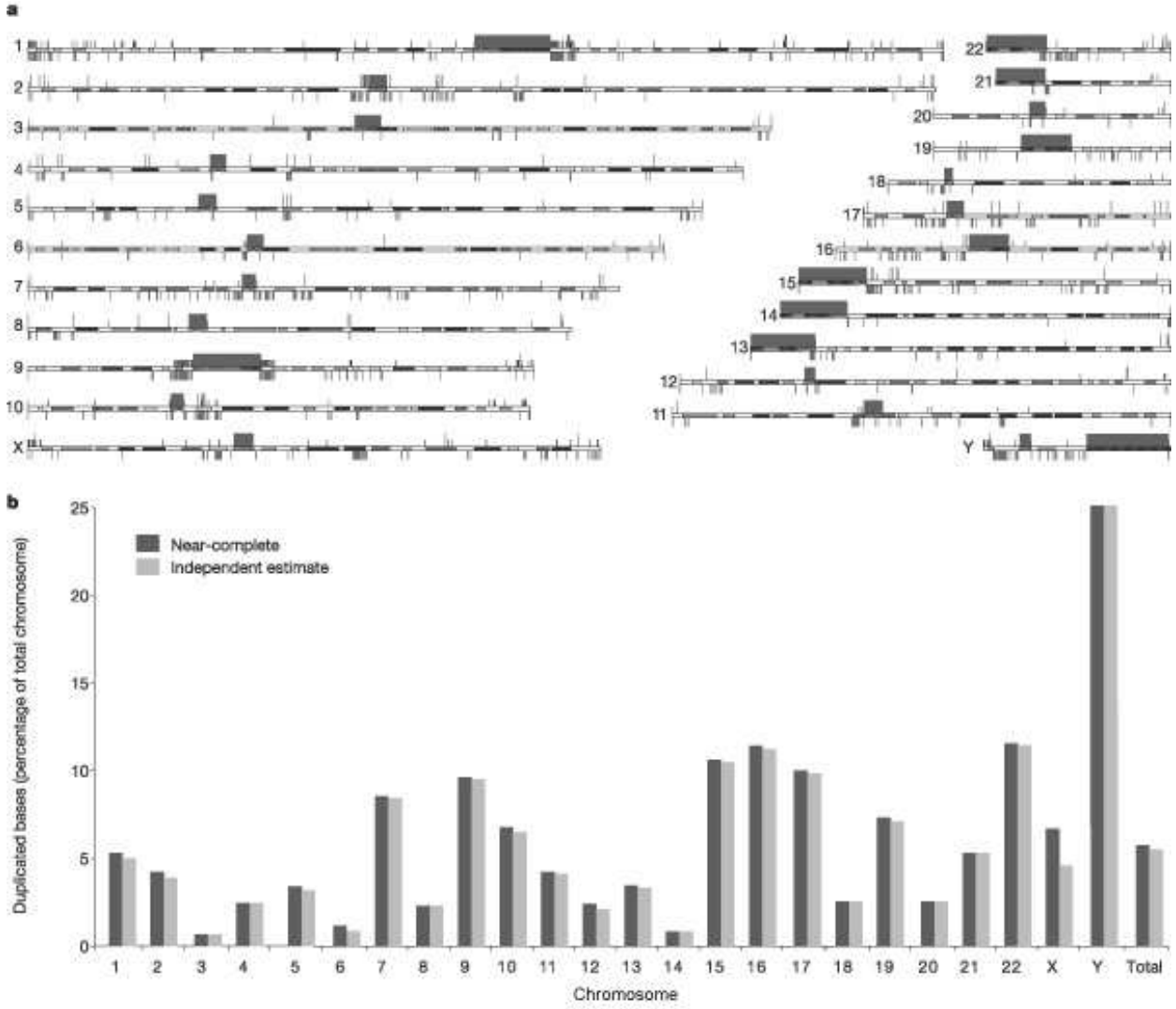
Evolučný vek rozptýlených repetícií v eukaryo-genómoch



Konzervované segmenty ľudského a myšieho genómu



Segmentové duplikácie v ľudskom genóme



Pozičné klonované gény s využitím HGP sekvencie

Table 26 Disease genes positionally cloned using the draft genome sequence

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