

Písemná zkouška obsahuje 10 otázek s krátkou odpovědí (vždy 3 body) a 5 otázek s obširnější odpovědí (vždy 10 bodů). Nejvyšší počet bodů, který můžete získat je tedy 80. K absolvování zkoušky je spolu se semestrálním testem a bonusem potřeba získat nejméně 50 bodů (stupeň hodnocení E), na stupně hodnocení D, C, B a A jsou stanoveny minimální počty bodů 60, 70, 80 a 90. Na vypracování odpovědí máte 90 minut. Na poslední stránce máte k dispozici tabulku genetického kódu.

ČÁST KRÁTKÝCH ODPOVĚDÍ

B01 Jak dlouhá by musela být sekvence oligonukleotidu, aby se s největší pravděpodobností vyskytoval v lidském genomu jen jednou (pro zjednodušení považujte lidský genom za náhodnou sekvenci nezávislých, stejně zastoupených nukleotidů) ?

B02 Co je to čtecí rámeček? Co znamená, že je otevřený? Proč ORF vyhledáváme?

B03 Zarovnáme sekvence AGTCA a ACTTA globálně algoritmem dynamického programování. Skóre za shodu je 1, neshodu -1, mezeru -2. Doplňte tabulku. Jak poznáme stupeň podobnosti těchto sekvencí?

	A	C	T	T	A	
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1				
T	-6					
C	-8					
A	-10					

B04 Co je to EST sekvence? Jaký má význam je shromážďovat?

B05 Jsou dány tyto mnohonásobně zarovnané sekvence

A: ACCTG GTTAC GGATC TTATA

B: AGCTG ATTAC AAATC TTAGA

C: AGCTG GTTAC AGATC TCATA

D: AGCTG CTTAG GGATC TCATA

Jakým jednoduchým způsobem můžeme odhadnout evoluční vzdálenosti mezi nimi pro účely

tvorby jednoduchého fylogenetického stromu typu UPMGA? Odhad vykonejte a načrtněte odpovídající strom.

B06 Z jakého materiálu jsou obvykle sondy na microarray nebo DNA čípech? Co se pomocí nich detekuje?

B07 Jaký je rozdíl mezi globálním, lokálním a semiglobálním zarovnáním sekvencí? Které by jste použili pro porovnání motivů pro vazbu nukleotidů v proteinech ze dvou různých rodin?

B08 Obvykle jeden gen kóduje jeden protein. Popište mechanismus, pomocí kterého jeden gen může kódovat několik různých proteinů. Jak se takové proteiny od sebe liší?

B09 Databáze CATH (a jiné podobné databáze, např. SCOP) obsahuje hierarchii proteinů známé struktury. Jak vznikla skratka CATH? Jak jsou proteiny rozděleny na nejvyšší úrovni?

B10 Jaké informace obsahuje databáze Gene Ontology a jak jsou tyto informace organizovány?

OTEVŘENÉ OTÁZKY

C01 Vyberte JEDNU z následujících experimentálních metod molekulární biologie. Popište princip (7b) a specifikujte styčné body s bioinformatikou (3b):

- i) aplikace restrikčních endonukleáz
- ii) PCR
- iii) hybridizace
- iv) sekvenace

C02 Co víte o možnostech identifikace proteinů pomocí hmotnostní spektrometrie? Jak se postupuje (8b)? Jak souvisí používané metody se sekvencí proteinů a v čem jsou na ní nezávislé (2b)?

C03 Co je to genetický kód? Vysvětlete v tomto kontextu i pojmy kodon a translace. Jaké důležité vlastnosti z biologického ale i informačního hlediska má genetický kód?

C04 Popište všechny Vám známe metody předpovídání terciární struktury proteinů ze sekvence.

C05 Napište co víte o algoritmech vyhledávání řetězců (jmenujte a uveďte princip alespoň dvou vyjma naivního postupu porovnávání všeho) (6b). Jak se liší časovou složitostí (2b), na jaký typ úloh se hodí (2b)?

Pokud jste tímto skončil(a) s IV107 gratuluji! Budu rád, když vložíte svůj názor na kurz do hodnocení v ISu. Příští semestr budu vyučovat navazující kurz IV108 a IV110 věnovaný vybraným algoritmům a studentským projektům.

POMOCNÉ TABULKY

Tabulka kodonů genetického kódu:

AAA AAG = Lys
AAU AAC = Asn
ACA ACC ACG ACU = Thr
AGA AGG = Arg
AGC AGU = Ser
AUA AUC AUU = Ile
AUG = Met

CAA CAG = Gln
CAC CAU = His
CCA CCC CCG CCU = Pro
CGA CGC CGG CGU = Arg
CUA CUC CUG CUU = Leu

GAA GAG = Glu
GAC GAU = Asp
GCA GCC GCG GCU = Ala
GGA GGC GGG GGU = Gly
GUA GUC GUG GUU = Val

UAA UAG UGA = STOP
UAC UAU = Tyr
UCA UCC UCG UCU = Ser
UGC UGU = Cys
UGG = Trp
UUA UUG = Leu
UUC UUU = Phe

