

Kolegij Molekularna biologija

ORGANIZACIJA I SEKVENCE STANIČNIH GENOMA

dr. sc. Gordana Maravić
 Zavod za biokemiju i molekularnu biologiju
 Farmaceutsko-biokemijski fakultet
 Sveučilišta u Zagrebu

GM 2004

- Sekvenciranje čitavih genoma
- Složenost eukariotskih genoma
- Kodirajuća i nekodirajuća DNA
- Kromosomi, kromatin, lokalizacija nekodirajuće DNA
- Nukleotidni sljedovi genoma različitih organizama

GM 2004

Genetička informacija organizma sadržana je unutar genoma

Genom

- ukupna DNA organizma
 - virusi - jednolančana ili dvolančana DNA ili RNA
 - svi stanični organizmi - dvolančana DNA
- veličina genoma raste prema višim organizmima (bakterija - 3 Mpb; čovjek - 3000 Mpb)

Gen

- funkcionalna jedinica genoma
- odsječak DNA koji kodira za polipeptidni lanac ili molekulu RNA

Ekspresija gena

- nastanak genskog produkta

Središnja dogma molekularne biologije

- prijenos genske poruke: DNA → RNA → PROTEIN

GM 2004

Sekvenciranje čitavih genoma viših organizama

- prije par desetljeća neostvariv pothvat (80-ih godina najveći sekvencirani genom 180 kb – Epstein-Barrov virus)
- danas realnost zahvaljujući automatskom sekvenciranju
- sekvenciranje humanog genoma – 2 skupine znanstvenika:
 - Eric Lander (International Human Genome Sequencing Consortium)
 - Craig Venter (Celera Genomics)
- skica humanog genoma objavljena 2001.
 - Nature, 409(6822) (2001)
 - Science, 291(5507) (2001)

GM 2004

2 pristupa sekvenciranju genoma

Shotgun Sequencing (sekvenciranje metodom "sačmarice")

- nasumično kidanje genoma u male preklapajuće fragmente
- višestruko sekvenciranje fragmenata da se pokrije cijeli genom (oko 10X)
- sklapanje preklapajućih sljedova daje kompletne kromosomske regije

Ordered Clone Sequencing (sekvenciranje uređenih klonova)

- podjela kromosoma u manje dijelove (*engl.* contigs)
- organizirano kartiranje dobivenih dijelova
- sekvenciranje svakog od tih dijelova klasičnim metodama (nije potrebno višestruko sekvenciranje jer su poznata mjesta preklapanja)
- sklapanje sekvenciranih dijelova u cjelinu

GM 2004

2 pristupa sekvenciranju genoma

Craig Venter

Shotgun Sequencing

Ordered Clone Sequencing

Eric Lander

GM 2004

Do danas sekvencirani genomi

Prema National Centre for Biology Information, National Institutes of Health, USA (<http://www.ncbi.nlm.nih.gov/>)

- **Virusi** – 1383 genoma
- **Bakterije i arheobakterije** – 197 genoma (266 u tijeku)
- **Gljive i kvasci** – 8 genoma
- **Eukarioti** – 21 genom (dvadesetak u tijeku)
- **Mitohondriji** – 621 genom
- **Plastidi** – 41 genom

GM 2004

Koliko se ljudi međusobno razlikuju prema sastavu genoma?

Manje od 0.1%

GM 2004

Čovjek ima oko 30 000 gena



Pan troglodytes
30 000 gena
98% homologije



Mus musculus
30 000 gena
90% homologije



Danio rerio
30 000 gena
85% homologije



D. melanogaster
13 600 gena
36% homologije



A. thaliana
26 000 gena
26% homologije



S. cerevisiae
6 275 gena
23% homologije



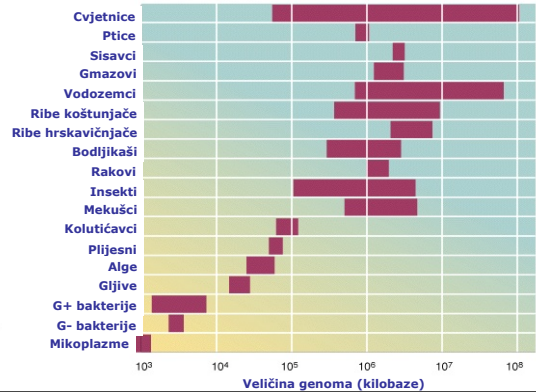
C. elegans
19 000 gena
21% homologije



E. coli
4 300 gena
7% homologije

GM 2004

Veličina genoma nije razmjerna složenosti organizma



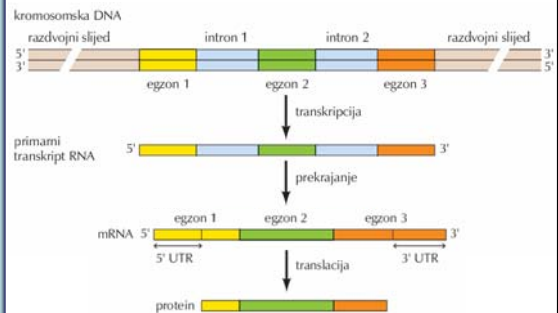
GM 2004

Velik dio eukariotskih genoma čini nekodirajuća DNA

- dijelovi eksona koji se ne prevode u protein (engl. *untranslated regions* –UTR)
- **introni**
- **razmaknice** (engl. *spacers*)
- **ponavljajući sljedovi DNA**
 - ponavljanja jednostavnih sljedova
 - **transpozoni**
 - DNA transpozoni
 - retrotranspozoni
- **pseudogeni**

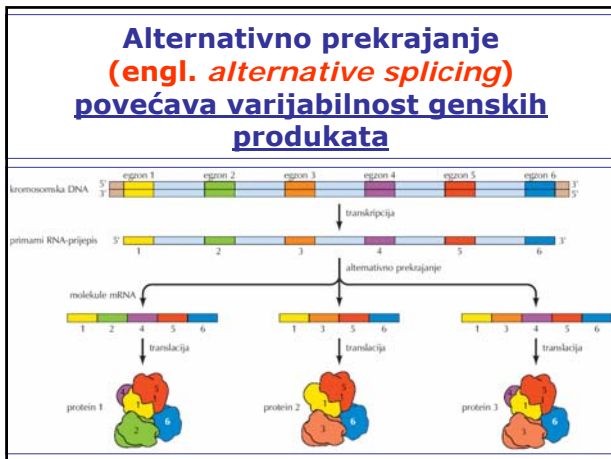
GM 2004

Struktura eukariotskih gena i izrezivanje introna



Introni čine oko 25% ukupne genomske DNA čovjeka

GM 2004



Ponavljajući sljedovi DNA

- čine čak preko 50% DNA sisavaca!
- nekoliko vrsta ponavljajućih sljedova

GM 2004

detekcija praćenjem kinetike reasocijacije DNA

Ponavljanja jednostavnih sljedova

- *engl. simple sequence repeats*
- uzastopno ponovljeni nizovi nekoliko tisuća kopija kratkih sljedova duljine 1-500 nukleotida (npr. ACAAAT u vinskoj mušici)
- **satelitna DNA**
- čini oko 10% ukupne DNA

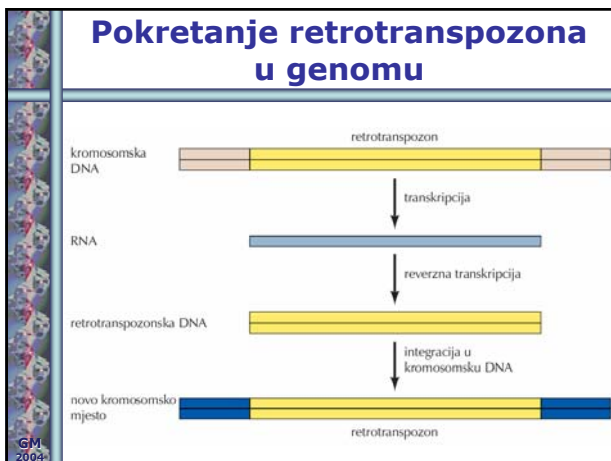
Zbog jedinstvenog sastava satelitna DNA može se razdvojiti iz ostatka genomske DNA ravnotežnim centrifugiranjem u gradijentu gustoće CsCl

GM 2004

Pokretni genetički elementi - transpozoni

- čine oko 45% ukupnog genoma čovjeka
- raštrkani po genomu
- **DNA transpozoni**
 - 300 000 kopija veličine 80-3000 pb (oko 3% humane DNA)
- **retrotranspozoni**
 - **Kratki raspršeni elementi** (*engl. short interspersed elements, SINE*)
 - 1,5 milijun odsječaka od 100-300 pb raštrkano po genomu (oko 13% genomske DNA)
 - **Dugi raspršeni elementi** (*engl. long interspersed elements, LINE*)
 - 850 000 ponavljanja veličine oko 1 kb (oko 21% genomske DNA)
 - **Retrovirusu slični elementi** (*engl. retrovirus-like elements*)
 - 450 000 odsječaka veličine 2-10 kb (oko 8% genomske DNA)

GM 2004

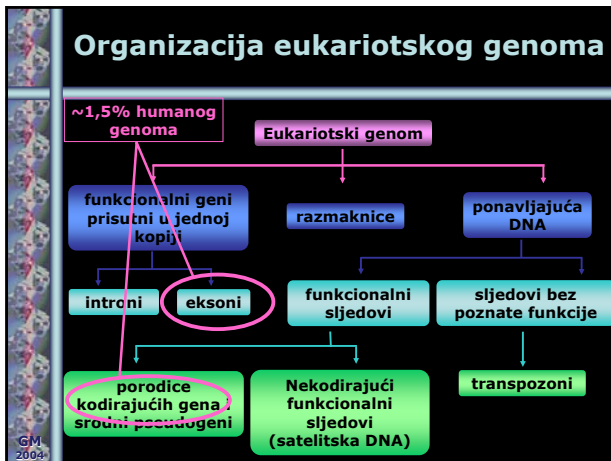


Duplikacija gena i pseudogeni

- mnogi geni postoje u više kopija od kojih su neke često nefunkcionalne - **pseudogeni** (oko 5% ljudskog genoma)
- mogu nastati duplikacijom i mutacijskom inaktivacijom gena pretka

Porodice globinskih gena

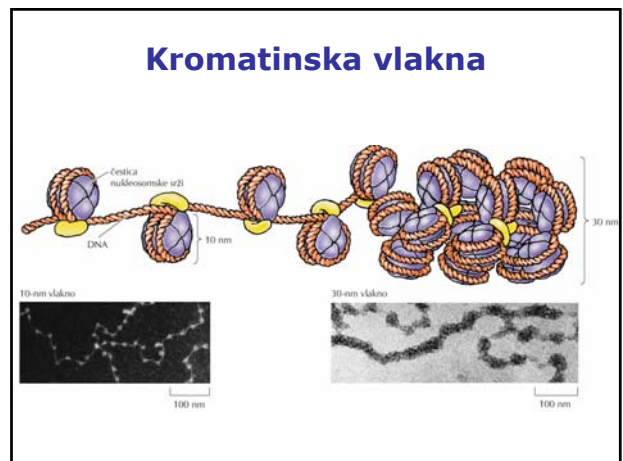
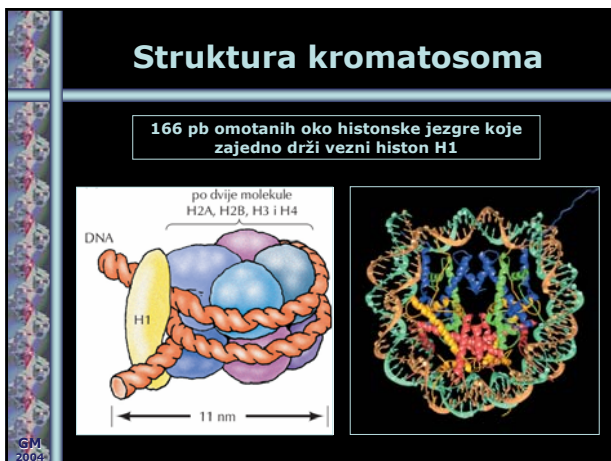
GM 2004

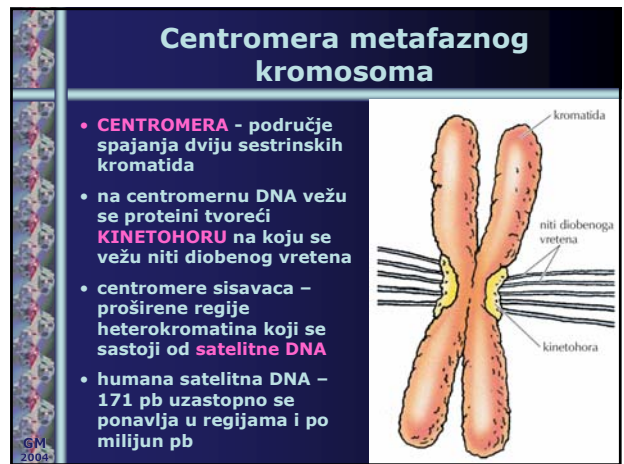
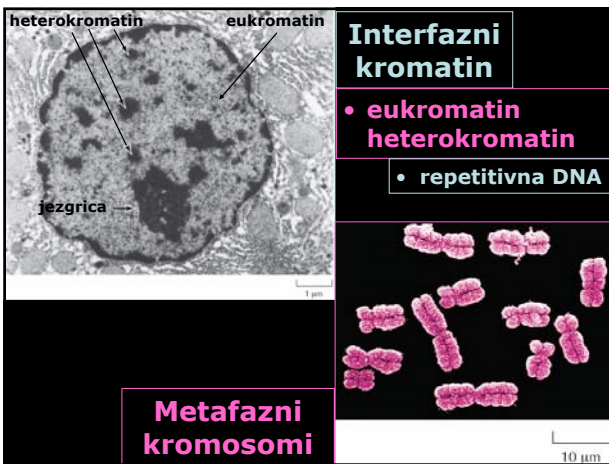
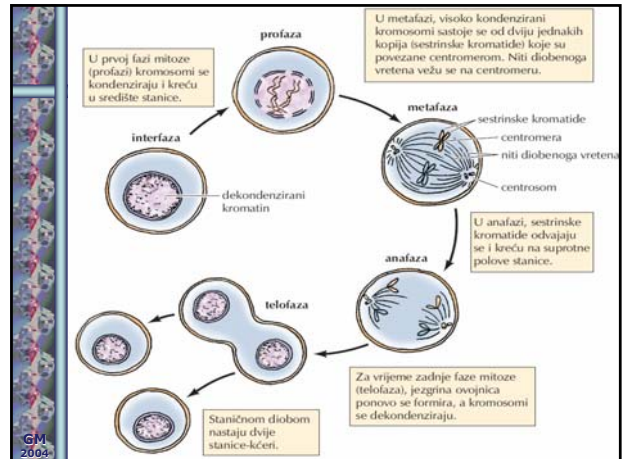
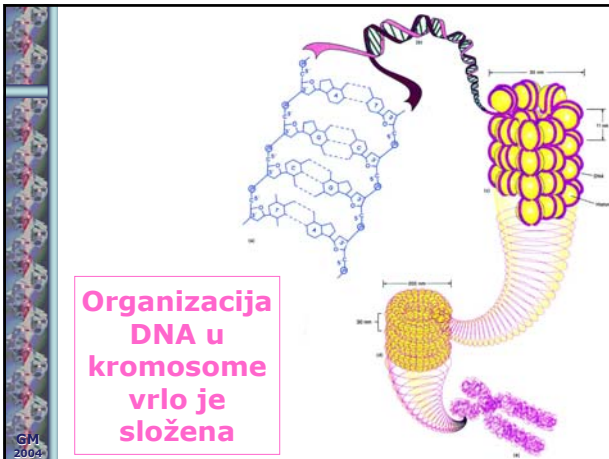


- Sekvenciranje čitavih genoma
- Složenost eukariotskih genoma
- Kodirajuća i nekodirajuća DNA
- Kromosomi, kromatin, lokalizacija nekodirajuće DNA
- Nukleotidni sljedovi genoma različitih organizama

- Sekvenciranje čitavih genoma
- Složenost eukariotskih genoma
- Kodirajuća i nekodirajuća DNA
- Kromosomi, kromatin, lokalizacija nekodirajuće DNA
- Nukleotidni sljedovi genoma različitih organizama

- ## DNA je vrlo dugačka molekula
- haploidni genom čovjeka ima 3 164 700 000 parova baza ≈ 1 m (promjer jezgre je 5-10 μ m)
 - zbog slaganja parova baza DNA je razmjerno kruta molekula
 - grupa bazičnih proteina (**histona**) sudjeluje u organizaciji strukture DNA
 - kompleks između eukariotske DNA i proteina - **kromatin**
 - osnovna strukturna jedinica je **nukleosom**
 - po dva histona H2a, H2b, H3 i H4
 - 140 pb dva puta je omotano oko histona
 - histon H1 veže se na DNA dužine 40 - 100 pb između dva nukleosoma
- 10^{13} stanica \times 2 m = 182 000 000 000 km
(610 puta do sunca i nazad)





I telomere sadrže ponavljajuće sljedove

Organizam	Telomerni ponovljeni slijed
Kvasci <i>Saccharomyces cerevisiae</i>	G_1_3T
<i>Schizosaccharomyces pombe</i>	G_2_3TTAC
Protozoa <i>Tetrahymena</i>	GGGGTT
<i>Dictyostelium</i>	G_1_3A
Biljka <i>Arabidopsis</i>	AGGGTTT
Sisavac Čovjek	AGGGTT

Usporedba genoma različitih organizama

Organizam	Veličina genoma (Mb) ^a	Broj gena	Sljedovi koji kodiraju proteine
Bakterije			
<i>Mycoplasma genitalium</i>	0,58	470	88%
<i>H. influenzae</i>	1,8	1.743	89%
<i>E. coli</i>	4,6	4.288	88%
Kvasci			
<i>S. cerevisiae</i>	12	6.000	70%
<i>S. pombe</i>	12	4.800	60%
Beskralježnjaci			
<i>C. elegans</i>	97	19.000	25%
<i>Drosophila</i>	180	13.600	13%
Biljke			
<i>Arabidopsis thaliana</i>	125	26.000	25%
Riža	440	30.000-50.000	≈ 10%
Sisavci			
Čovjek	3.200	30.000-40.000	1-1,5%

^aMb = milijuni parova baza

GM 2004

Važnost analize cjelokupnih genoma - uloga bioinformatike -

- otkrivanje evolucijskih odnosa
 - arheobakterije i eukarioti dijele zajedničku evolucijsku liniju
- otkrivanje novih gena
- otkrivanje genske funkcije prema sličnosti s genima iz drugih organizama
 - komplementarni pristup eksperimentalnim istraživanjima (inaktivacija gena)
- otkrivanje regulatornih sljedova
- otkrivanje gena uključenih u bolesti

