

Biologický korespondenční seminář



Biozvěst

Ročník 0

Serie 2

Milí řešitelé,

tak zdárně proběhla historicky zcela první série Biozvěstu. Moc rád bych poděkoval Vám všem, kteří jste se zapojili a doufám, že se s chutí pustíte i do nových úloh. Jako motivací budiž Vám skutečnost, že od 16. do 21.5.2013 se uskuteční terénní výprava Biozvěstu do šumavského předhoří pro 12 nejlepších řešitelů. Díky podpoře Přírodovědecké fakulty UK budou mít všichni účastníci ubytování zdarma.

**Jak řešit**

Veškeré pokyny k řešení semináře získáte na internetové stránce Biozvěstu

<http://web.natur.cuni.cz/~vosolsob/krouzek/bios.html>,

nebo zadejte „Biozvěst“ do Google. Nejdříve je třeba přihlásit se ke Google skupině „Řešitelé Biozvěstu“

Biozvest-resitele@googlegroups.com,

<https://groups.google.com/d/forum/biozvest-resitele>,

vyplnit přihlášku a následně odesílání úloh se provádí prostřednictvím služby „Disk“ (dřívější „Dokumenty“) na Google.

Kdybyste měli jakýkoliv problém s uvedeným postupem, pošlete řešení na adresu biozvest@gmail.com.

V případě, že byste se ocitli bez internetu, můžete využít i klasickou poštu

Stanislav Vosolsobě

Katedra experimentální biologie rostlin

Přírodovědecká fakulta Univerzity Karlovy v Praze

Viničná 5

128 44 Praha 2

opravené řešení však dostanete naskenované e-mailem.

Nelekejte se, když Vám přijdou úlohy na první pohled příliš těžké, ponořte se do informačních zdrojů a uvidíte, že na vše lze někde nalézt odpověď. Dobré tipy k řešení naleznete také na stránce Biozvěstu v sekci „Návody“. Není nutné abyste vyřešili všechny úlohy, stačí odeslat libovolně velký fragment. Oceňujeme, pokud přiložíte jakékoliv připomínky (např. úloha byla příliš lehká/těžká, nesrozumitelná, nudná), úlohy se pokusíme tvořit k Vaší maximální spokojenosti.

Veškeré dotazy či připomínky směřujte na adresy

biozvest@gmail.com či vosolsob@natur.cuni.cz

Uzávěrka 2. série proběhne v neděli 6.1.2013 o půlnoci.

Úlohy můžete vypracovat přímo do zadání jednotlivých úloh, které se objeví ve Vašich sdílených dokumentech. Hotovou úlohu pojmenujte

Ročník-Série-Úloha-Jméno_Příjmení,

např. **0-1-2-Bioslav_Biomilný** v případě druhé úlohy první série aktuálního ročníku. Poslední Vámi provedená změna by měla být ze dne uzávěrky (odevzdání po uzávěrce je možné jen při zvláště závažných okolnostech). Z technických důvodů ukládejte úlohu ve formátu .doc, abychom Vám mohli přidávat do řešení komentáře.

Neboť nelze systémově blokovat kooperaci mezi řešiteli a navíc kooperace je základem úspěchu při vědecké práci, akceptujeme i skupinově řešené úlohy, v názvu a hlavičce úlohy však vyjmenujte všechny řešitele a tito řešitelé mohou odevzdat dohromady pouze jednu úlohu, všem bude přičten identický počet bodů. Případné spoluautory mimo řešitelů semináře taktéž uvádějte do hlavičky úlohy. Zdroje informací taktéž do řešení připište.

V každé sérii se můžete těšit na jednu „zelenou“, jednu „bílou“, studijní, praktickou a naučnou úlohu, ta naučná je v tomto ročníku věnována bioinformatice.

Mnoho zdaru při řešení Vám za kolektiv autorů přeje

Stanislav Vosolsobě

Úloha 1: Neukázněný biotop

Autor: Stanislav Vosolsobě

Počet bodů: 15

Bioslav dostal k narozeninám jako dárek let balonem, naneštěstí se udělalo špatné počasí a vzdušný koráb ztroskotal na neznámém místě. Bioslav však zachoval chladnou hlavu, vytáhl terénní zápisníček a zapsal si druhy, které spatřil v okolí: *Vaccinium myrtillus*, *Nardus stricta*, *Luzula sylvatica*, *Calamagrostis villosa*, *Polytrichum formosum*, *Sphagnum* sp., *Lycopodium annotinum*, *Dryopteris dilatata*, *Athyrium distentifolium*, *Listera cordata*, *Homogyne alpina*, *Trientalis europaea* a *Oxalis acetosella*. Poté se jal přeměřovat obvod mohutných stromů, které ho v biotopu zaujaly. Nakonec začal usilovně lovit jedince druhu *Apodemus agrarius*, protože dostal hlad. Bylo štěstí, že si hned po ztroskotání naplnil kapsu kvalitními žulovými kameny.

- 1) Z výčtu druhů určete v jakém biotopu Bioslav ztroskotal a kde to mohlo být (zkuste určit místo s největší možnou přesností, bylo to v České republice). Jaký druh stromu pravděpodobně měřil?
- 2) Jakými faktory je vymezen tento biotop (horní i spodní hranice)?
- 3) Jeden druh živočicha je pevně svázán s přirozeným vývojem tohoto biotopu. O kterého živočicha se jedná? Popište podrobně jeho vliv na ekosystém.
- 4) Jaký je jeho celkový vliv na biodiverzitu našeho ekosystému? Zjistěte příklady hmyzích druhů (zejména brouků), které se vyskytují pouze ve správně fungující verzi tohoto ekosystému a zjistěte, proč jsou tak vzácné.
- 5) Řada druhů z výčtu má u nás podobný, blízké příbuzný druh. Identifikujte příslušné dvojice a popište, jak byste druhy od sebe odlišili a jak se liší ekologicky a rozšířením.

Úloha 2: Nukleové kyseliny

Autor: Stanislav Vosolsobě

Počet bodů: 20

V biosféře se nachází velké množství různých druhů nukleových kyselin, které jsou základními stavebními kameny života. Buď jsou tvořeny DNA, nebo RNA. Konkrétní typy se pak rozlišují malým písmenkem na začátku (např. mRNA). Různých typů je celý zvěřinec.

- 1) Vaším prvním úkolem bude připojit následující označení

k DNA či RNA:

ds, c, m, mt, chl, t, r, si, mi, ss, hn, g, sn, sno, pre-m, pi.

- 2) Každý typ stručně charakterisujte a popište jejich funkci. Nastiňte případné souvislosti mezi různými typy.
- 3) Který typ se na Zemi začíná vyskytovat až od druhé poloviny 20. století?
- 3) Jak se liší významem “si” a “mi”? Co dělá a kdo nebo co jsou “Dicer” a “Argonaut”.
- 4) Rozeberte význam “mir165” při formování tvaru (tj. patterning) listu.

Úloha 3: Kladistika

Autor: Stanislav Vosolsobě

Počet bodů: 10

V minulé sérii jsme se zabývali sběrem znaků pro fylogenetickou analýzu a stanovili jsme si dva základní požadavky, kterými se řídí fylogenetika - musíme přísně dbát na to, abychom porovnávali pouze homologické znaky mezi organismy (tj. neporovnávat například křídlo hmyzu s křídlem ptáků) a snažit se nashromáždit co největší počet znaků a z těch teprve rekonstruovat fylogenezi, neboť pokud postavíme rekonstrukci na malém počtu znaků, můžeme omylem pracovat se znaky nehomologickými či konvergentními.

- 1) Proč v současnosti dominuje molekulární fylogenetika nad klasickou, která pracovala s morfologickými znaky (podobně jako Vy v praktické úloze minulé serie)?
- 2) Vlastní metodika výpočtu fylogenetického stromu zcela radikálně ovlivňuje spolehlivost výsledku. Popište, čím se liší strategie výpočtu u metody algoritmické (např. Neighbor-joining) od metod heuristických (např. Maximum parsimony či Maximum likelihood).
- 3) Nastiňte rozdíl mezi distančními metodami a znakovými. Základem řady metod vycházejících z molekulárních dat je evoluční model, k čemu slouží?
- 4) Naše data z minulé praktické úlohy zpracujeme metodou maximální parsimonie. Výpočet proveďte pomocí Wagnerova algoritmu (popsán je podrobně například v <http://ibot.sav.sk/usr/Karol/docs/Prednasky>, soubor kladistika9.doc). Výpočetní náročnost roste s počtem taxonů v analýze, tudíž bude stačit, pokud provedete analýzu pro menší počet taxonů (5-6 včetně outgroupu). Čím více znaků použijete, tím bude analýza zajímavější. Pokud bude výsledný strom vycházet nejednoznačně, nevadí, je to jen doklad obtížnosti fylogenetické práce způsobené konvergentními a nehomologickými znaky (nejednoznačné uzly zvýrazněte). U každé větve spočítejte kritérium parsimonie, tedy ke kolika evolučním změnám v konkrétní větvi muselo dojít (větve je linie spojující dvě větvení stromu).
- 5) Věrohodnost stromu se dá statisticky vyhodnotit pomocí metod bootstrapování či jack-knife. Nastiňte jejich princip a popište, jak byste je prováděli s Vašimi daty.
- 6) Co je to artefakt “long branch attraction” a jak vzniká?
- 7) Nepovinný úkol za bonusové body: Můžete zkusit analyzovat vaše data metodou Neighbor-joining.

Úloha 4: Miliardy přátelských bakterií

Autor: Stanislav Vosolsobě

Počet bodů: 10

Řada mléčných výrobků se honosí obsahem živých mléčných bakterií. Vaším úkolem bude experimentálně odhadnout počet živých bakterií vztažený na jednotku objemu výrobku (objem stačí určovat relativně). Srovnajte „přátelskost“ různých produktů. Jedna z použitelných metod se v praxi používá třeba při stanovování čistoty pitné vody. Podrobně popište metodiku a diskutujte výsledky s literaturou.

Úloha 5: Blast

Autor: Stanislav Vosolsobě

Počet bodů: 10

V minulé sérii jsme se zabývali získáváním bioinformatických dat z databází. Nyní pokročíme o krok dále. Od minule máme k dispozici sekvence základních typů lidských keratinů. Nyní prozkoumáme evoluci této genové rodiny. Bude nás zajímat, na které úrovni evoluce (tj. to co se dříve označovalo jako čeleď, řád, třída, kmen atd.) došlo k největšímu rozrušení této skupiny a případně jakému jinému proteinu je keratin nejpřibuznější.

Budeme muset u jiných organismů hledat sekvence podobné lidským keratinům. Nebudeme je však hledat podle názvu, neboť u ostatních organismů nemusí být geny a proteiny tak dobře pojmenovány jako u člověka, ale podle sekvenční podobnosti. Nástrojem nám bude program Blast, který nalezneme také na serveru NCBI

(<http://blast.ncbi.nlm.nih.gov>). Tento program pracuje se stejnou databází všech známých sekvencí, se kterou jste pracovali již minule. My mu zadáme známou sekvenci (“query”), program ji rozseká na krátké úseky (“slova”) a k těm hledá shody v databázi. Pokud uspěje, snaží se rozšiřovat původní slovo přidáváním dalších písmen ze zadané sekvence a kontroluje, zda-li dochází stále ke shodě s nalezenou sekvencí. Jako výsledek nám vyhodí ty sekvence, u kterých našel nejvíce a nejdelší podobnosti.

Program pracuje v různých módech - vyhledává nukleotidové sekvence podle zadané nukleotidové nebo může podobně hledat na proteinové úrovni. Pro naše účely je lepší

proteinová úroveň, neboť je tam lépe konzervována příbuznost. Dále je možné hledat např. i v nukleotidové databázi dle zadaného proteinu - program si databázi nejdříve translatuje ve všech šesti hypotetických čtecích rámcích a pak ji prohledává - toto je praktické při hledání podobných sekvencí v málo prozkoumaných organismech, neboť nukleotidové sekvence jsou přímým výstupem ze sekvenace, kdežto proteinové sekvence jsou teprve dedukovány z nukleotidových a tudíž nemusí být k dispozici.

kovém bloku. Pro přehlednost zjednodušte názvy sekvencí (je úplně jedno, co je v prvním řádku sekvence za znakem ">"). V BioEditu pak odstraňte redundantní sekvence, které se opakují. Výsledný soubor fasta přidejte do řešení úlohy.

The screenshot shows a BLAST search interface. At the top, there's a 'Distribution of 24 Blast Hits on the Query Sequence' chart with a color key for alignment scores: <40 (black), 40-60 (blue), 60-80 (green), 80-100 (yellow), 100-120 (orange), 120-140 (red), >140 (magenta), >200 (dark red). Below the chart is a table of sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
XP_016163235.1	predicted protein [Nematostella vectensis] >gi ED040287.1	161	204	99%	1e-42	31%
XP_016163024.1	predicted protein [Nematostella vectensis] >gi ED047959.1	22.6	1276	96%	6e-18	30%
XP_016163026.1	hypothetical protein kpo_459p.3 [Vanderwaltozyma polyspora]	22.8	402	99%	1e-04	37%

Below the table, there's an 'alignments' section with a 'Get selected sequences' button. The alignment details show the query sequence and its alignment with the top hit (XP_016163235.1).

Pro vlastní analýzu si tedy nejdříve na výchozí stránce Blastu vyberete možnost "blastp" a pak již můžete zadat vlastní hledanou sekvenci. Na prvním obrázku vidíte možnost omezení hledání jen na konkrétní taxony - v tomto případě zkoumáme příbuzné sekvence k sekvenci z huseníčku v rámci brukvovitých a nechceme, aby nám vypadávala ta samá sekvence z huseníčku. Vlastní hledání trvá určitou dobu a poté spatříte v případě úspěchu výsledky uspořádané jako na druhém obrázku. Na stránce s výsledky jsou nejprve graficky seřazeny nalezené sekvence a podle jejich barvy můžete usuzovat na podobnost se zadanou sekvencí (nejpodobnější jsou červené a růžové). Poté následuje tabulka nalezených sekvencí s parametry podobnosti. Nejinformativnější je "E-value", která udává pravděpodobnost, se kterou by mohla být podobnost sekvencí čistě náhodná (ideální hodnoty jsou v řádu 10⁻¹⁰, naopak hodnoty typu 0,001 nám naznačují, že se danou sekvencí nemusíme zabývat). Poslední část výstupu tvoří vzájemné přiřazení (alignment) zadané a nalezených sekvencí s vyznačením shody. Sekvence, které nás zajímají si zaškrtneme a získáme pomocí "Get selected sequences". Na další stránce pak specifikujete "Display Settings" na "Fasta (text)" a výsledek si zkopírujeme a uložíme pomocí poznámkového bloku jako soubor s příponou .fst. Výsledek můžeme obdivovat opět v BioEditu. A nyní pár úkolů:

1. Nejdříve prozkoumejte seznamy kompletně sekvenovaných organismů a vyberte druhy, které budou zajímavé, pokud chceme zkoumat historii lidských keratinů (abyste pokryli základní evoluční stupně vývoje živočichů).
2. Pro každý druh proveďte blastování postupně se všemi základními typy lidských keratinů. Vyberte vždy několik nejpodobnějších sekvencí a ty shromažďujte v poznám-

