**Laboratoriniai darbai, kuriuos reiktų atlikti prieš atliekant šį darbą:** nėra.

**1. Nežinomų sekų atpažinimas naudojantis BLAST paieška**

**Teorinis įvadas**

Pagrindinis bioinformatikos įrankis yra duomenų bazės. Jose gali būti saugomos DNR sekos (geno nukleotidų seka), RNR sekos (taip pat nukleotidų seka), baltymų pirminės struktūros (aminorūgščių seka) ir bet kokia kita biologinė informacija. Dažniausiai naudojamos genų ir baltymų bazės priklauso JAV mokslo centrui NCBI *(National Center for Biotechnology Information).* Paieška jose atliekama naudojantis programa BLAST – tai bioinformatiko „Google“. Būtent šiais įrankiais ir naudositės išspręsti problemai apie mikroorganizmus.

Išoriškai visi mikroorganizmai panašūs – daugumos rūšių negalima atpažinti net ir per mikroskopą, o žiūrint plika akimi dauguma mikroorganizmų kolonijų atrodo vienodai. Dėl to jų atpažinimas remiasi genetika. Kiekvieno organizmo genai ir baltymai skiriasi, taigi palyginus savo turimo organizmo molekules su duomenų bazėmis galima nustatyti, kokia tai rūšis.

Mokslininkai bandė užauginti *Vibrio cholerae* bakteriją, kuri žmonėms sukelia cholerą. Ja galima užsikrėsti geriant nešvarų vandenį arba valgant gyvūnus, augusius tokiame vandenyje. Iš dviejų aptiktų kolonijų buvo išskirtos tam tikro geno sekos, o jums pateiktos sekos baltymų, kuriuos koduoja tie genai. (Prisiminkite, kad baltymo seką galima lengvai sužinoti pagal geno seką ir genetinį kodą.) Naudodamiesi BLAST paieška, turėsite patikrinti, ką mokslininkams pavyko užauginti.

**Tikslas**: atpažinti dvi mikroorganizmų rūšis remiantis jų baltymų sekomis ir BLAST paieška.

**Reikalingi failai**: baltymas1.fasta, baltymas2.fasta

**Eiga**

Baltymų sekos pateiktos failuose baltymas1.fasta ir baltymas2.fasta. Nors failo formatas neįprastas (.fasta), jis nedaug skiriasi nuo paprasto teksto.

* Atsidarykite aplanką, kuriame yra šie failai. Spragtelkite ant baltymas1.fasta failo dešiniu pelės mygtuku, pasirinkite „Open with“, ir iš pasirodžiusio sąrašo išrinkite programą „Notepad“ (jei jos sąraše nėra, pirma paspauskite „Choose default program“).
* Apžiūrėkite failo struktūrą – ji būdinga visiems FASTA formato failams.

>Baltymo seka iš pirmosios kolonijos

MKKTKIVCTIGPKTESVEKLTELVNAGMNVMRLNFSHGDYVEHGTR

ITNFRKVMEVTGKQLAILLDTKGPEIRTIKLENGDDVDLVAGQEFT

FTTDTKVVGNKERVAVTYSGFTKDLNVGNRILVDDGLIEMEVLATT

DTEVKCKVLNNGALGENKGVNLPGVSVNLPALSEKDKNDLKFGCEQ

Pav. 1. FASTA formato struktūra: pirma eilutė visad prasideda simboliu „>“ – joje aprašoma, kokio tai geno ar baltymo seka. Pati seka pateikiama tolesnėse eilutėse, kur kiekviena raidė žymi nukleotidą arba aminorūgštį, pvz., pirma šio baltymo aminorūgštis yra M – metioninas.

K1. Kiek skirtingų raidžių reikia koduoti baltymo sekoms? DNR sekoms? RNR sekoms?

|  |
| --- |
| Baltymo sekoms - 20:  A - Alaninas  R - Argininas  N - Asparaginas  D - Asparto rūgštis  C - Cisteinas  E - Glutamo Rūgštis  Q - Glutaminas  G - Glicinas  H - Histidinas  I - Izoleucinas  L - Leucinas  K - Lizinas  M - Metioninas  F - Fenilalaninas  P - Prolinas  S - Serinas  T - Treoninas  W - Triptofanas  Y - Tirozinas  V - Valinas  DNR sekoms - 4:  A - Adeninas  T - Timinas  G - Guaninas  C - Citozinas  RNR sekoms - 4:  A - Adeninas  U - Uracilas  G - Guaninas  C - Citozinas |

Sekanti užduotis – nustatyti, kam priklauso šis baltymas. Pirmiausia patikrinkime, ar tai *Vibrio cholerae*.

* Naršyklėje atsidarykite BLAST programos tinklapį (<http://blast.ncbi.nlm.nih.gov/>). Po antrašte „Basic BLAST“ rasite kelis BLAST programos variantus. Šiuo atveju atidarykite „protein blast“, kadangi ieškosime baltymų sekų. Turėtumėte matyti tokį vaizdą:



Pav. 2. BLAST programos darbo laukas.

* Visą baltymas1.fasta failo turinį nukopijuokite į laukelį A („Enter <...> FASTA sequences“) arba pasirinkite failą spustelėję ant „Choose File“ mygtuko.
* Kad apribotume paiešką tik baltymuose iš mus dominančios choleros bakterijos, laukelyje B („Organism“) įrašykite „Vibrio cholerae“.
* Galiausiai laukelyje C uždėkite varnelę, kad rezultatai būtų atverti naujame lange, ir spauskite mygtuką „BLAST“.
* Paieška gali užtrukti keletą minučių – būkite kantrūs. Jei praėjo daugiau kaip 5 minutės, galite paspausti mygtuką „BLAST“ iš naujo. Kai pasirodys rezultatai, lauke „Graphic Summary“ matysite daug raudonų juostų – jos rodo, kad rastos sekos labai panašios į jūsų užklausą. Žemiau, lauke „Descriptions“, rasite ir sekų pavadinimus. Jų formatas bus maždaug toks:

Succinate dehydrogenase [Vibrio cholerae HE-16]

Pav. 3. BLAST sekos aprašo formatas. Pirma pateiktas baltymo pavadinimas, o laužtiniuose skliaustuose nurodyta rūšis, kuriai priklauso ši seka. Sekos gali kartotis, nes ištirta daug šios bakterijos atmainų – kartais jos nurodomos greta rūšies pavadinimo, pvz. aukščiau pateikta HE-16 kamieno seka.

* Pažvelkite dešiniau sekos pavadinimo – stulpelyje „Ident“ rasite skaičių procentais, kiek aminorūgščių šioje sekoje ir jūsų užklausoje sutampa. Didelis sekų sutapimas rodo, kad baltymai evoliuciškai artimi: pavyzdžiui, žmogaus ir beždžionės baltymų sekos panašios, nes šios rūšys atsiskyrė neseniai, o žmogaus ir mielių – labai skirtingos.

K2. Pasirinkite vieną rastą seką su geriausiu sutapimo procentu (ji bus rezultatų lentelės viršuje) ir įrašykite jos duomenis žemiau:

|  |  |  |
| --- | --- | --- |
| **Baltymo pavadinimas** | **Rūšis** | **Sutapimo procentas** |
| Pyruvate Kinase | Vibrio cholerae | 100% |

* Visą procedūrą (žingsniai 4-8) pakartokite su seka iš baltymas2.fasta failo.

K3. Iš šios paieškos taip pat parinkite seką su geriausiu sutapimo procentu ir įrašykite jos duomenis žemiau:

|  |  |  |
| --- | --- | --- |
| **Baltymo pavadinimas** | **Rūšis** | **Sutapimo procentas** |
| Pyruvate Kinase | Vibrio cholerae | 96% |

Kadangi šiuo atveju gauti rezultatai prastesni, galime įtarti, kad baltymas priklauso kitam organizmui, tad dabar paieškoje neribosime organizmų pasirinkimo.

* Pakartokite paiešką (žingsniai 4-8) su antrąja seka, tik laukelį „Organism“ palikite tuščią.

K4. Ir iš šios paieškos parinkite seką su geriausiu sutapimo procentu ir įrašykite jos duomenis žemiau:

|  |  |  |
| --- | --- | --- |
| **Baltymo pavadinimas** | **Rūšis** | **Sutapimo procentas** |
| Pyruvate Kinase | Saccharomyces cerevisiae | 100% |

K5. Kaip lietuviškai vadinama rūšis, kuriai priklausė jūsų pasirinkta seka?

|  |
| --- |
| Alaus Mieliagrybių |

K6. Žemiau pateikti keli šiai rūšiai giminingų organizmų pavadinimai, o taip pat nurodyta, ar jie šiai rūšiai evoliuciškai artimi.

* Įrašykite, kiek šis baltymas sutampa su atitinkamu *Vibrio cholerae* baltymu (šį skaičių nurodėte atsakydami į klausimą K3).
* Paskutinės paieškos rezultatuose suraskite po vieną seką iš likusių organizmų ir nurodykite jų sutapimo procentą.
* Padarykite išvadą, ar *Vibrio cholerae* evoliuciškai artima rūšiai, kuriai priklauso šis baltymas (t.y. rūšiai, kurią nurodėte atsakydami į K5 klausimą).

|  |  |  |
| --- | --- | --- |
| **Rūšis** | **Ar evoliuciškai artima rūšiai iš praėjusio klausimo?** | **Sutapimo procentas** |
| *Candida glabrata* | Labai artima | 92% |
| *Kluyveromyces marxianus* | Artima | 85% |
| *Candida tropicalis* | Tolima | 71% |
| *Yarrowia lipolytica* | Tolima | 0% |
| *Aspergillus clavatus* | Labai tolima | 0% |
| *Vibrio cholerae* | Labai artima | 96% |

K7. Padarykite išvadą apie sekų panašumą gimininguose organizmuose.

|  |
| --- |
| Sekų fragmentai sutampa evoliuciškai gimininguose organizmuose. Kuo organizmai artimesni evoliuciškai, tuo didesnis jų sekų sutapimo procentas. |

K8. Pabaikite šio tyrimo išvadas, įrašydami pavadinimus arba parinkdami tinkamą variantą:

Palyginus baltymo aminorūgščių sekas iš dviejų kolonijų, nustatyta, kad pirmąją koloniją sudaro mikroorganizmai *neturintys branduolių(prokariotai, bakterijos)*, o antrąją *turintys branduolius(eukariotai, mielės)*. Vadinasi, mokslininkus dominanti bakterija buvo *pirmoje kolonijoje/antroje kolonijoje/abiejose kolonijose.*