

A tárgy címe: Bioinformatika

Kötelezően választható tárgy IV. és V. évfolyamos biológus hallgatók számára;
heti 2+3 óra

Előkövetelmény:

Biokémia főkéllégium; genetika főkéllégium; alapszintű számítógépes ismeretek:
szövegszerkesztés, táblázatkezelés, e-mail, web-böngésző használata

Részvevő oktatók:

Dr. Patthy László, egyetemi magántanár
Dr. Ponyi Tamás, egy. adjunktus, Genetikai Tanszék
Dr. Tóth Gábor, tud. munkatárs, MBK Gödöllő
Dr. Barta Endre, tud. munkatárs, MBK Gödöllő

Tantárgyi tematika

Elmélet

1. **Bevezetés**

A bioinformatika története. A bioinformatika tárgya, alkalmazási területei ma. Genom projektek. Trendek. A molekuláris biológiában leggyakrabban használt bioinformatikai módszerek, programok, programcsomagok rövid áttekintése. Internet alapok: e-mail, telnet, ssh, ftp, WWW. Bioinformatika a világhálón. EMBnet, EBI, NCBI.

2. **Bioinformatikai adatbázisok**

Molekuláris biológiai adatbázisok. Elsődleges szekvencia-adatbázisok. Nukleotidszekvencia adatbázisok: EMBL, GenBank, DDBJ. Fehérjeszekvencia adatbázisok: PIR, SWISS-PROT, TrEMBL. Összetett („nem redundáns”) fehérjeszekvencia adatbázisok. Szekvencia-adatbázis formátumok. Másodlagos vagy származtatott adatbázisok. Komplex adatbázisok: térképezési–genomikai adatbázisok (genomprojektek), taxonómiai–filogenetikai adatbázisok (NCBI/Taxonomy, COG), funkcionális megközelítések („Gene Ontology”), humángenetikai adatbázis (OMIM). Bibliográfiai adatbázis: PubMed.

3. **Információkeresés az adatbázisokban**

Keresés az adatbázisok annotációiban: SRS. Integrált információkeresés: NCBI–Entrez. Integrált keresés a genom-projektek eredményeiben. Genom-böngészők: Ensembl, UCSC.

4. **A biológiai szekvenciaelemzés alapjai**

Szekvenálási adatok kezelése, ellenőrzése. Kontigok összeállítása. Restriktions endonukleáz hasítóhelyek keresése. Primertervezés. Új szekvencia elhelyezése az elsődleges adatbázisban.

5. **Szekvencia-összehasonlítások**

Páronkénti összehasonlítások: „dot-plot”. Pontozási módszerek, helyettesítési mátrixok. PAM, BLOSUM mátrixok. Páronkénti szekvenciaillesztések: optimális illesztések. Globális és lokális illesztés; dinamikus programozási algoritmusok:

Needleman–Wunsch és Smith–Waterman algoritmusok. Hézagok pontozása.

6. **Hasonlósági keresések szekvencia-adatbázisokban**

Keresés optimális illesztési algoritmusokkal: webes implementációk. Heurisztikus keresési módszerek: FASTA és BLAST algoritmusok. Statisztikai alapok: a találatok szignifikanciájának becslése. A FASTA3 programcsomag programjai. A BLAST programok. Mikor melyik programot használjuk? A jel/zaj arány javítása. A fals pozitív találatok okai és kiszűrésük: kis összetételi komplexitású régiók, repetitív szekvenciák, vektorszennyezés.

7. **Többszörös szekvenciaillesztés**

Módszerek. A progresszív illesztés módszere a ClustalW program példáján. Szegmens-alapú illesztés: dialign2. Motívum-alapú illesztés: MEME. Többszörös illesztések vizualizálása.

8. **Molekuláris filogenetikai elemzések I.**

A filogenetikai elemzés áttekintése: filogenetikai szignál, filogenetikai fák. Evolúciós változások, távolságok becslése: szubsztitúciós modellek (aminosav és nukleotid). A filogenetikai rekonstrukció módszerei. Távolság-alapú módszerek: UPGMA, legkisebb négyzetek, „minimum evolution”, „neighbor joining”.

9. **Molekuláris filogenetikai elemzések II.**

Karakter-alapú módszerek. Maximális parszimónia (MP) módszerek. Konszenzus fa. "Maximum likelihood" (ML) módszerek. MP- és ML-fák heurisztikus keresési módszerei: "branch-and-bound", NNI, SPR, TBR, SD. Statisztikai tesztek. Fák összehasonlítása, topológiai távolságok. A PHYLIP és PAUP programcsomagok.

10. **Nukleinsav-szekvenciák alapján végzett predikciók**

Funkcionális helyek, régiók detektálása a DNS-ben. Kódoló régiók, exon–intron határok predikciója. RNS másodlagos szerkezet predikciója.

11. **Távoli fehérjeszekvencia-hasonlóságok kimutatása**

Fehérjecsald, -domén, funkcionális hely adatbázisok és az azokban való homológia-keresés. Funkcionális predikció. Reguláris kifejezések, mintázat-adatbázisok: PROSITE „patterns”. Motívum-adatbázisok: PRINTS, BLOCKS. Pozícióspecifikus pontozómátrixok (PSSM), profil módszerek. Iterációs keresések: PSI-BLAST, PHI-BLAST. Profil és profil-HMM (rejtett Markov modell) adatbázisok: PROSITE profile, Pfam, SMART. Klaszterező adatbázisok: ProDom. Integrált adatbázisok és keresőrendszerek: InterPro, DART.

12. **Fehérjeszerkezet, fehérjeszerkezet előrejelzés I.**

A fehérjeszerkezet különböző szintjei. Fehérjegeometria. Fehérje családok, szupercsaládok. Feltekeredés (folding): szekvenciában kódolt információk, rejtett információk, fold családok. Strukturális osztályozás. Fehérjeszerkezet adatbázisok: PDB, MMDB, SCOP, CATH. Fehérje szerkezet összehasonlító módszerek: PRIDE, genetikus algoritmusok. Strukturális hasonlóság keresések, illesztés szerkezet alapján. A térszerkezeti adatok megjelenítése. Reprezentációs formák. A leggyakrabban használt megjelenítő szoftvercsomagok.

13. **Fehérjeszerkezet előrejelzés II. Modellezés**

Modellezés a gyakorlatban. Statisztikai módszerek: Chou & Fasman előrejelzés; másodlagos szerkezet előrejelzés, ideghálózaton alapuló rendszerek; motívum és domén felismerés; kis komplexitású régiók felismerése; transzmembrán régiók keresése. Homológia modellezés. Molekula mechanikai/dinamikai módszerek. *Ab initio* módszerek. A modellek megbízhatósága. Alkalmazási terület.

14. **Génexpressziós elemzések. Proteomika**

EST projektek, EST klaszterezés. DNS-chip adatok klaszteranalízise. Fehérjeazonosítás: 2D gélelektroforézis, tömegspektrometriai eredmények értékelése. Fehérjeinterakciós térképek.

Gyakorlat

- 1. A UNIX/Linux operációs rendszer használata**
Távoli elérés: ssh, telnet. Állománykezelés; állományok mozgatása számítógépek között: scp, sftp, ftp, e-mail. Programfuttatás; program be- és kimenetének átirányítása. Számítógéphálózati alapismeretek.
- 2. Bioinformatikai adatbázisok**
A három elsődleges adatbázis honlapjának a megismerése; mezőstruktúrák összehasonlítása. A fehérjeadatbázisok megismerése, összehasonlítása. A legjelentősebb doménkönyvtárak bemutatása. Példák komplex adatbázisokra.
- 3. Információkeresés az adatbázisokban I.**
Az NCBI-Entrez rendszer bemutatása. Keresés a PubMed bibliográfiai adatbázisban. Linkek a PubMed adatbázisból más Entrez adatbázisokba. Keresés egyéb Entrez adatbázisokban.
- 4. Információkeresés az adatbázisokban II.**
Az SRS keresőrendszer működésének bemutatása. Egyszerű keresések az SRS-ben. Bonyolultabb keresések az SRS-ben. Külső programok használata az SRS-ből. Szekvenciakinyerés az SRS segítségével.
- 5. A számítógépes szekvenciaelemzés alapjai**
Szekvenciakezelés, különböző szekvenciaformátumok. Számítógépes szekvenciaelemző programok, programcsomagok működésének bemutatása, indításuk, kezelésük. Szekvenciák kezelése az EMBOSS programcsomagban.
- 6. Szekvenáláshoz kapcsolódó számítógépes feladatok**
Számítógépes primertervezés bemutatása. Restrikciós hely meghatározó programok bemutatása. Automata szekvenátor adatainak feldolgozása. Szennyező szekvenciák eltávolítása. Szekvenciák összerakása. Új szekvenciák beküldése weben keresztül egy elsődleges adatbankba.
- 7. Genom-adatbankok használata**
Teljes genomsekvenciák a weben. Az Ensemble genom-böngésző rendszer bemutatása; Keresések teljes genomokban. Genom-régiók összehasonlítása.
- 8. Szekvenca-összehasonlítások, szekvenciaillesztések**
Dot-plot módszerek. Globális és lokális illesztések (DNS–DNS ill. fehérje–fehérje). Exon–intron határok megállapítása cDNS és genomi DNS illetve fehérje és genomi DNS illesztése révén. Programok az EMBOSS csomagban és a világhálón.
- 9. Hasonlósági keresések szekvenca-adatbázisokban**
A FASTA3 és BLAST programok használata, paraméterezése. Lokális BLAST adatbázis formázása. Pontozómátrixok, hézagbüntetések megválasztása. Kis komplexitású és ismétlődő szekvenciák maszkolása. Keresések lokális és távoli adatbázisokban. Az eredmények értékelése.
- 10. Többszörös szekvenciaillesztés. Fehérjék doménszerkezetének megállapítása**
A számítógépes programok bemutatása, alkalmazhatóságuk összehasonlítása. Nukleinsav- és fehérjeszekvenciák illesztése. Konzervált régiók keresése az illesztett

szekvenciákban.

A fehérjedomén-, mintázat- és motívumadatbázisok használatának bemutatása. Ismert domének, motívumok keresése fehérjékben. Illesztett fehérjedomén szekvenciákból új profil-HMM készítése, az adott motívumot tartalmazó fehérjék keresése a SWISS-PROT adatbázisban. PSI-BLAST keresések.

11. **Molekuláris filogenetikai elemzések**

A PHYLIP programcsomag és a PAUP program működésének bemutatása. Más filogenetikai programok megismertetése. Filogenetikai rekonstrukció elvégzése különböző (távolság-alapú, MP, ML) módszerekkel. Bootstrap elemzés.

12. **Fehérjeszerkezet**

Ismerkedés a PDB adatbázissal: szerkezetkeresés, letöltés. Térszerkezeti adatok megjelenítésére alkalmas programok: ICMLite, SwissPDBViewer, Rasmol. Az egyes régiók, aminosavak, oldalláncok kiemelése.

A gyakorlaton használt szoftverek: EMBOSS (Chou Fasman predikcióhoz), ICMLite (Linux), SwissPDBViewer (Linux), Rasmol (Windows).

13. **Fehérjeszerkezet előrejelzés**

Szerkezet előrejelzés homológia modellezéssel: SwissPDBViewer alkalmazása homológia modell létrehozására; Ismeretlen fehérjeszekvenciából modell létrehozása; A modell minőségének jellemzése (Ramachandran plot).

Szerkezet előrejelzés szekvencia homológia hiányában: Transzmembrán régió keresése DAS algoritmus alapján; Alacsony komplexitású részek keresése SEG programmal; Domén felismerés az SBASE programmal; Másodlagos szerkezet előrejelzés a PHD programmal; A fenti eredmények összevetésével a szekvencia és szerkezet illesztése; Fold felismerés (HOMSTRAP).

A gyakorlaton használt szoftverek: SwissPDBViewer (ICMLite, Ramachandran), DAS (transzmembrán régió előrejelzés), SEG (alacsony komplexitású régió előrejelzés), SBASE (domén felismerés), PHD (másodlagos szerkezet előrejelzés), HOMSTRAP (fold felismerés).

A gyakorlatok értékelése:

Minden gyakorlatról jegyzőkönyv benyújtása elektronikus formában.