

Összehasonlító bioinformatika

Miklós István

MTA-ELTE Elméleti Biológiai és Ökológiai Kutatócsoport

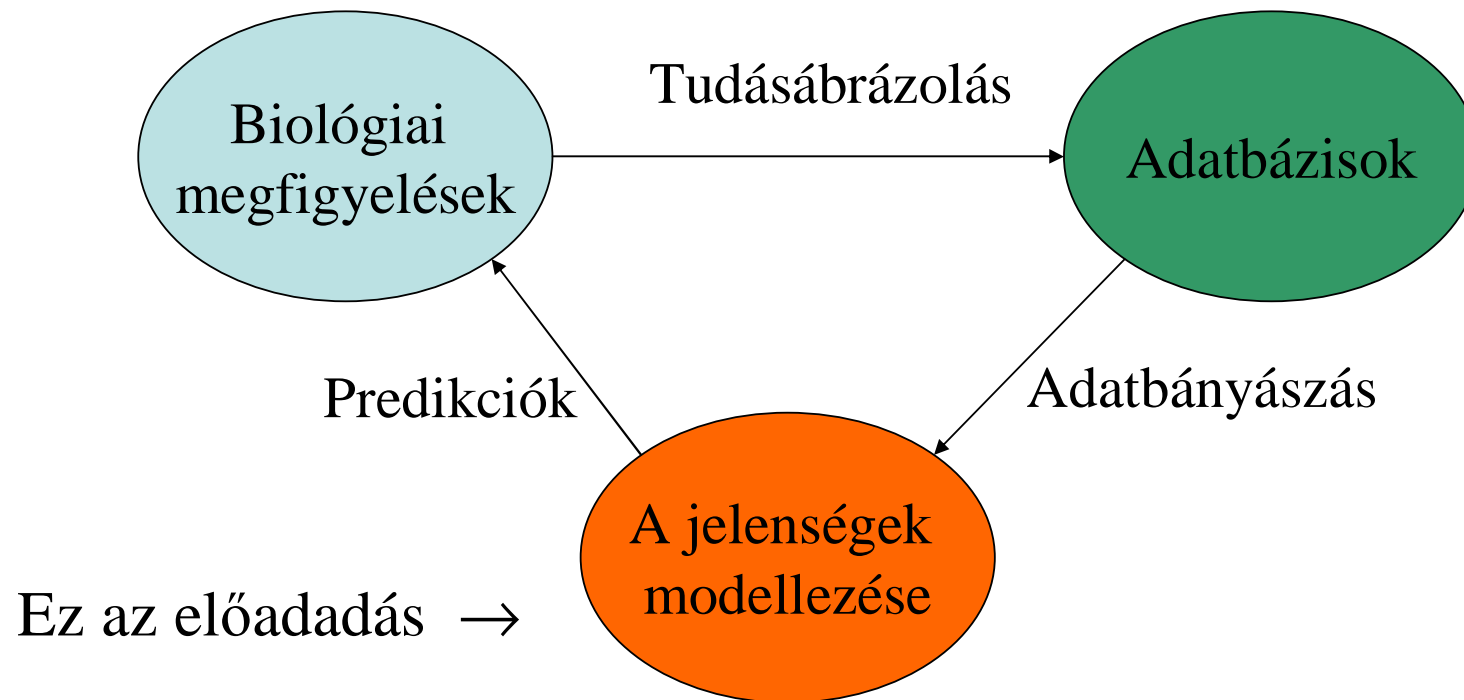
Bioinformatika szeminárium
Rényi Intézet, 2004 október 1.

Tartalom

- Mi is az összehasonlító bioinformatika?
 - Fehérjék másodlagos térszerkezetének predikciója
 - RNS-ek másodlagos térszerkezetének predikciója
 - Génkeresés
- Kihívások a bioinformatikai modellezésben
 - A mutációk egymástól nem függetlenek
 - CpG szigetek
 - Térszerkezetfüggő szubsztitúciók
 - Genomátrendeződés mitochondriumban
 - „Computationaly hard” problémák
 - Pseudoknotok
 - Transzpozíciók
 - Hosszú beszúrás-törlés
 - Összetett modellek
 - Átfedő gének
 - RNS térszerkezet + kódolás mRNSben
 - Kotranszkripcionális folding
- Matematikai kihívások
 - Markov lánc Monte Carlo
 - Statisztikai problémák
 - Algoritmuseleméleti problémák

Mi a bioinformatika?

"All aspects of gathering, storing, handling, analyzing, interpreting and spreading vast amounts of biological information in databases. The information involved includes gene sequences, biological activity/function, pharmacological activity, biological structure, molecular structure, protein-protein interactions, and gene expression. Bioinformatics uses powerful computers and statistical techniques to accomplish research objectives, for example, to discover a new pharmaceutical or herbicide."



Összehasonlító bioinformatika

Központi dogmája:

A struktúra konzervatívabb, mint a szekvencia

Rosetta kő:



Arthur Lesk:

„What one or two homologous sequences whisper, a full multiple alignment shouts out loud.”

Richard Durbin példája:

AYTGTHISSQKLIISCLPNOTKSI AIHIDDENAWYA

AYTGTHISSQKLIISCLPNOTKSI AIHIDDENAWYA
DEFYTHISPSQALISCOMPLETELY IHIDDENYWAE

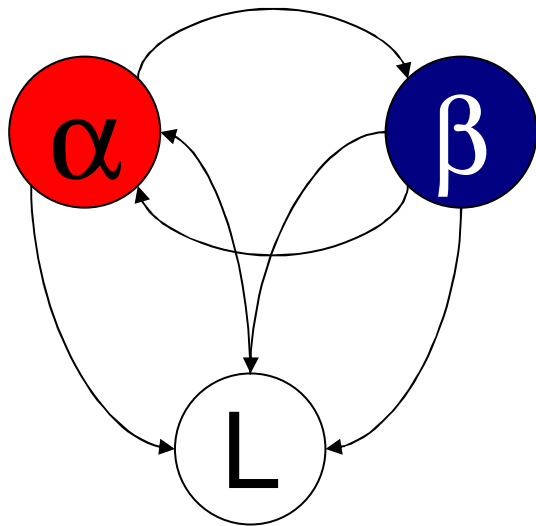
Ez az előadás

- Összehasonlító bioinformatika
 - Struktúrák predikciója
 - Evolúciós leszármazási kapcsolatok vizsgálata
- Sztochasztikus modellezéssel. Ennek előnyei:
 - Paraméterek vizsgálata
 - Becslést adhatunk a predikció jóságára
 - Ez a becslés elősegíti a pontosabb predikciót!
- Evolúciós modellek
 - Egyes paramétereknek biológiai jelentésük van

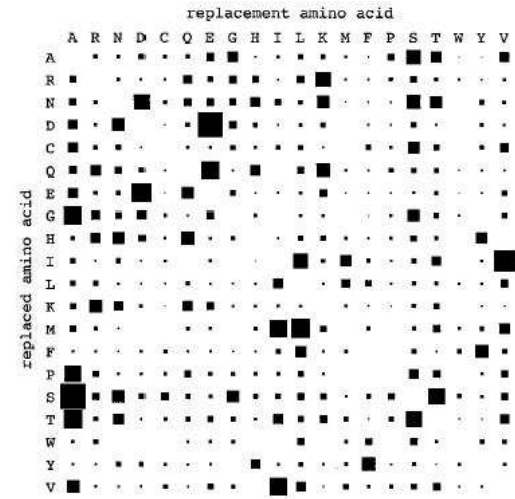
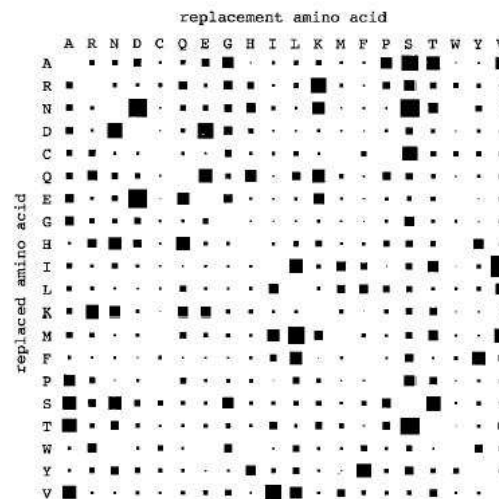
Első alkalmazása az evolúciós információknak

Jones, Thorne, Goldman (1996) *J. Mol. Biol.* **263**:196-208.

Illesztett protein szekvenciák közös struktúrájának a meghatározása

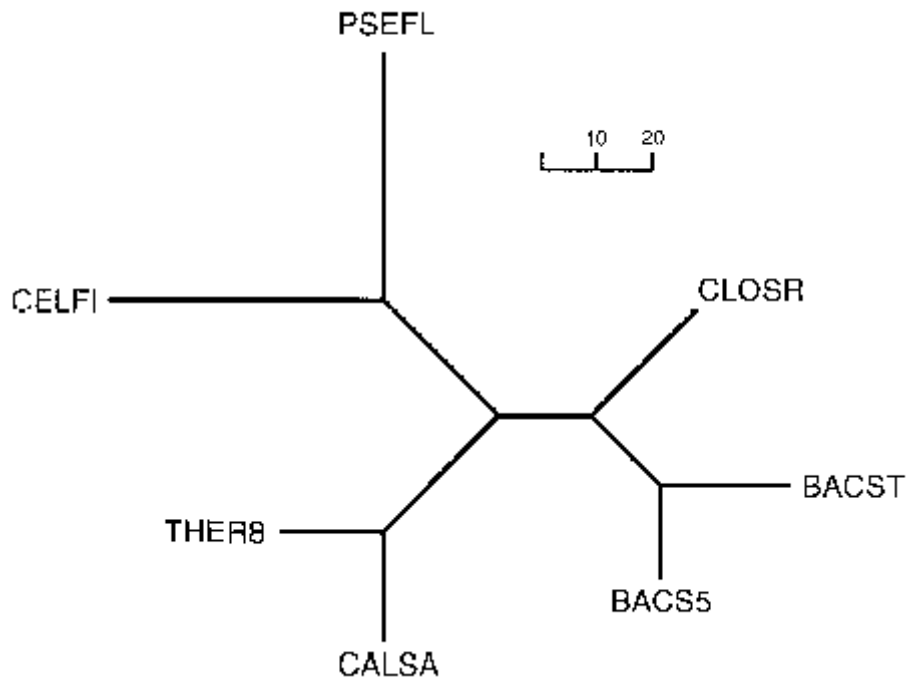


Rejtett Markov Modell

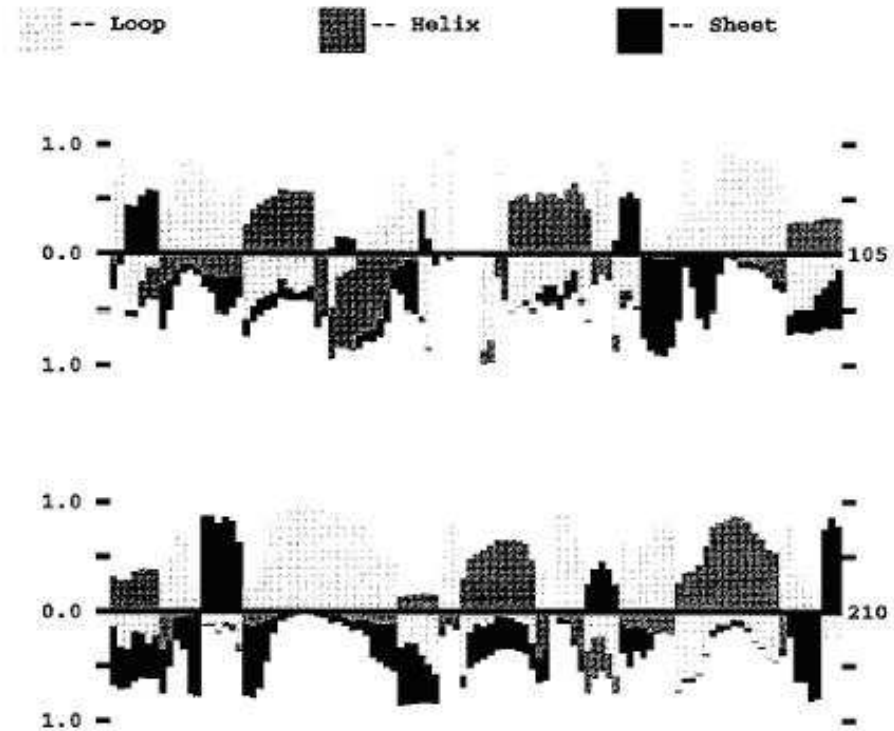


A kibocsájtási valószínűségeket egy időfolytonos Markov modell adja meg.

Eredmények I.



ML Evolúciós fa



Posterior decoding

Eredmények II.

Analysis method

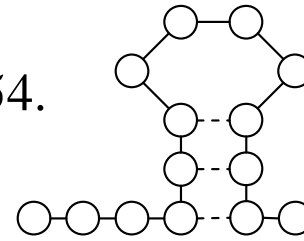
Data set	Analysis method							
	HMM ^a		HMM (no tree) ^b		HMM (no tree) +2PSEFL ^c		HMM (no tree) + 9PSEFL ^d	
PSEFL only	65.7		65.7		64.7		56.6	
	72.5	79.7	72.5	79.7	82.4	77.8	61.8	81.6
	5.8	95.7	5.8	95.7	23.1	86.8	53.8	75.5
	81.3	65.6	81.3	65.6	67.1	81.2	54.2	78.6
close 3	74.4		68.0		64.7		56.6	
	82.4	88.4	80.4	78.7	81.4	81.2	58.8	82.6
	38.5	94.2	25.0	92.6	36.5	83.3	55.8	74.7
	81.3	74.0	74.2	76.6	63.2	82.5	55.5	78.6
med 5	71.5		64.7		61.2		55.0	
	73.5	89.9	66.7	85.5	63.7	87.4	53.9	82.6
	59.6	89.5	51.9	81.7	55.8	75.1	55.8	73.2
	74.2	74.0	67.7	79.2	61.3	80.5	55.5	77.9
all 7	69.6		61.5		58.9		53.7	
	66.7	91.3	59.8	86.5	58.8	87.0	52.9	82.6
	63.5	84.0	55.8	77.8	53.8	75.1	51.9	72.8
	73.5	77.3	64.5	77.9	60.6	76.6	54.8	76.0

A filogenetikai információ általában javítja a becslés pontosságát

PFold: RNS folding filogenetikai információval

A Knudsen-Hein nyelvtan:

Knudsen & Hein (1999) *Bioinformatics* **15**:446-454.



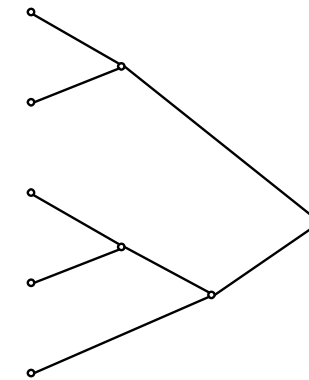
$S \rightarrow SL \mid L$

$L \rightarrow s \mid dFd$

$F \rightarrow S \mid dFd$

$S \rightarrow LLLLL \rightarrow sssdFds \rightarrow sssdddSdddS \rightarrow sssdddLLLLdddS$
 $\rightarrow sssdddssssdddS$

AAAGACGACAUCAUGA---UACG
 CACGACGACGUCAAG-----ACG
 C--GACCACGUCAUGACGGUACG
 AACGACGA---CAUGAUCCUACC
 AACUAUGA-----AUGCUGGUACG



PFold: eredmények I.

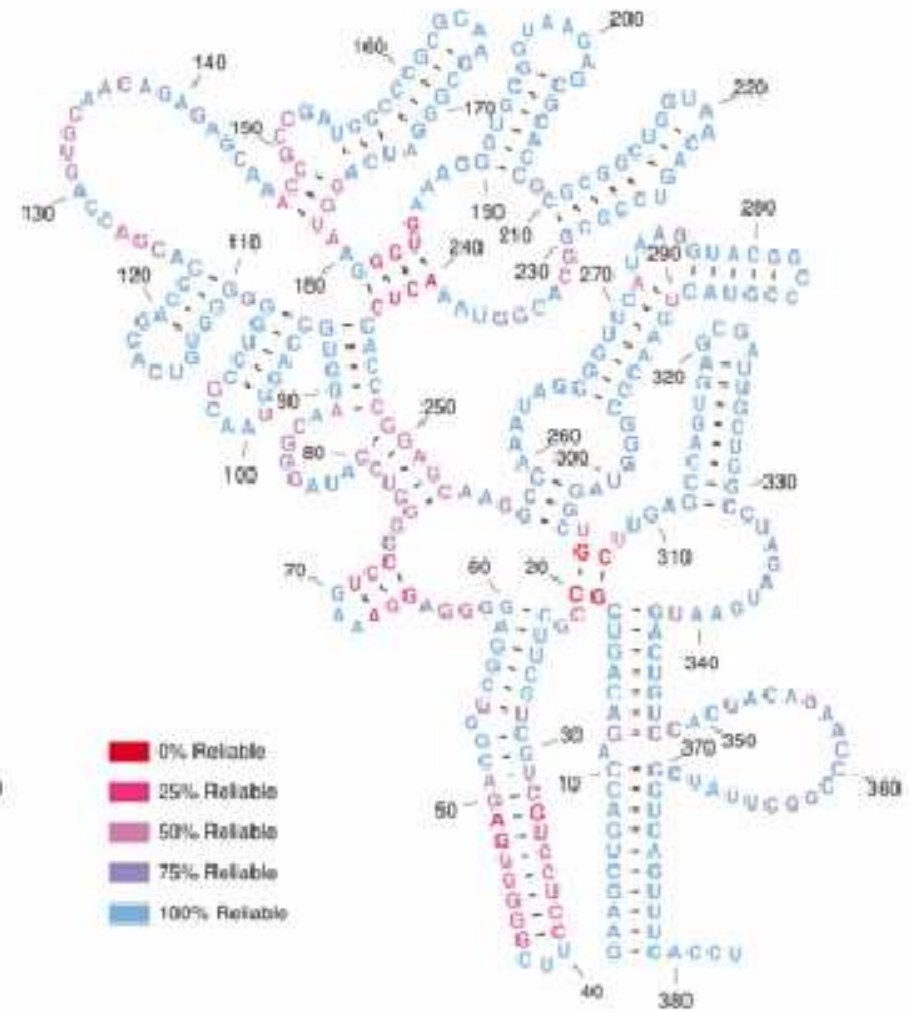
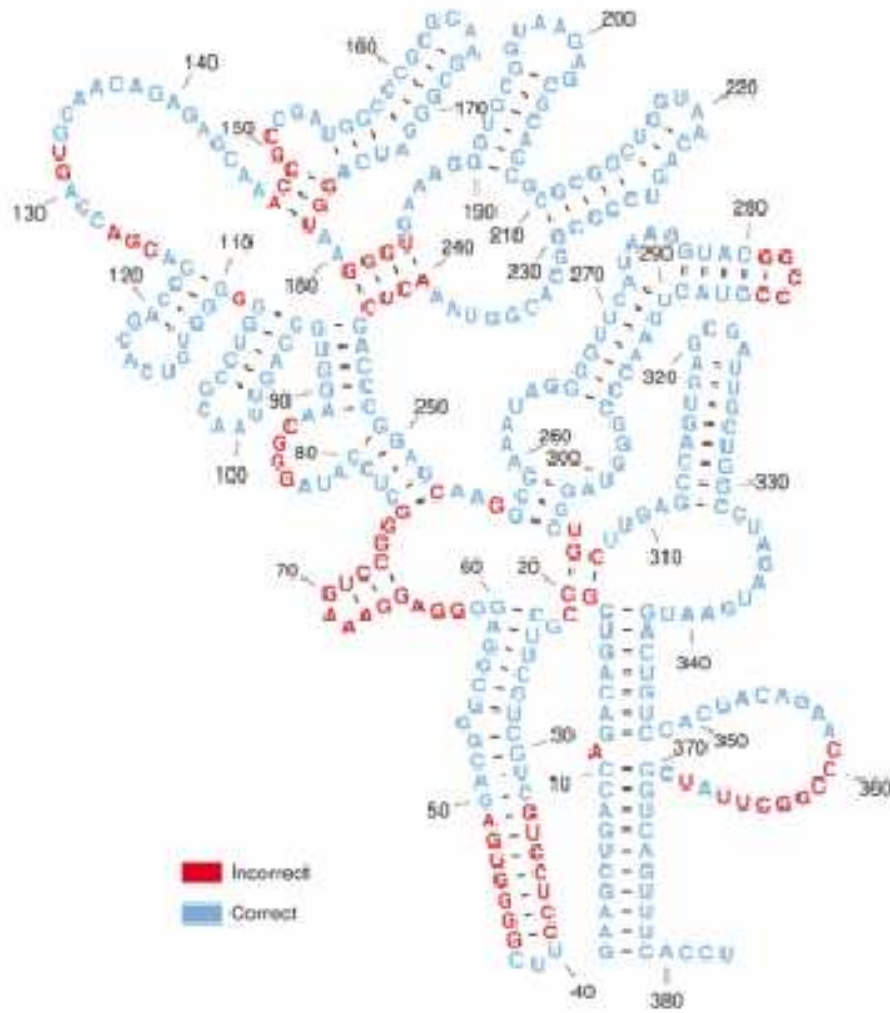
No. of sequences	Structural alignment			
	1	2	3	4
Min result	41.2%	65.2%	73.9%	79.2%
Max result	57.7%	82.1%	79.6%	79.2%
Average	48.3%	73.6%	77.8%	79.2%

No. of sequences	CLUSTAL W alignment			
	1	2	3	4
Min result	41.2%	54.9%	60.1%	73.8%
Max result	57.7%	69.1%	76.9%	73.8%
Average	48.3%	64.4%	68.5%	73.8%

No. of sequences	Structural alignment, no phylogeny			
	1	2	3	4
Min result	41.2%	59.9%	67.7%	76.2%
Max result	57.7%	76.6%	76.6%	76.2%
Average	48.3%	68.9%	72.2%	76.2%

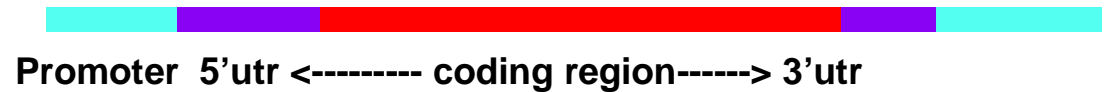
PFold: eredmények II.

Knudsen, B. and Hein, J. (2003) *NAR*, 31:3423-3428.

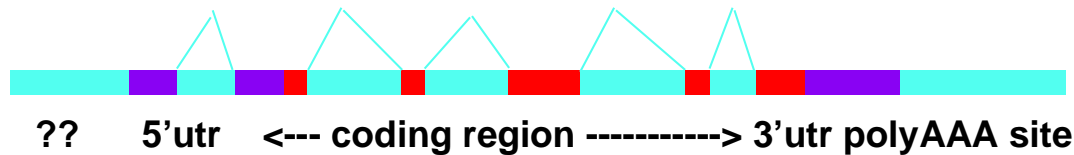


Génkeresés/annotáció

A feladat megkeresni a kódoló régiókat a genomban és annotálni ezeket



Bacterial gene: continuous coding region, known signals

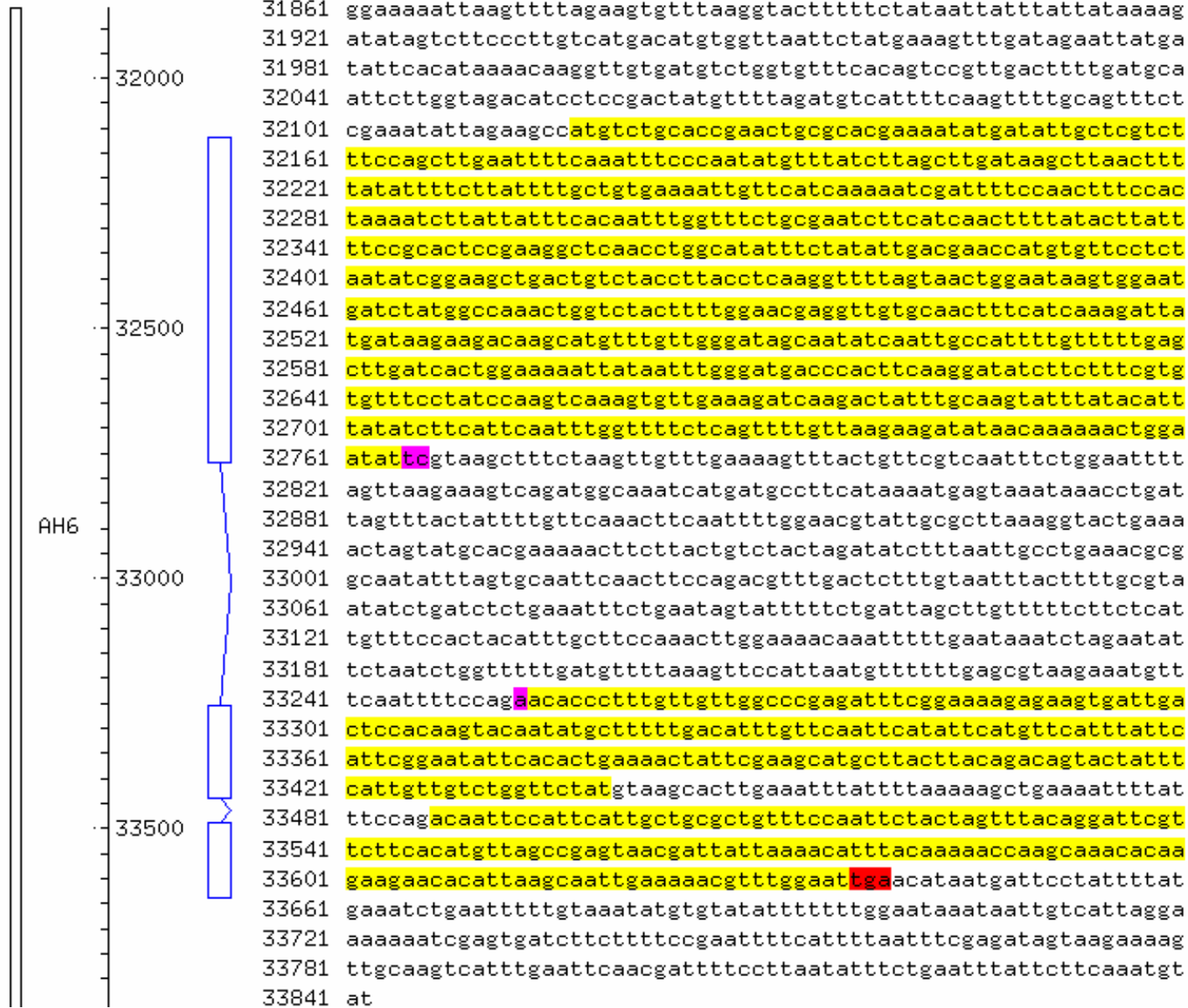


Human gene: fragmented coding region, unknown signals, contained in much more DNA

AHG

32000
32500
33000
33500

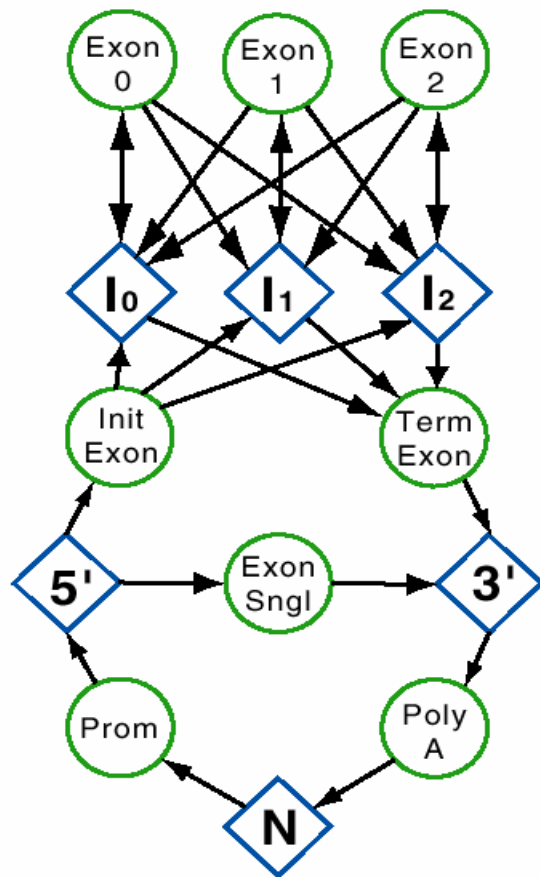
31861 ggaaaaattaagttttagaagtgtttaaggtaactttttctataattatctattataaaaag
31921 atatagtcctcccttgcatgacatgtggtttaattctatgaaagtttgatagaaattatga
31981 tattaacataaaaacaaggcttgatgtotgggttboacagtcogttgaactttgatgca
32041 attcttggtagacatccctccgactatgttttagatgtcattttcaagtttgcagttct
32101 cgaatattagaagccatgtctgcaccgaactgcgcacgaaatgatattgctcgtct
32161 ttocagottgaattttcaaatttoccaatagtgttatcttagottgataagottaacttt
32221 tatatctcttattttgctgtgaaaattgttcatcaaaaatcgattttccaactttccac
32281 taaaatottattttccaaatttggtttctgogaantcttoatcaacttttataactbatt
32341 ttccgcaactccgaaggctcaacctggcatatttctatattgacgaaccatgtgttctct
32401 aatatacgaagctgactgctaccttacctcaaggcttttagtaactggaatagtggaat
32461 gatctatggccaaactggctacttttggaaagaggttgctgcaactttcatcaagatta
32521 tgataagaagacaagcatgtttgttgggatagcaatatcaattgccattttgttttgag
32581 cttgatcaactggaanaattataatttgggatgaccacttcaaggatctctcttctgtg
32641 tgtttcctatccaagtcaaaagtgttgaagatcaagactatttgcaagtatttatcatt
32701 tatatcttcaatttggttttctcagttttgttaagaagatatacaaaaaactgga
32761 atattcgtaagctttctangttgtttgaaaagtttaactgttcgtoaatttctggaatttt
32821 agttaagaaagtcagatggcaaatcatgatgccttcataaaatgagtaaaataaacctgat
32881 tagtttactatatttgttcaaaacttcaattttggaacgtattgcgcttaaaggtaactgaaa
32941 actagtatgcacgaaaacttcttactgtctactagatatcttcaattgcctgaaacgag
33001 gcaatatttagtgcaatcaacttccagaogttgactctttgtaattbaacttttgogta
33061 atatctgatctctgaaatttctgaatagtattttctgattagcttgttttcttctcat
33121 tgtttccactacatttgccttccaaacttggaaaacaaatttttgaataaatctagaatct
33181 tctaatotggtttttgatgttttaagttccattaatgttttttgagogtaagaaatggt
33241 tcaatttccagaacacctttgttgttggccogagatttcggaagagaagtgattga
33301 ctccacaagtaacaabatgcttttgaacatttgttcaattcattatcattgctcatttatto
33361 attcggaaatattcacactgaaaactattcgaagcatgcttacttacagacagtaactattt
33421 cattgttctcggttctatagtaagcaacttgaatttatttttaaaaagctgaanaatttat
33481 ttocagadaattccattcattgctgcogctgttccaattctactagtttaoaggattogt
33541 tcttcacatgttagccgagtaacgattattaaaacatttcaaaaaaccaagcaaacacaa
33601 gaaagaacccatgaagcaattgaaaacgtttggaattgaacataatgattcctattttat
33661 gaaatctgaatttttgtaaatatgtgtatattttttggaataaataattgtcatttagga
33721 aaaaaatcgagtgatctcttttccgaattttcatttttaatttcgagatagtaagaaaag
33781 ttgcaagtcatttgaattcaacgattttccttaattttctgaatttatttccaantgt
33841 at



Rejtett Markov modellek génkeresésre

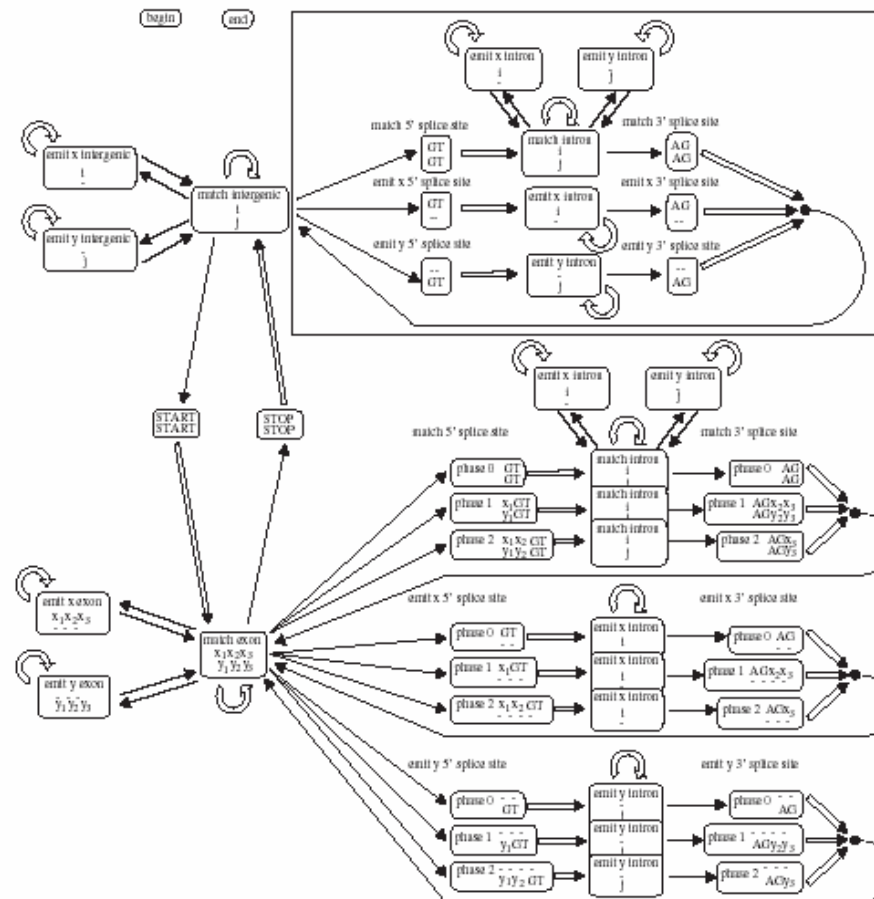
Genescan (HMM)

Burge & Karlin (1997) *J. Mol. Biol.* **268**:78-94.



Doublescan (pair-HMM)

Meyer & Durbin (2002) *Bioinformatics* **18**:1309-1318



Genescan-Doublescan

	DOUBLESCAN without UTR-splicing	DOUBLESCAN	DOUBLESCAN including post-processing	GENSCAN
Gene				
Sensitivity	0.51	0.57	0.57	0.47
Specificity	0.35	0.43	0.50	0.46
Genes overlapping	0.42	0.44	0.46	0.53
Genes missing	0	0	0.01	0
Genes wrong	0.23	0.14	0.04	0.01
Start Codon				
Sensitivity	0.77	0.78	0.75	0.73
Specificity	0.64	0.67	0.78	0.91
Stop Codon				
Sensitivity	0.86	0.91	0.89	0.88
Specificity	0.70	0.74	0.86	0.97
Exon				
Feature Level				
Sensitivity	0.79	0.81	0.80	0.84
Specificity	0.68	0.74	0.79	0.82
Exons overlapping	0.16	0.15	0.15	0.12
Exons missing	0.03	0.03	0.05	0.03
Exons wrong	0.16	0.10	0.06	0.06
Nucleotide Level				
Sensitivity	0.97	0.97	0.96	0.98
Specificity	0.97	0.98	0.99	0.94

Tartalom

- Mi is az összehasonlító bioinformatika?
 - Fehérjék másodlagos térszerkezetének predikciója
 - RNS-ek másodlagos térszerkezetének predikciója
 - Génkeresés
- **Kihívások a bioinformatikai modellezésben**
 - A mutációk egymástól nem függetlenek
 - CpG szigetek
 - Térszerkezetfüggő szubsztitúciók
 - Genomátrendeződés mitochondriumban
 - „Computationaly hard” problémák
 - Pseudoknotok
 - Transzpozíciók
 - Hosszú beszúrás-törlés
 - Összetett modellek
 - Átfedő gének
 - RNS térszerkezet + kódolás mRNSben
 - Kotranszkripcionális folding
- Matematikai kihívások
 - Markov lánc Monte Carlo
 - Statisztikai problémák
 - Algoritmuseleméleti problémák

Life is complicated

CpG szigetek

A metilált citozin könnyebben mutálódik, mint a metilálatlan!

A szomszédhatás tovaterjed → kompikált modell!

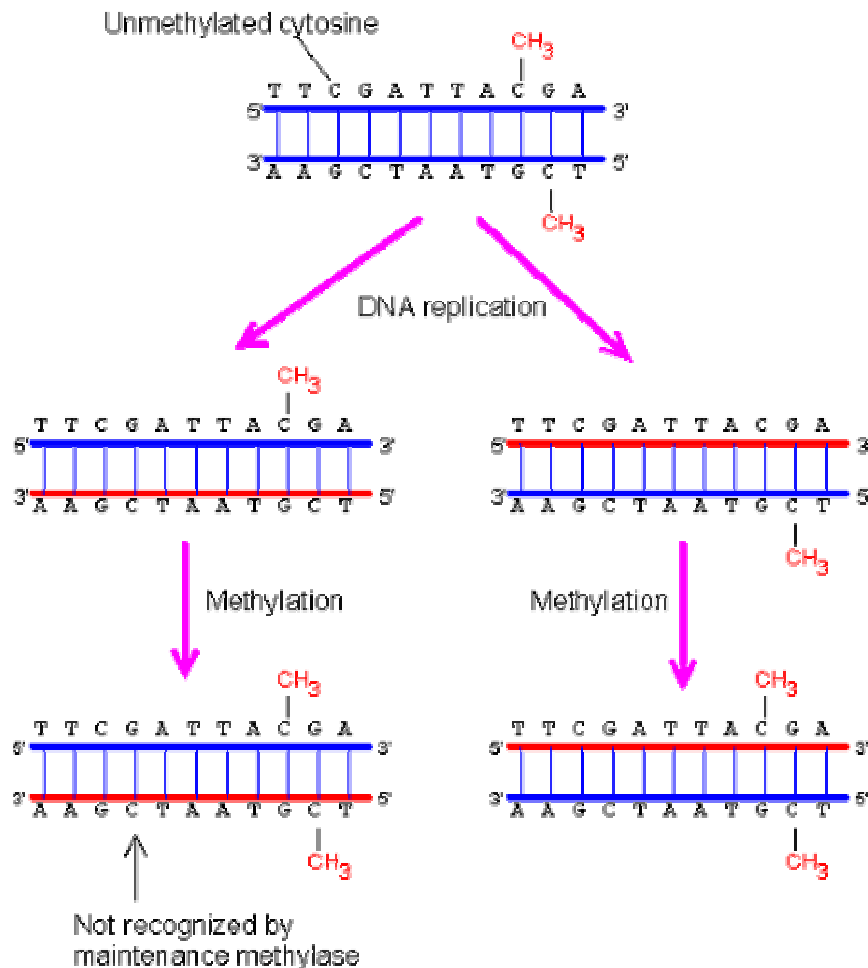
Jensen, Pedersen (2000) *Adv. in Appl. Probab.* **32**:499–517.

Ising modelhez hasonló model, MCMC trajectory sampling.

Lunter & Hein (2004) *Bioinformatics* **20**:i216-i223.

Kombinatorikus approximáció + MCMC

A szomszédhatás fontos, nem csak CpG mintázatra



Szubsztitúció 3D térszerkezettel

Robinson et al. (2003) Mol. Biol. Evol. **20**:1692-1704.

Szubsztitúciós modell:

$$R_{i,j} = \begin{cases} u\pi_h & \text{syn. transverzio} \\ u\pi_h\kappa & \text{syn. tranzicio} \\ u\pi_h\omega e^{(E_s(i)-E_s(j))s+(E_p(i)-E_p(j))p} & \text{non - syn. transverzio} \\ u\pi_h\kappa\omega e^{(E_s(i)-E_s(j))s+(E_p(i)-E_p(j))p} & \text{non - syn. tranzicio} \end{cases}$$

Ahol

$E_s()$: solvent accesibility

$E_p()$: szekvencia-struktúra összehasonlíthatóság

κ, ω, u : hagyományos kodonmodell

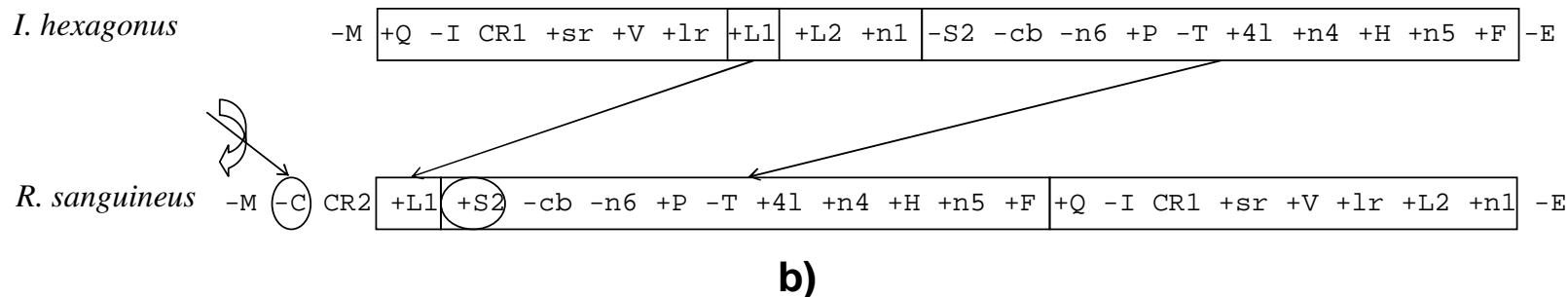
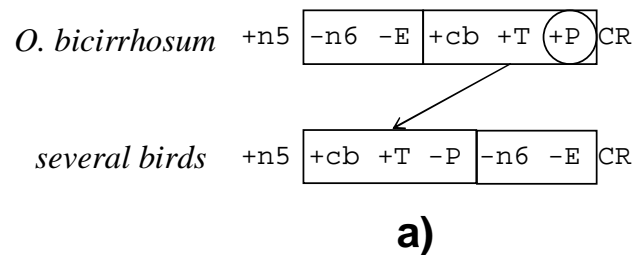
π_h : a j kodonban változó h nukleinsav egyensúlyi gyakorisága

MCMC trajectory sampling

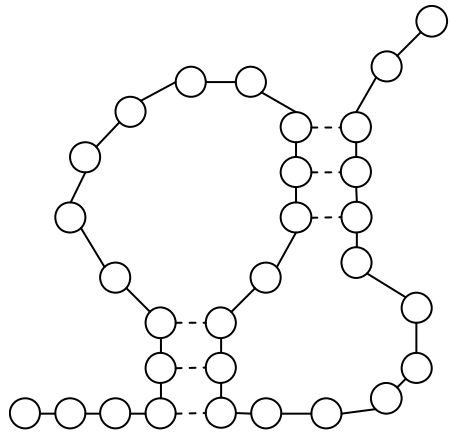
Genomátrendeződés mitochondriumban

Miklós & Hein (2004) *Lecture Notes in Bioinformatics* to appear.

A mutációs ráta nem egyenletes a mitochondriumban, a kontrol régió (CR) környékén gyakoribbak a mutációk



Pseudoknotok (álcsomók)

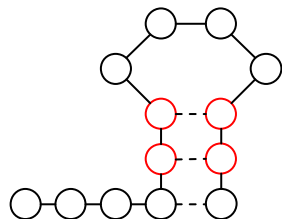


Álcsomókkal rendelkező RNS-ek

- Group I, Group II intronok
- RNas-P
- 16S rRNS

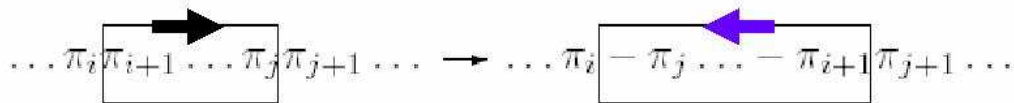
Az álcsomók algoritmuselmélete:

- MWM: polinomiális algoritmus, de nem vesz figyelembe semmilyen sztérikus kényszert
- Speciális osztályokra vannak polinomiális algoritmusok Tree-adjointing grammars, stb.
- Ha a stacking loopok számát akarjuk maximalizálni: NP teljes probléma



Genomátrendeződések

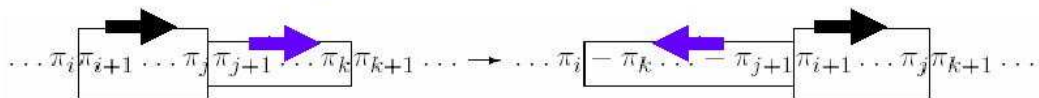
Inversions



Transpositions



Inverted Transpositions



Inverziók:

Előjeles permutációkban
 $O(n^2)$ időben legrövidebb mu-
 tációsorozat

$O(n)$ időben a sorozat hossza

Előjelek nélkül
 NP-teljes probléma!

Transzpozíciók: Ismeretlen komplexitású!

Összetett modellek: Ismeretlen komplexitású!

Inverziós medián: NP teljes előjeles permutációkkal is!

Hosszú beszúrás-törlés modell

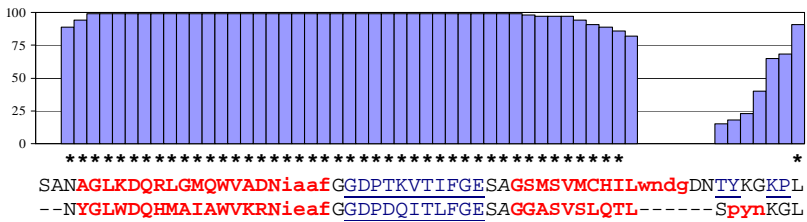
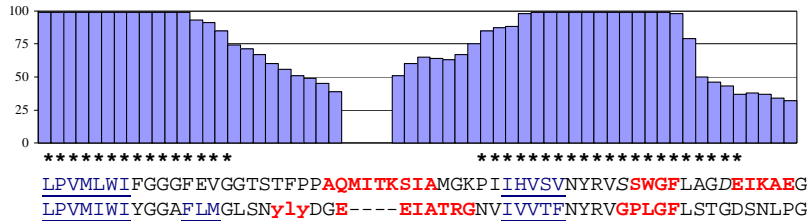
- Adott súlyfüggvény mellett a probléma már régen megoldott (Waterman et al. 1976). Speciális súlyfüggvényekre léteznek hatékonyabb algoritmusok (Gotoh, 1982; Galil & Giancarlo, Eppstein, '80 évek vége)
- A sztochasztikus modellezése azonban bonyolult

$$\begin{aligned} \frac{\partial P(\xi, t)}{\partial t} &= \frac{\partial P(\xi, t)}{\partial \xi} \left[\lambda \frac{\xi(1-\xi)}{1-(1-r)\xi} - \mu \frac{r(1-r)\xi}{1-r-\xi} \right] + \\ &+ P(\xi, t) \left[\lambda \frac{1-\xi}{1-(1-r)\xi} - \mu \frac{(1-r)(1-\xi)^2}{(1-r-\xi)^2} \right] + P(1-r, t) \mu \frac{(1-r)(1-\xi)^2}{(1-r-\xi)^2} \end{aligned}$$

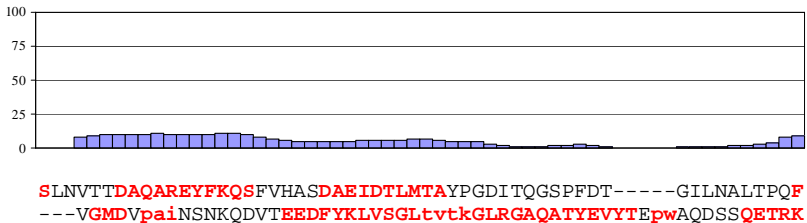
$$P(\xi, 0) = 1$$

A szükséges átmeneti valószínűségeket $\partial P(\xi, t)$ ξ szerinti Taylor sorának együtthatói adják meg. (Miklós (2001) doktori értekezés)

Approximáció: Trajectory likelihood + Dinamikus programozás



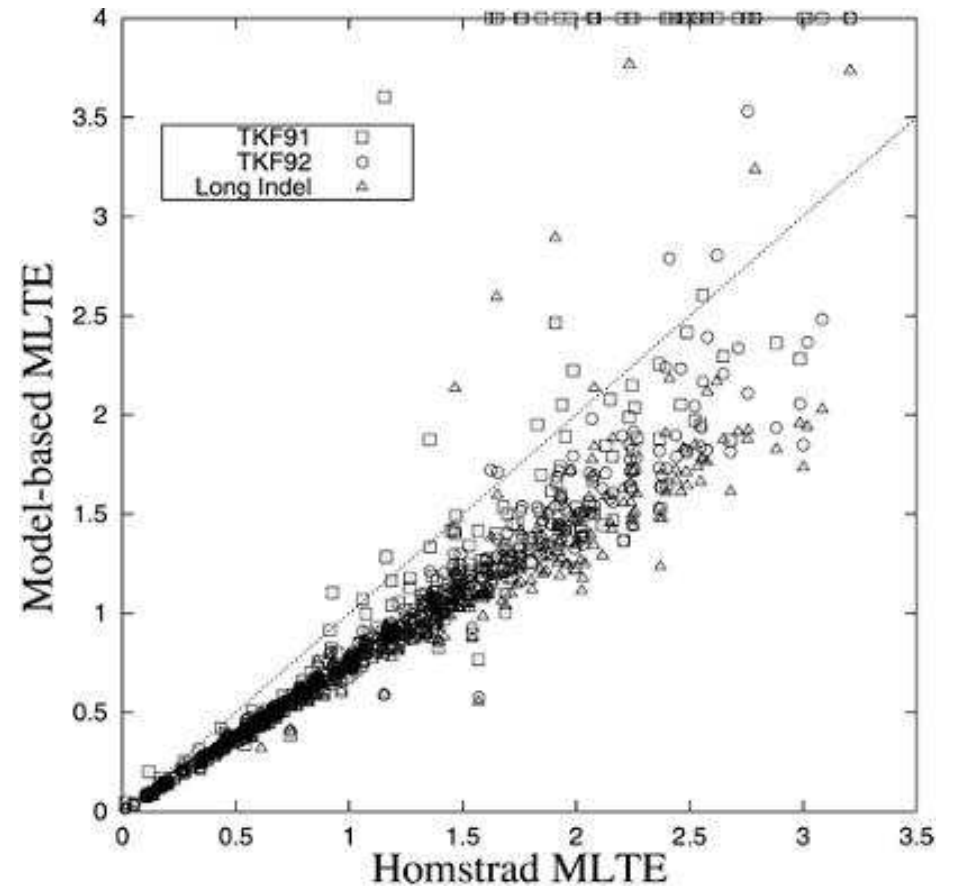
(a)



(b)

GtssLNVT-----TDAQAREYFKQS FVHASDAEIDTLMTAYPGDITQGS PFDTGILNALTE
 VGMDVpaiNSNKQDVT EEDFYKLVSGL-tvtkGLRGAQATYEVYTEpwA-----QDSSQE

(c)

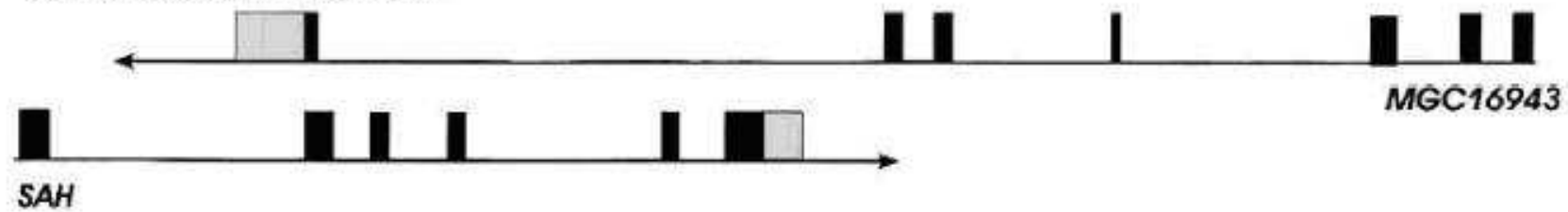


DMGPQPRAEASWQFFMS-DKPLRLAVSL

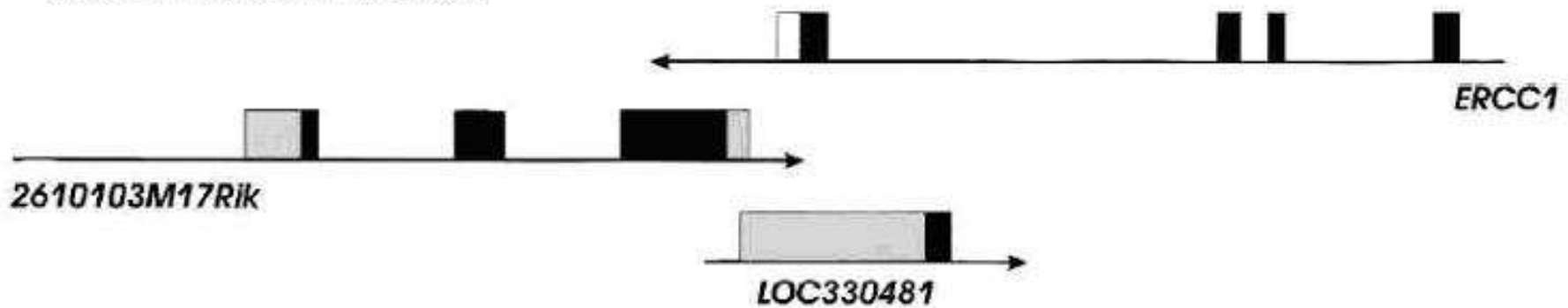
--PGPQPTAETTRQFLMSDKPLHLEASL

Átfedő gének

SAH locus in human

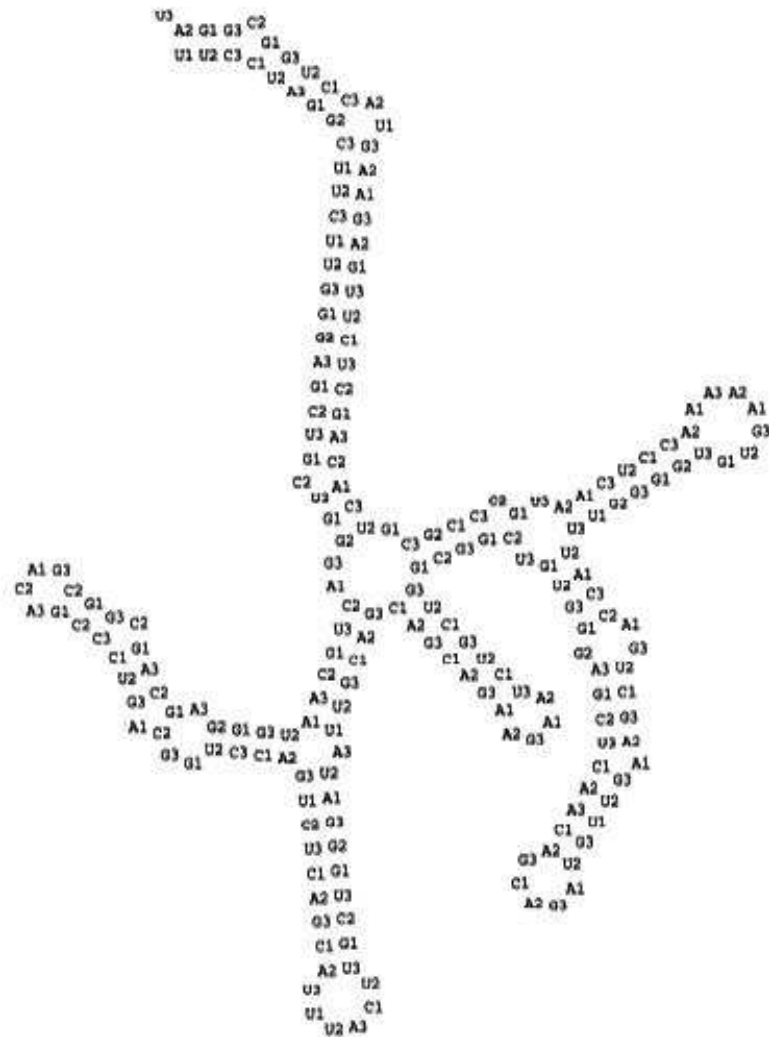


ASE-1 locus in mouse



Fekete: kódoló régió, Szürke: át nem íródó régió

Térszerkezet a mRNS-ben



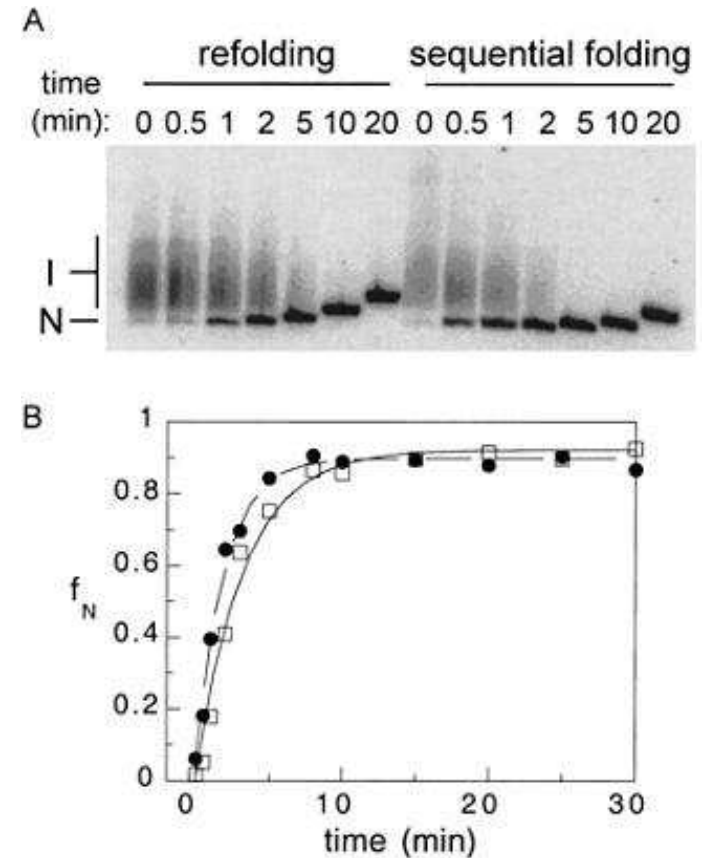
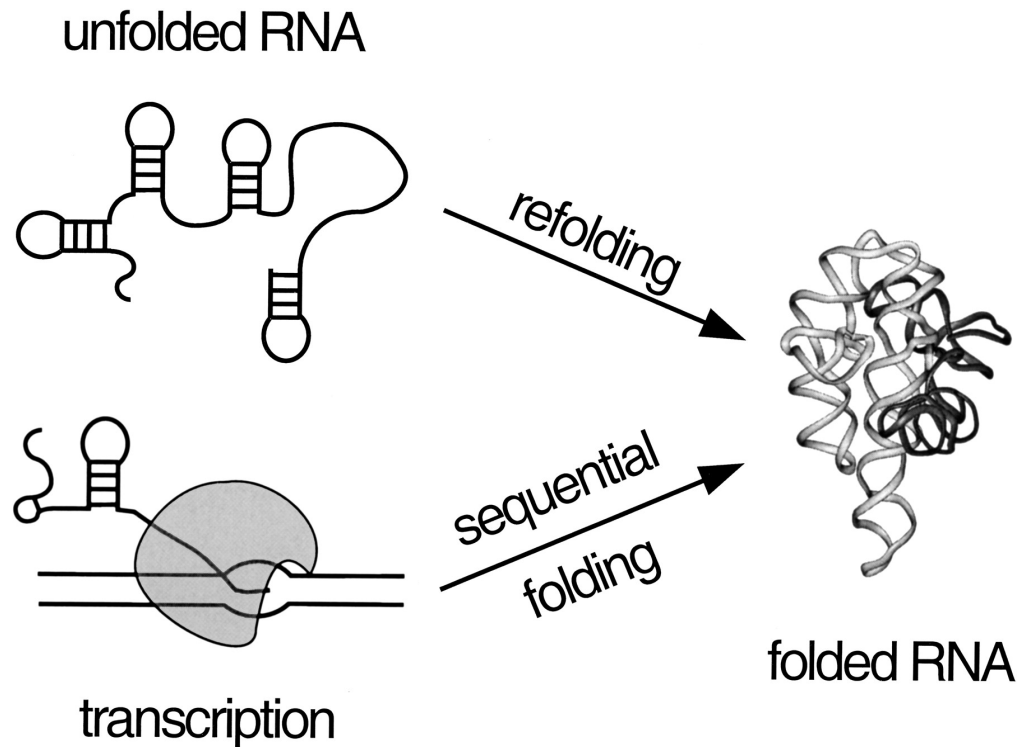
Konecny et al. (2000) *J. Mol. Evol.* **50**:238-242.

SIVAGM 155 RRE gén (lentivírus)
Retrovírus, azaz a vírus RNS-ben
tárolja az információt, ez íródik
vissza DNS-sé (reverz transzkripció)
→Az RNS térszerkezet fontos.

A mutációk befolyásolják mind a
kódolást, mind az RNS térszer-
kezetet!

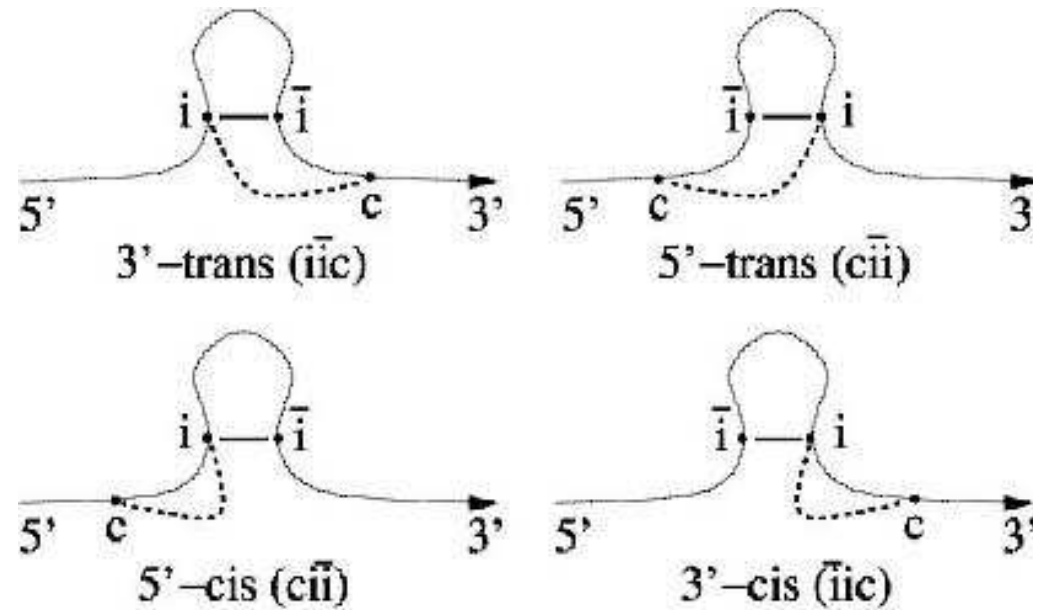
Kotranszkripcionális folding

Heilman-Miller & Woodson (2003) *RNA* **9**:722-733.



Meyer & Miklós (2004) *BMC Mol. Biol.* 5:10.

Taxonomic unit	all 16S rRNA	23S rRNA	Group I	Group II
Data set A				
Archea	28	22	6	0
Bacteria	277	232	45	0
Eukaryotes	41	35	6	0
Chloroplasts	6	6	0	0
Mitochondria	9	9	0	0
Sum	361	304	57	0
Data set B				
Eukaryotes	15	0	0	15
Bacteria	5	0	5	0
Chloroplasts	5	0	5	0
Mitochondria	23	0	17	6
Sum	48	0	27	6



dataset	A		B	
	p-value for t-test	p-value for pos	p-value for t-test	p-value for pos
\overline{Cis}_p	< 0.0001	< 0.0001	0.5733	0.6137
\overline{Cis}_g	< 0.0001	< 0.0001	0.5650	0.6137
\overline{Trans}_p	0.0012	< 0.0001	0.3093	0.8068
\overline{Trans}_g	0.0021	< 0.0001	0.3011	0.5000

Tartalom

- Mi is az összehasonlító bioinformatika?
 - Fehérjék másodlagos térszerkezetének predikciója
 - RNS-ek másodlagos térszerkezetének predikciója
 - Génkeresés
- Kihívások a bioinformatikai modellezésben
 - A mutációk egymástól nem függetlenek
 - CpG szigetek
 - Térszerkezetfüggő szubsztitúciók
 - Genomátrendeződés mitochondriumban
 - „Computationaly hard” problémák
 - Pseudoknotok
 - Transzpozíciók
 - Hosszú beszúrás-törlés
 - Összetett modellek
 - Átfedő gének
 - RNS térszerkezet + kódolás mRNSben
 - Kotranszkripcionális folding
- **Matematikai kihívások**
 - Markov lánc Monte Carlo
 - Statisztikai problémák
 - Algoritmuseleméleti problémák

Markov lánc Monte Carlo (MCMC)

Metropolis, Rosenbluth, Rosenbluth, Teller, Teller (1953) *J. Chem. Phys.* **21**:1087-1091.

$T(Y|X)$ ergodikus Markov lánc, $\forall X$ -re $\pi(X) > 0$ eloszlás, akkor a következő algoritmussal definiált Markov lánc konvergál $\pi(X)$ -hez:

- 1. (proposal) Végy egy random Y -t $T(\cdot|X)$ -ből.
- 2. (acceptance) A Markov lánc következő eleme Y

$$\min \left\{ 1, \frac{T(X|Y)\pi(Y)}{T(Y|X)\pi(X)} \right\}$$

valószínűséggel, és X ennek komplementerével

Első alkalmazások: mintavételezés Boltzmann eloszlásból

Bayes statisztikában szeretik: nem kell a normalizációs konstans ismerni!

$$P(\Theta | D) = \frac{P(D | \Theta)P(\Theta)}{\int_{\Theta'} P(D | \Theta')P(\Theta')} \quad \text{Mintavételezés } P(\Theta | D) \propto P(D | \Theta)P(\Theta) \text{-ből}$$

MCMC a bioinformatikában

Központi kérdés: Mi a jó $T(Y|X)$?

- $T(Y|X)$ könnyen számolható kell, hogy legyen
- $T(X|Y)$ is!
- Gyors mintavételezés $T(\cdot|X)$ -ből.
- A 'jó' megoldások legyenek szomszédok

Bonyolult térben akarunk mintavételezni:

- Szekvenciaillesztés
- RNS térszerkezet (pseudoknotokkal)
- Evolúciós fák
- Evolúciós trajektóriák

Metropolizált részleges fontossági mintavételezés

Vágj ki X -ből egy kis dimenziót, és ezt mintavételezzed függetlenül az aktuális állapottól. A minta egy fontossági eloszlásból kell, hogy származzon, ami közelíti a kívánt eloszlást

Példák a dimenzionálásra:

- A szekvenciaillesztés egy része
- Az RNS térszerkezet egy része
- A trajectória egy része
- Az evolúciós fa egy része

Példák a fontossági eloszlásokra

- Forward-Backward HMM
- Sztochasztikus iteratív szekvenciaillesztés
- Sztochasztikus mohó algoritmus

Első alkalmazás:

Metzler, Fleißner, von Haeseler, Wakolbinger (2001) *J. Mol. Evol.* **53**:660-669

Részleges Gibbs sampling páros rejtett Markov modellből

Statisztikai kérdések

- Milyen gyors a konvergencia?
 - Elméleti bizonyítása annak, hogy egy módszer gyorsabban konvergál/keveredik, mint egy másik.
 - Estimated Sample Size (ESS): A minta hány független mintavételezésnek felel meg (Csúnyán alulbecsülhető!!!)
 - Autokorreláció (hasonlóan alulbecsülhető)
- Goodnes-of-fit testing
- Priorok jogossága
- Torzítatlanság
- Konzisztencia
- Power
- Robusztusság

Algoritmikai kérdések

- Mik legyenek az X -ek?
 - Mennyi időbe telik kiszámolni $\pi(X)$ -et?
 - Ha az állapottérben csoportokat tudunk definiálni, a mintavételezési variancia *általában* csökken (*Rao-Blackwellizáció*) (Kérdés: Általában vagy mindig? Bizonyítás?)

Példa I.:

Trajectory likelihood vs. Path likelihood: Az egy állapotban való tartózkodás idejei ki vannak integrálva vs. nincs kiintegrálva. Trajectory likelihoodot pontosan számoló algoritmus:

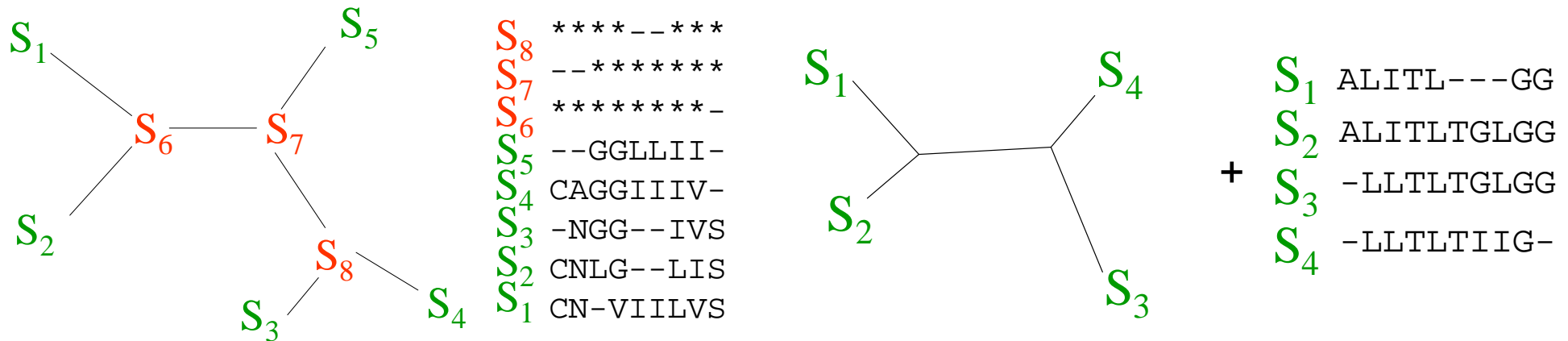
Miklós, Lunter, Holmes (2004) *Mol. Biol. Evol.* **21**(3):529-540.

$O(n^2)$ futási idejű, ahol n az átmenetek száma.

Kérdés: van gyorsabb?

Példa II.:

Szekvenciaillesztés fán. Adatkiterjesztéssel vagy anélkül



Lunter, Miklós, Drummond, Jensen, Hein (2003) *Lecture Notes in Bioinformatics*, **2812**:228–244.

Adatkiterjesztés nélkül is számolható a likelihood, lineáris időben!

Ún. *one-state recursion*, nem igényli a rejtett Markov modell állapotait.

Rejtett Markov Modellel a futási idő $O(\sqrt{5}^n L)$ lenne, ahol n a szekvenciák száma, L az illesztés hossza, a *one-state recursion*-nel $O(nL)$.

Algoritmikai kérdések (folyt)

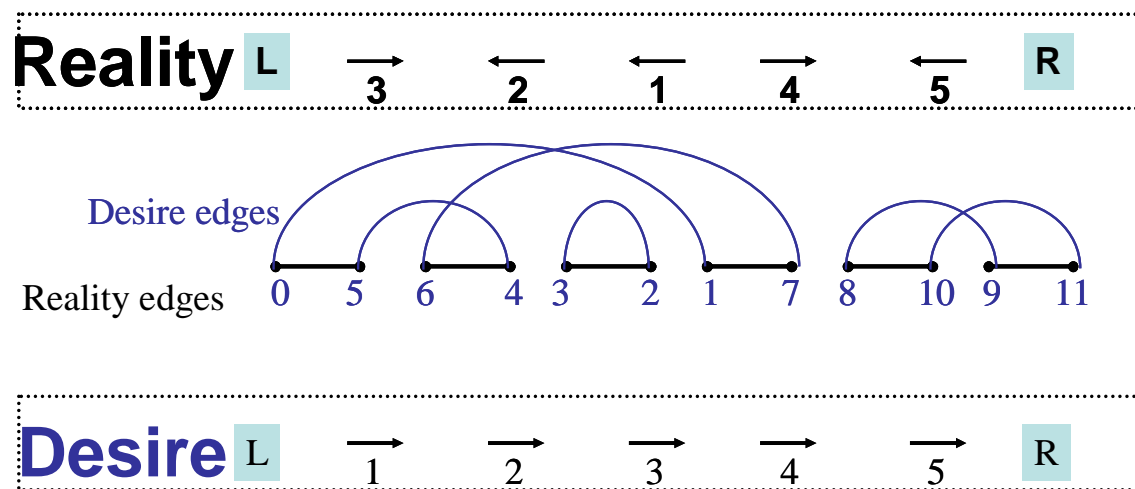
- Mi legyen $T(\cdot|X)$?
 - A 'jó' szomszédok legyenek közel egymáshoz
 - A mintavételezés legyen gyors

Példa:

Trajektória mintavételezés genomátrendeződésre, a trajektóriák inverziókból állnak.

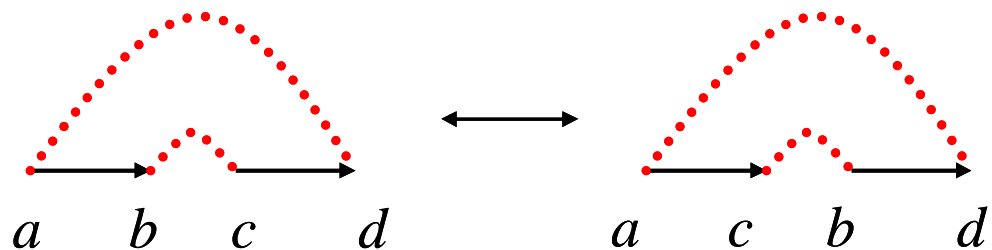
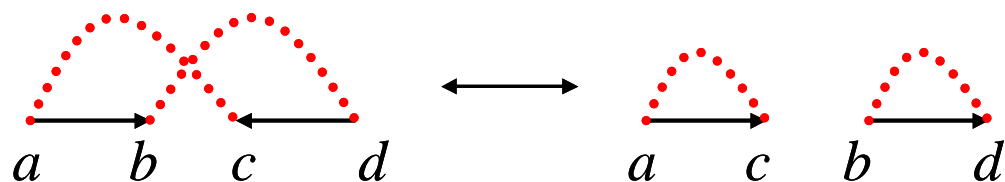
Tudjuk, melyek a rendező inverziók, de az ebből való mintavételezés lassú!

Helyette a körök számának a változását nézzük a kívánalom és valóság gráfjában:



Rendező inverziókon alapuló mintavételezés $O(n^2L)$ idejű, ahol n a gének száma, L a mintavételezett trajektória.

A körök száma alapján való mintavételezés $O(nL)$ idejű, mivel könnyen karakterizálható, mely inverziók csökkentik, növelik, ill. hagyják változatlanul a körök számát.



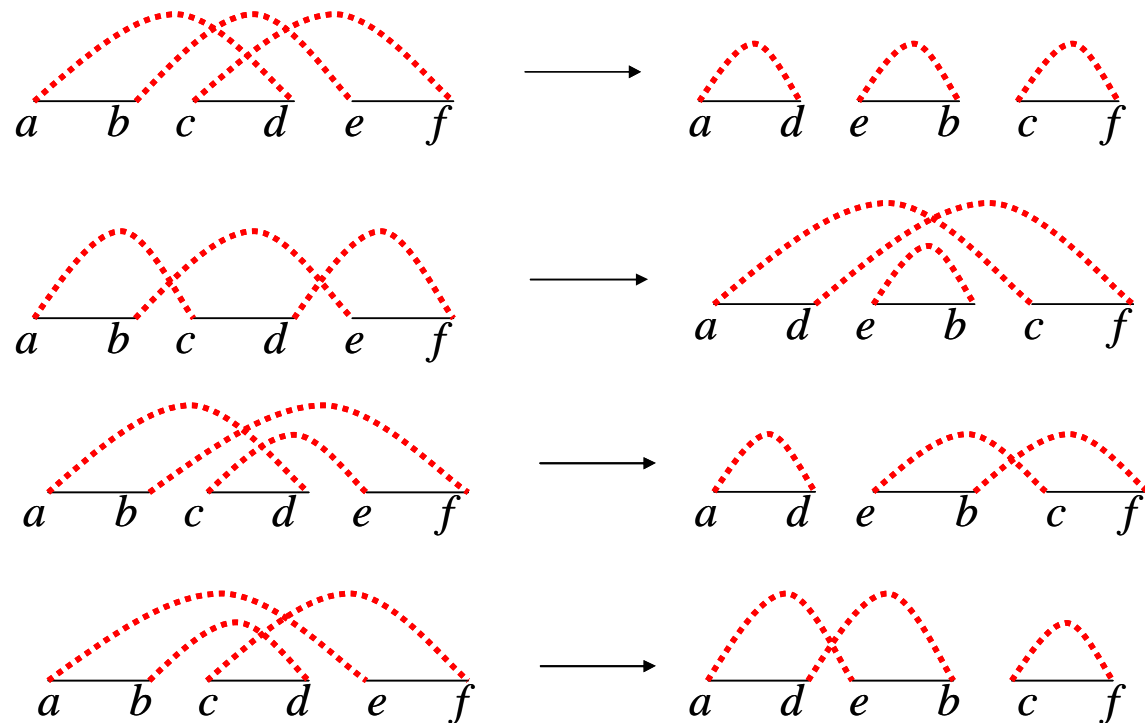
Az algoritmus alapja: súlyozott mintavételezése a köröknek, majd a körökből súlyozott mintavételezése a valóságéleknek, amelyen az inverzió hat.

Miklós, Ittész, Hein (2004) *Bioinformatics* to appear,

Miklós, Hein (2004) *Lecture Notes in bioinformatics* to appear.

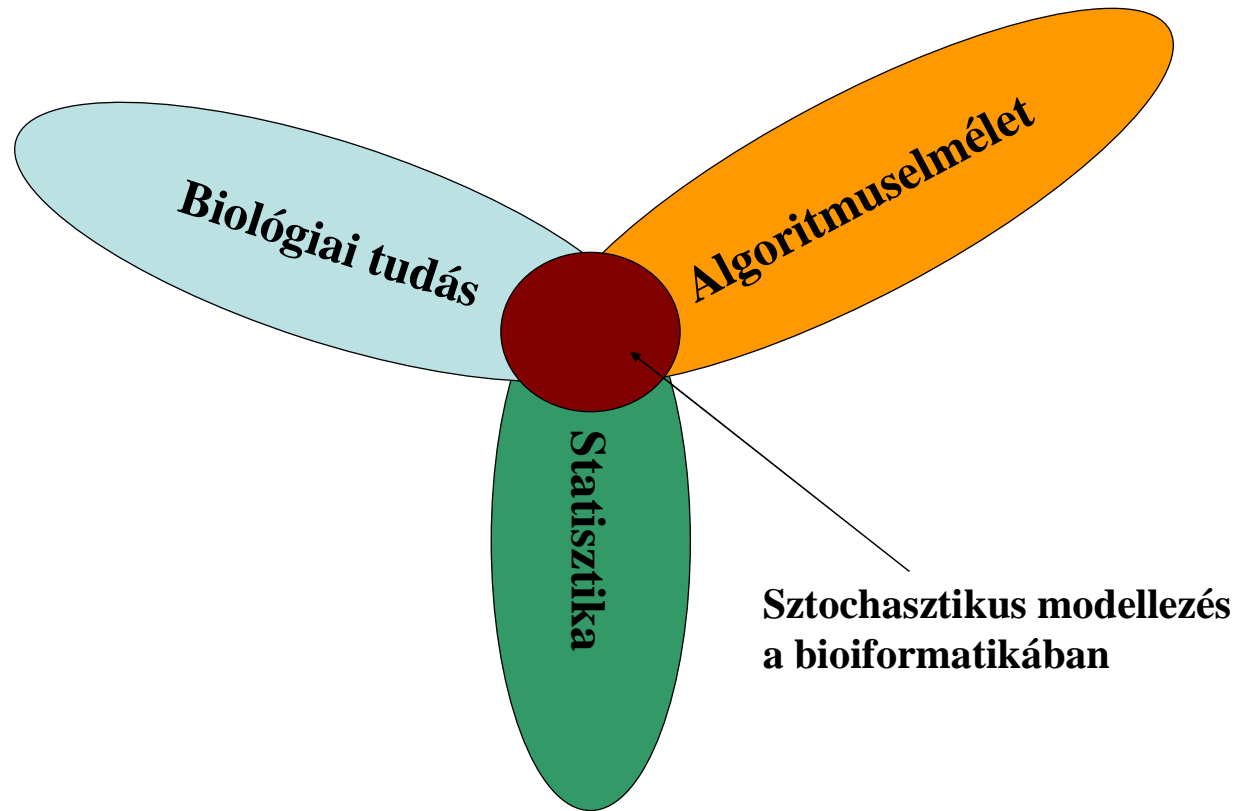
Kérdések:

- Kell-e minden lépésben a köröket újraszámolni, vagy van-e $O(n+L)$ futási idejű mintavételezés?
- Mi a helyzet a transzpozíciókkal? Jelenleg: teljes enumerációja a 'jó' transzpozícióknak, ebből mintavételezés, ill. a rosszakból a *rejection* módszerrel (Miklós (2003) *Bioinformatics*, **19**:ii130–ii137.) Van jobb?



Konklúzió

A modern bioinformatikai modellezés interdiszciplinális:



→ Szükség van összefogásra. Ez a szeminárium remek lehetőség lenne erre!