

Bioinformatica, 29 February 2012

Naam:

Collegekaartnummer:

Question 1 (20 points):

| | |
|----------|------------|
| Elephant | LSP-RSPDED |
| Human | RSPT-SPD-E |
| Bovine | LSP--SPDDE |
| Opposum | RSPAGSPDED |
| Rat | RSP-GSPEED |

Myosin kinase adds a phosphate (PO_4^{3-}) group to myosin molecules. The above alignment shows the active sites (part of the enzyme where substrates bind and are phosphorylated) of few mammalian myosin kinase molecules. *Myosin kinase voegt een fosfaat-groep (PO_4^{3-}) toe aan myosine moleculen. De bovenstaande alignment laat de "active site" (een deel van het enzyme waar een substraat kan binden, welke vervolgens gefosforyleerd wordt) van myosine kinase moleculen van een aantal zoogdieren.*

A) Build the **most biologically likely** Hidden Markov Model (HMM) to detect this active site in myosine kinase molecules. Explain your choice of insertion and deletion states with two sentences.

*A) Maak het **biologisch gezien meest waarschijnlijke** Hidden Markov Model (HMM) om deze "active site" te ontdekken. Verklaar je keuze van insertie en deletie "states" in twee zinnen.*

B) The myosin kinase active site in fish is given below as a sequence logo. Can the HMM you built in A) also recognize the fish myosin kinase active sites? Explain your answer with two sentences.

B) De "active site" van myosine kinase in vissen staat hieronder weergegeven als een sequentie-logo. Kan de HMM die je gemaakt hebt in A) ook de myosine kinase "active sites" in vissen herkennen? Verklaar je antwoord in twee zinnen.



Bioinformatica, 29 February 2012

Naam:

Collegekaartnummer:

Question 2 (20 points):

The below abstract is adapted from Roch & Sherwood, “Stanniocalcin has deep evolutionary roots in eukaryotes”, *Genome Biol Evol.* 2011;3:284-94.

Hieronder staat een samenvatting op basis van het artikel van Roch & Sherwood, “Stanniocalcin has deep evolutionary roots in eukaryotes”, Genome Biol Evol. 2011;3:284-94.

Vertebrates have a large glycoprotein hormone, stanniocalcin, which inhibits calcium uptake from the environment. Performing a series of bioinformatic analyses, we here show that humans, other mammals, and teleost fish have two forms of stanniocalcin (STC1 and STC2) that are widely distributed in many tissues. Investigating evolution of stanniocalcin beyond the vertebrates, we identified novel STC homologs in a diverse array of multicellular and unicellular organisms. Within the eukaryotes, almost all major taxonomic groups except plants and algae have STC homologs, although some groups like echinoderms and arthropods lack STC genes. The critical structural feature for recognition of stanniocalcins was the conserved pattern of ten cysteines, even though the amino acid sequence identity was low. Predicted signal peptides in STC sequences suggest they are secreted from the cell of synthesis. We predict that large secreted stanniocalcin homologs appeared in evolution as early as single-celled eukaryotes.

Name three different bioinformatics analyses these researchers might have performed to obtain these results and explain shortly how they could have performed these analyses.

Noem drie verschillende bioinformatische analyses die de onderzoekers gedaan kunnen hebben om hun resultaten te verkrijgen. Leg kort uit hoe ze deze analyses gedaan zouden kunnen hebben.

Bioinformatica, 29 February 2012

Naam:

Collegekaartnummer:

Question 3 (35 points):

An unrooted phylogenetic tree of the interleukin (IL)-10/20/24 genes was constructed by the neighbor-joining method and is given below (the tree is redrawn from *Hughes & Freidman, Mol. Phy. Evol. 2005, 35(1):186-95*). YLDV, LSDV and SPPV stands for different poxviruses (Yaba-like disease virus, Lumpy skin disease virus, Sheeppox virus, respectively). Fugu is a fish.

Een “unrooted” phylogenetische boom van interleukine-10/20/24-genen was gemaakt volgens de “neighbor-joining”-methode en staat hier beneden (de boom is gereproduceerd van Hughes & Freidman, Mol. Phy. Evol. 2005, 35(1):186-95). YLDV, LSDV en SPPV zijn afkortingen voor verschillende pokken virussen (Yaba-like disease virus, Lumpy skin disease virus, Sheeppox virus, respectievelijk). Fugu is een vis.

A) According to this tree, how old are IL-10/20/24 genes?

A) Hoe oud zijn de IL-10/20/24 genen volgens deze boom?

B) How many IL-10 genes had the common ancestor of mammals and fish?

B) Hoeveel IL-10 genen had de gemeenschappelijke voorouder van zoogdieren en vissen?

C) Indicate where possible gene duplications, and gene losses have occurred.

C) Geef aan waar mogelijke gen duplicaties hebben plaatsgevonden en waar genen mogelijk zijn verdwenen.

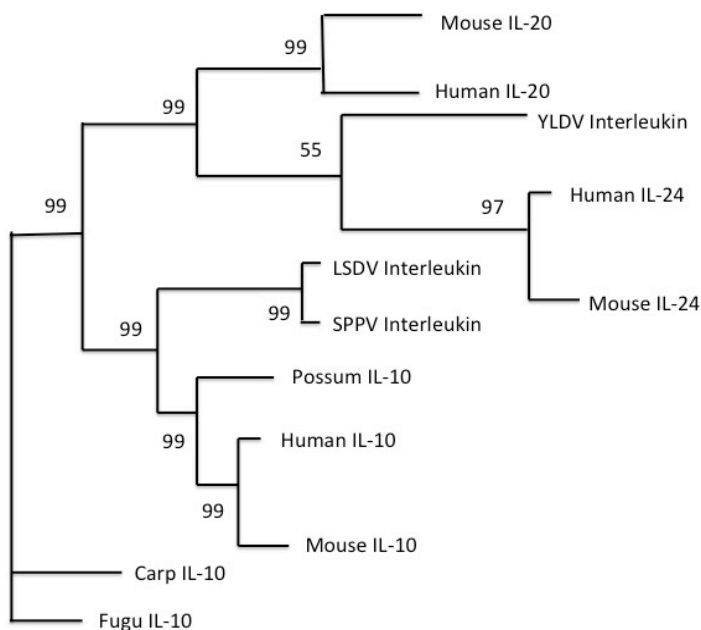
D) Does this tree suggest horizontal gene transfer(s), and if so how many separate events could have happened?

D) Suggereert deze boom dat “horizontal gene transfers” hebben plaatsgevonden? Zo ja, hoe vaak zou dit onafhankelijk kunnen hebben plaatsgevonden?

E) The bootstrap values given in the figure are the result of 100 bootstrap experiments. In how many

(maximum) of the bootstrap trees could we have seen the viral genes form one group without any other genes?

E) De “bootstrap”-waarden in het figuur zijn het resultaat van 100 “bootstrap”-experimenten. In maximaal hoeveel “bootstrap”-bomen kunnen de virale genen in één groep clusteren, zonder andere genen.



Bioinformatica, 29 February 2012

Naam:

Collegekaartnummer:

Question 4 (25pnts):

The expression of genes A-E is measured in four time points during a cell cycle experiment. Below you see the distance matrix generated using the Manhattan distance and a clustering of genes based on gene expression using the complete linkage method.

De expressie van genen A-E is gemeten op vier tijdstippen in een cel-cyclus-experiment. Hieronder zie je een afstands-matrix die gemaakt is met behulp van de "Manhattan"-afstand, en een clustering van de genen op basis van hun expressie volgens de "complete linkage"-methode.

A) Cluster the genes A-E based on their expression using the single linkage method. Indicate the branch lengths.

A) Cluster de genen A-E op basis van hun gen expressie volgens de "single linkage"-methode. Geef de "branch"-lengte aan.

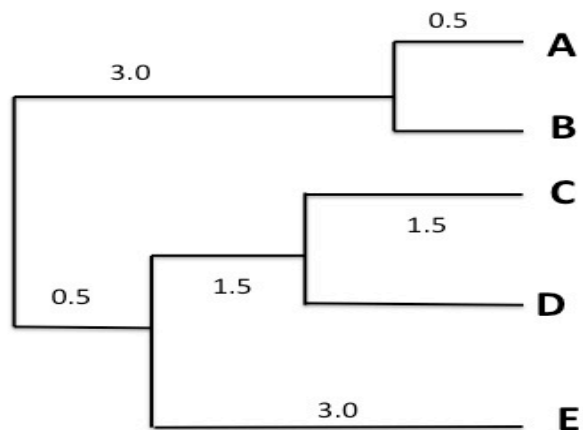
B) If you know already that gene B and E are expressed **only** during the G1 phase of the cell cycle, which clustering (single linkage or complete linkage) seems to be biologically more correct?

*B) Als je zou weten dat gen B en gen E alleen tot expressie komen **gedurende** de G1-fase van de cel cyclus, welke clustering (single linkage or complete linkage) zou dan biologisch meer correct zijn?*

C) Can you conclude from your clustering that gene C and D are results of a gene duplication, ie. that gene C and gene D had once a common ancestor? If yes, explain your answer in max 3 sentences, if no propose a bioinformatics analysis that can test if gene C and D are homologous.

C) Kun je concluderen dat de clustering van de genen C en D het resultaat is van een gen duplicatie, dat gen C en gen D ooit een gemeenschappelijke voorouder hadden? Zo ja, verklaar je antwoord dan in maximaal 3 zinnen. Zo nee, stel een bioinformatische analyse voor waarmee je zou kunnen testen of gen C en gen D homologe genen zijn.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | - | 1 | 5 | 4 | 7 |
| B | | - | 6 | 7 | 2 |
| C | | | - | 3 | 6 |
| D | | | | - | 5 |
| E | | | | | - |

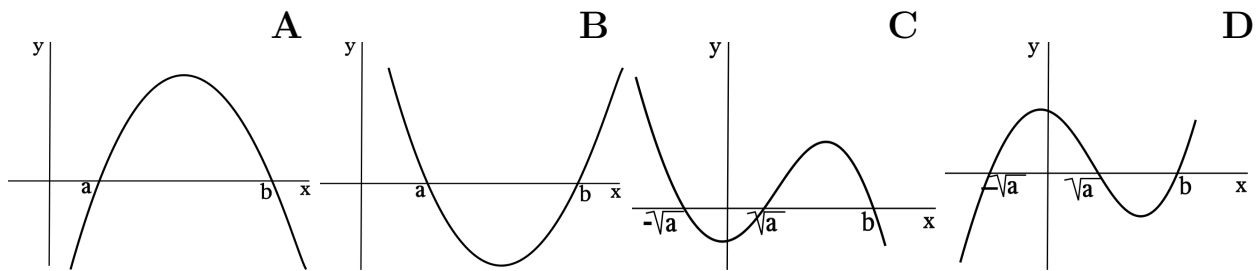


Naam:

Collegekaartnummer:

Vraag 1: Dit zijn multiple choice vragen. Om-cirkel het meest correcte antwoord.

1.1 Beschouw de functie: $y = (a - x^2)(x - b)$, $a < b$;
Welke grafiek komt overeen met deze functie?



- a. Figuur A hierboven
- b. Figuur B hierboven
- c. Figuur C hierboven
- d. Figuur D hierboven

1.2 Beschouw het volgende predator prooi model:

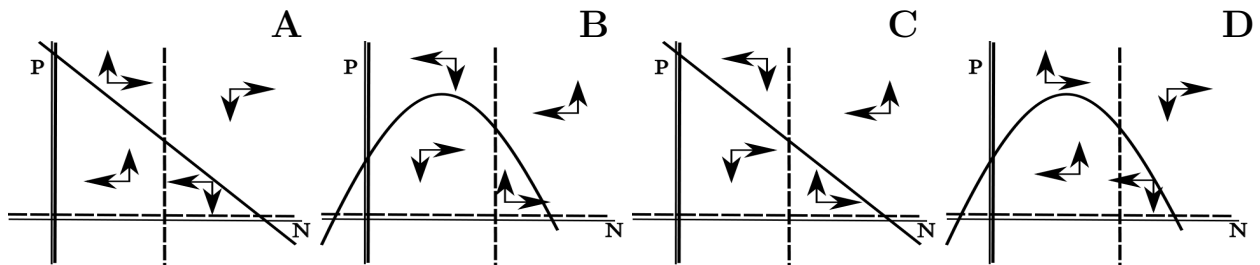
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) - d\frac{NP}{h + N}$$

$$\frac{dP}{dt} = e\frac{NP}{h + N} - fP$$

$$N \geq 0, P \geq 0$$

$$r, K, d, h, e \geq 0$$

Welke van de volgende faseplaatjes hoort bij dit model?



- a. Figuur A hierboven
- b. Figuur B hierboven
- c. Figuur C hierboven
- d. Figuur D hierboven

1.3 Beschouw de volgende matrix:

$$\begin{pmatrix} 2 & -2 \\ -6 & 3 \end{pmatrix}$$

Welke eigenwaardes horen bij deze matrix?

- a. $\lambda_1 = 6, \lambda_2 = 3$
- b. $\lambda_1 = 6, \lambda_2 = -1$
- c. $\lambda_1 = 3, \lambda_2 = -1$
- d. $\lambda_1 = -6, \lambda_2 = 1$

1.4 Beschouw de functie: $f(x, y) = 3xy - 5y^3 + 6x + 2$

Wat is de lineaire benadering van deze functie in het punt $(1, 1)$?

- a. $f(x, y) \approx 6 + 9x - 12y$
- b. $f(x, y) \approx 9 - 9x + 12y$
- c. $f(x, y) \approx 6 + 9x + 12y$
- d. $f(x, y) \approx 9 + 9x - 12y$

1.5 Beschouw de volgende twee complexe getallen: $a = 5 - 3i$ en $b = 6 + 7i$

Wat is $a \times b$?

- a. $51 + 17i$
- b. $30 - 21i^2$
- c. $30 + 17i$
- d. $9 + 17i$

Naam:

Collegekaartnummer:

Vraag 2: Dit zijn multiple choice vragen. Om-cirkel het meest correcte antwoord.

2.1 Een eenvoudig model voor de productie van rode bloedcellen is $dB/dt = m - dB$, waar B het totaal aantal rode bloedcellen in het lichaam is. Dit model heeft een beenmergproductie van m cellen per dag en de cellen hebben een levensverwachting van $1/d$ dagen. De totale productie en de *per capita* sterfte in dit model zijn:

- beide dichtheidsafhankelijk.
- beide dichtheidsonafhankelijk.
- De productie is dichtheidsonafhankelijk en de sterfte is dichtheidsafhankelijk.
- De productie is dichtheidsafhankelijk en de sterfte is dichtheidsonafhankelijk.

2.2 Beschouw de concentratie messenger-RNA, M , en het corresponderende eiwit, P , van een bepaald gen, wat we modelleren met de vergelijkingen:

$$\frac{dM}{dt} = c - dM \quad \text{en} \quad \frac{dP}{dt} = lM - \delta P .$$

Hier is c de transcriptie-snelheid (in moleculen per uur), l de translatie-snelheid (per uur) en d en δ de verval-snelheid van messenger-RNA en eiwit (ook per uur). Als een cel geactiveerd wordt neemt de transcriptie-snelheid c met een factor k toe (d.w.z. c wordt kc). Na verloop van tijd stelt zich een nieuw evenwicht in, waarvoor we verwachten dat

- De concentratie mRNA is toegenomen terwijl die van het eiwit gelijk gebleven is.
- De concentratie eiwit is toegenomen terwijl die van het mRNA gelijk gebleven is.
- De concentratie mRNA neemt meer toe dan die van het eiwit.
- De concentratie mRNA en eiwit nemen allebei met deze factor k toe.

2.3 De Heckrunderenpopulatie in de Oostervaardersplassen bleek zich goed te laten beschrijven met het volgende model

$$\frac{dN}{dt} = N[b(1 - N/k_1) - d(1 + N/k_2)]$$

met een maximaal geboortecijfer van $b = 0.4$ per jaar, een minimaal sterftecijfer van $d = 0.03$ per jaar, en met $k_1 = 800$ en $k_2 = 50$ runderen. De biologische interpretaties van k_1 is:

- k_1 is het aantal runderen waarbij het geboortecijfer gehalveerd is.
- k_1 is het aantal runderen waarbij het geboortecijfer nul is.
- k_1 is het aantal runderen waarbij het sterftecijfer gehalveerd is.
- k_1 is het aantal runderen waarbij het geboortecijfer verdubbeld is.

2.4 Op een Hopfbifurcatie

- sterft van de variabelen uit.
- ontstaat een limiet cycles.
- verdwijnt het evenwicht.
- treedt een catastrofe op.

2.5 Een quasi steady state aanname betekent dat we de desbetreffende variabele

- Constant veronderstellen omdat zij zo snel is.
- Constant veronderstellen omdat zij zo langzaam is.
- In evenwicht veronderstellen omdat zij zo snel is.

d. In evenwicht veronderstellen omdat zij zo langzaam is.

2.6 We hebben een model behandeld voor het Lac-operon van *E. coli*. Dit model vertoonde hysteresis in het intra-cellulaire metabolisme van lactose als de concentratie lactose in het externe milieu gevarieerd werd. Dit betekent dat:

- de bacteriën altijd optimaal zijn aangepast.
- dat bacteriën bij veel lactose in het medium juist weinig lactose opnemen.
- dat bacteriën bij weinig lactose in het medium maximaal lactose opnemen.
- dat verschillende bacteriën in hetzelfde milieu verschillende hoeveelheden lactose kunnen gebruiken.

2.7 In het boek hebben we de interactie tussen twee transcriptie-factoren, P_1 en P_2 (dit zijn twee eiwitten) en hun messengers, M_1 en M_2 , beschreven als:

$$\frac{dM_1}{dt} = \frac{c_1}{h_2 + P_2} - d_1M_1, \quad \frac{dM_2}{dt} = c_0 + \frac{c_2P_1}{h_1 + P_1} - d_2M_2, \quad P_1 = kM_1 \quad \text{en} \quad P_2 = kM_2,$$

waar de c_i parameters voor de transcriptie-snelheden zijn, de d_i parameters verval-snelheden zijn en de h_i parameters verzadiging-constanten. Omdat de turnover van eiwit sneller is dan die van hun messenger nemen we aan dat de concentratie van iedere transcriptie factor proportioneel is aan die van hun messenger ($P_i = kM_i$). Dit model neemt aan dat P_1 en P_2 transcriptie factoren zijn waarbij

- P_1 een activator is en P_2 een repressor.
- P_1 een repressor is en P_2 een activator.
- Beide activatoren zijn.
- Beide repressoren zijn.

2.8 We hebben het volgende model voor een chronische virusinfectie behandeld

$$\frac{dT}{dt} = \sigma - \delta_T T - \beta TV, \quad \frac{dI}{dt} = \beta TV - \delta_I I - kEI, \quad \frac{dV}{dt} = pI - \delta_V V, \quad \frac{dE}{dt} = \alpha EI - \delta_E E,$$

waar T target-cellen, I geïnfecteerde cellen, V virus-partikels en E immune effector-cellen zijn. De set-point virus load, \bar{V} , in het evenwicht is:

- $\bar{V} = \frac{p\delta_E}{\alpha\delta_V}$
- $\bar{V} = \frac{p}{\delta_V} I$
- $\bar{V} = pI - \delta_V V$
- $\bar{V} = \frac{\sigma}{\beta T} + \frac{\delta_T}{\beta}$

Naam:

Collegekaartnummer:

Vraag 3.

Beschouw het volgende systeem van twee differentiaalvergelijkingen:

$$\frac{dx}{dt} = x(25 - x^2 - y)$$

$$\frac{dy}{dt} = y(x - 3)$$

$$x \geq 0, y \geq 0$$

- a. Teken het faseplaatje (null clines en vectorveld) voor dit systeem
- b. Bepaal (indien mogelijk) het type en de stabiliteit van de evenwichten
- c. Bepaal de coördinaten (x en y waarden) van de evenwichten
- d. Bepaal de algemene Jacobiaan van dit systeem
- e. Bepaal de Jacobiaan in elk evenwicht en gebruik hem om type en stabiliteit van het evenwicht te bepalen

Naam:

Collegekaartnummer:

Vraag 4.

Epitheelcellen, E , bekleden de maagwand en delen zichzelf als de bekleding van de maagwand dun wordt. Als er k epitheelcellen zijn is de maagwand ruimschoots bekleed en delen de epitheelcellen niet meer. Infectie van de maag met de bacterie *Helicobacter pylori*, B , tast de maagwand aan door infectie (β) van epitheelcellen. Geïnfecteerde epitheelcellen, I , dragen niet langer bij aan de bekleding van de maagwand en produceren bacteriën B . Beschouw het volgende model

$$\frac{dE}{dt} = E(b(1 - E/k) - d) - \beta EB, \quad \frac{dI}{dt} = \beta EB - \delta I \quad \text{en} \quad \frac{dB}{dt} = pI - cB,$$

waar p de dagelijkse productie van *Helicobacter pylori* is per geïnfecteerde cel en d , δ , en c sterfte-snelheden zijn.

- a. Wat is de dimensie van de parameter b ?
- b. Hoeveel epitheelcellen bekleden een gezonde maag?
- c. Hoeveel epitheelcellen zitten er in een chronisch geïnfecteerde maag?
- d. Medicijnen die de zuurgraad verlagen verlengen de levensverwachting van gezonde epitheelcellen. Verbeteren deze middelen de bekleding van de maagwand tijdens een chronische infectie?
- e. Antibiotica verhogen de sterfte-snelheid c van *Helicobacter pylori*. Verbetert het slikken van antibiotica de bekleding van de maagwand?
- f. Vereenvoudig het model door gebruikt te maken van het feit dat de tijdschaal van de bacteriën veel sneller is dan die van de cellen. Schrijf de nieuwe differentiaal-vergelijkingen.
- g. Teken de nulisoclines van het vereenvoudigde model aannemende dat de infectie zich kan handhaven. Bepaal de stabiliteit van het niet-triviale evenwicht.

