

Bioinformatika: bevezetés



Gáspári Zoltán, 2019

gaspari.zoltan@itk.ppke.hu

Mi a bioinformatika?

Sokféle meghatározás létezik, melyek általában említik a számítógépeket és speciális területeket. Személyes véleményem szerint a legjobb meghatározás az, hogy a bioinformatika **a biológiai adatok feldolgozásának és értelmezésének a tudománya.**

Ehhez ma, a XXI. sz. elején **számítástechnikai eszközöket** használunk, azonban a gépek nem oldják meg helyettünk a feladatokat, csak **segítséget** nyújtanak hozzá.

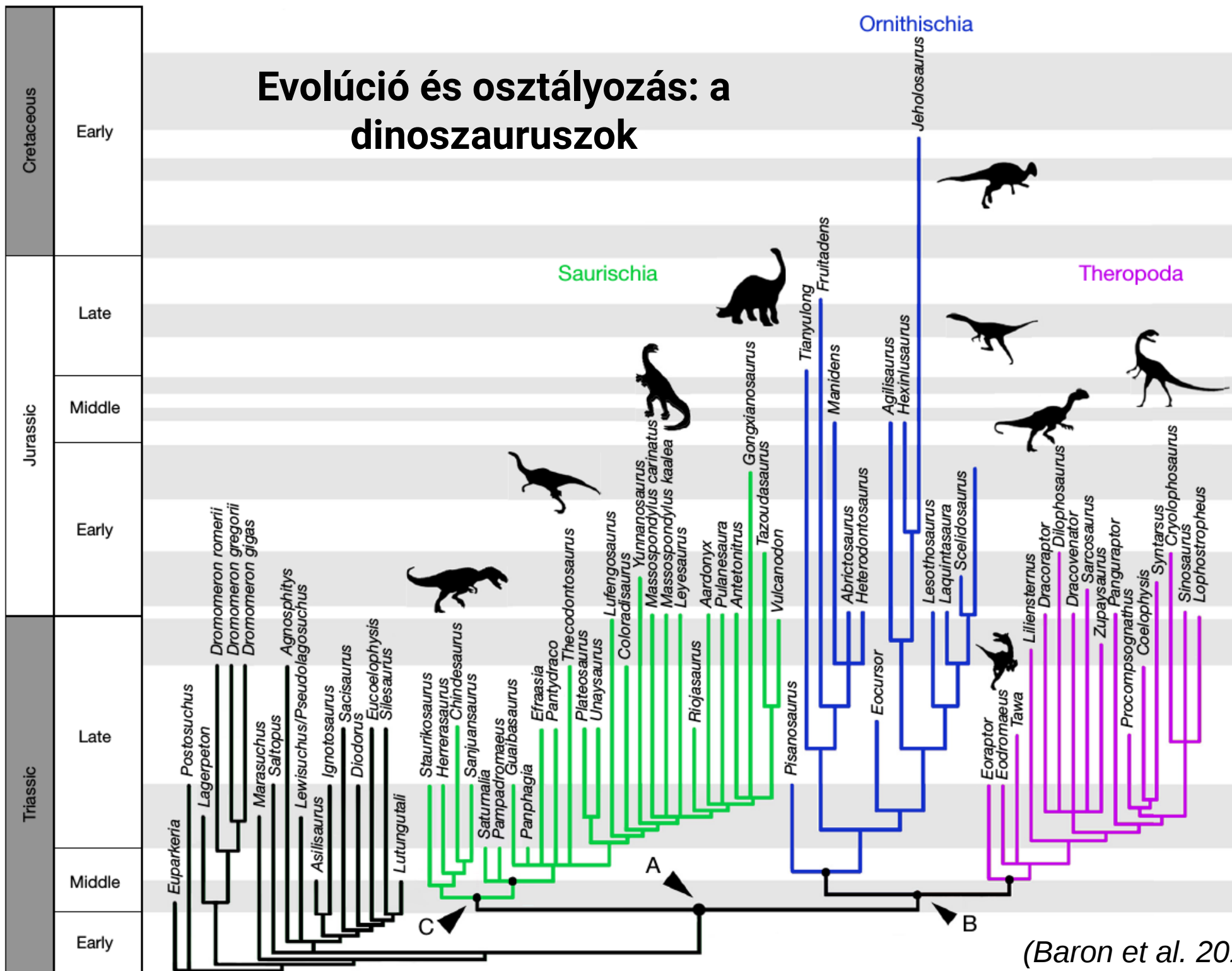
Gépekre az **adatok mennyisége** és a **számítások összetettsége** miatt van szükségünk.

A bemeneti adatok megfelelő **előkészítése** és a kimenet **biológiai** értelmének, **jelentőségének** meghatározása a **kutató feladata!**

A bioinformatikai elemzés nem ér véget az adott program futásának befejezésével, hanem valójában akkor kezdődik el a lényegi része:

- *Mit hihetek el a kapott adatokból és mit nem? (Biológiai tudás / algoritmusok korlátainak ismerete / józan ész!)*
- *Milyen újabb elemzésekkel tudom megerősíteni/megcáfolni az első vizsgálatok alapján kapott képet?*
- *Ha két módszer ellentmondó eredményeket ad, melyiknek higgyek? Esetleg egyiknek sem?*
- *Milyen, biológiailag értékes és használható új információhoz jutottam?*

Evolúció és osztályozás: a dinoszauruszok

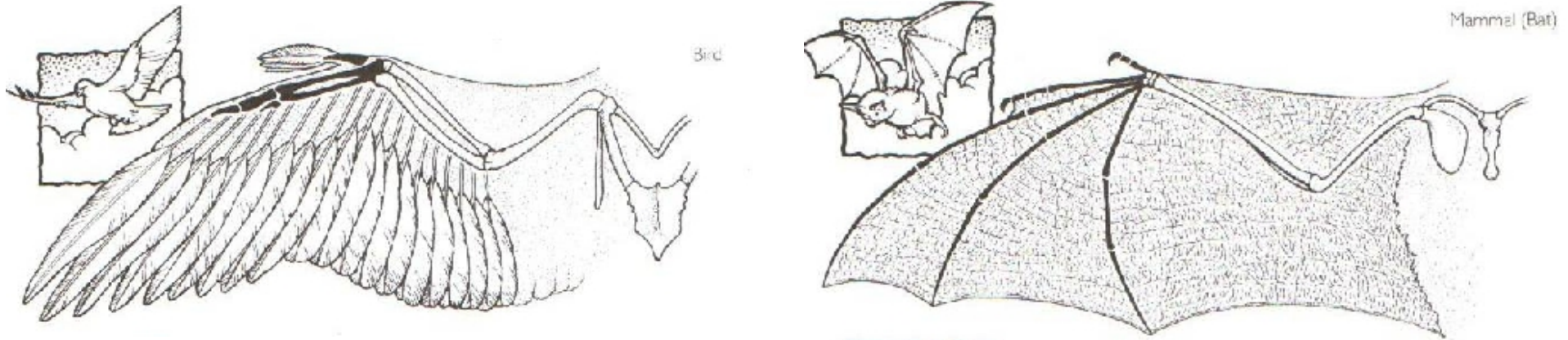


(Baron et al. 2017)

Homológia és analógia

A **homológia** evolúciós rokonságot jelent. Két szerv, csont, sejt, gén vagy fehérje akkor homológ, ha közös őstől származnak. Fontos, hogy ez önmagában nem feltétlenül jelent egyéb fajta, pl. funkcionális vagy alaki hasonlóságot, csak a vizsgált képletek történetére vonatkozik. A homológia megállapítása nem feltétlenül triviális feladat.

Az **analógia** ezzel szemben alaki vagy funkcionális hasonlóságot jelent, leszármazási történetétől függetlenül.

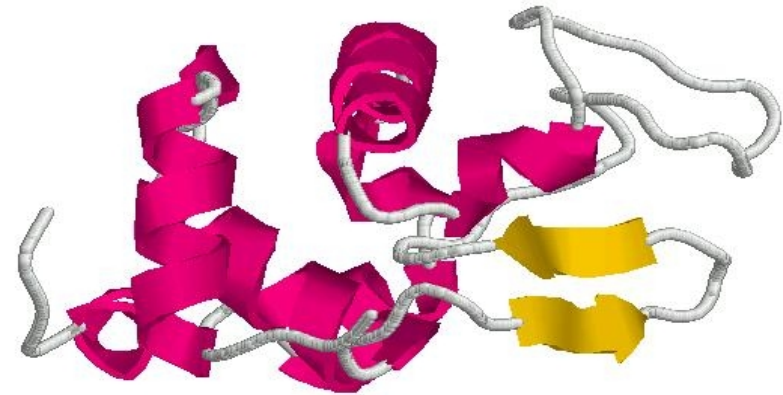
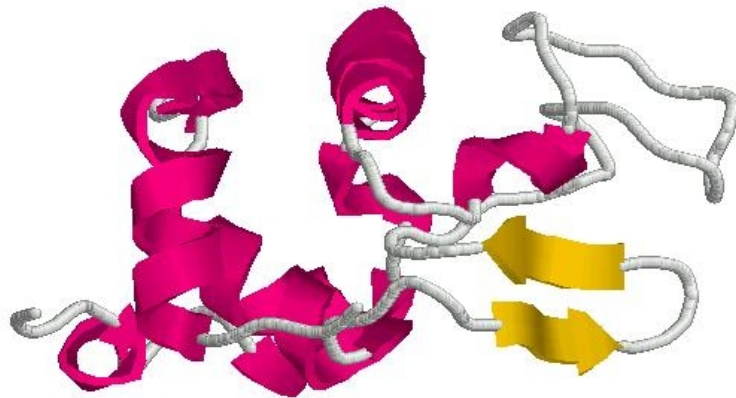


A homológia és analógia komplex viszonyban lehetnek egymással, attól függően, hogy milyen szinten vizsgáljuk az adott biológiai objektumot.

Példa: a madarak és a denevérek szárnya

- a két szárny **mint teljes végtag** egymással **homológ** (a két élőlény utolsó közös ősének mellső végtagjából származtathatóak)
- egyúttal, mint **teljes szárnyak**, **analógok** is (repülésre használatosak)
- ugyanakkor a szárnyak egyes **részei nem homológok** (pl. a szárnyak vége: toll vs. 3. ujj)
- és az egyes **homológ csontok nem analóg** helyen vannak a szárnyakon belül
- ennek oka, hogy a legutolsó közös ős mellső végtagja még nem volt szárny, a két csoportban ezek egymástól **függetlenül** fejlődtek szárnyá (konvergencia)

Homológia és analógia a gének és fehérjék világában



LALBA_HUMAN	1	MRFFVPLFLVGLFPAILAKQFTKCELSQLLK--DIDGYGGIALPELICTMFHTSGYDTQ	58
LYSC_HUMAN	1	MKALIVLGLV-LLSVTVQGVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTR	59
LALBA_HUMAN	59	AIVEN--NESTEYGLFQISNKLWCKSSQVPQSRNICDISCDKFLDDDITDDIMCAKKIL-	115
LYSC_HUMAN	60	ATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVR	119
LALBA_HUMAN	116	DIKGIDYWLAHKALCTEK	133
LYSC_HUMAN	120	DPQGIRAWVAWRNRCQNR	137

A lizozim (balra) és α -laktalbumin (jobbra) kb. 40%-os **szekvenciaazonosságot** mutatnak. Ezt a **homológia** jelének tekintjük, csakúgy, mint a nagyon **hasonló térszerkezetet**: a közös eredet a legegyszerűbb tudományos magyarázat.

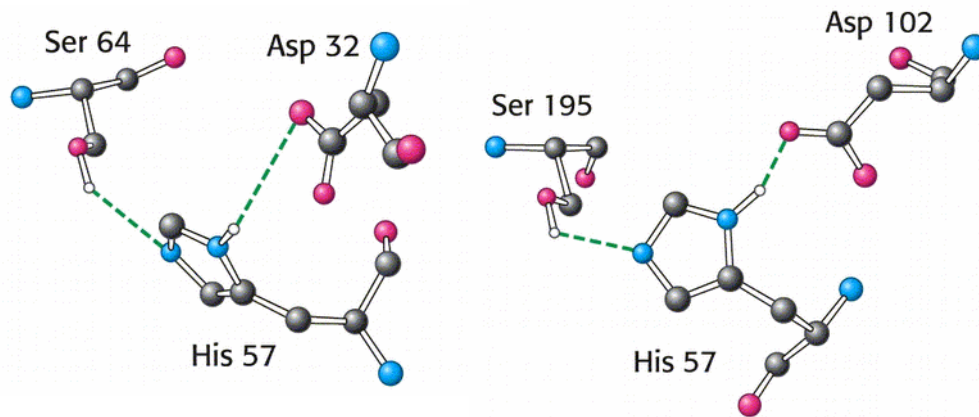
A lizozim is **enzim**: baktériumok sejtfalát bontja a védekező mechanizmusok részeként. Az α -laktalbumin a galaktoziltranszferáz enzim egyik **regulációs alegysége**, mely lehetővé teszi, hogy a tejmirigyben glükózt is felismerjen az enzim, amely ezáltal képes laktózt előállítani. Az α -laktalbumin a tejbe is átkerülő fehérje.

Homológia és analógia a gének és fehérjék világában

Különböző 3D szerkezet, de hasonló lokális elrendeződés az aktív centrumban



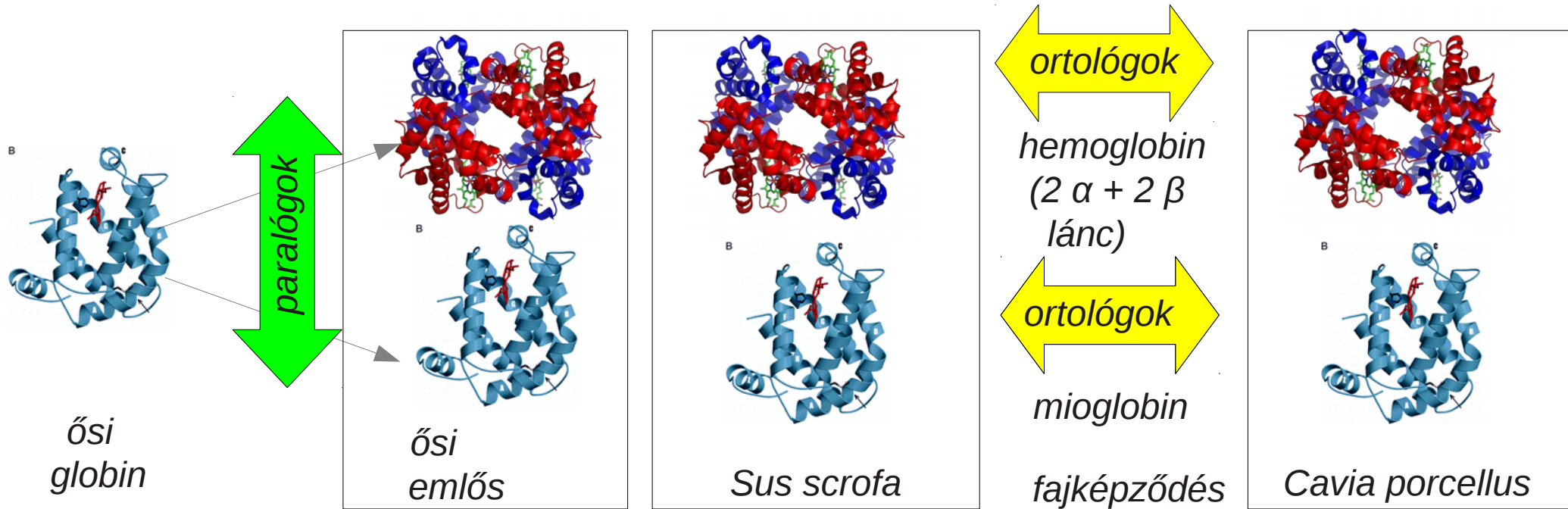
eltérő evolúciós eredet, de hasonló funkció (konvergencia)



A szubtilizin (balra) és a kimotripszin (jobbra) **szerin protázok**, melyek jellegzetes katalitikus triáddal rendelkeznek: Ser, His és Asp. A két fehérje azonban evolúciósan nem rokon, erre egyértelműen utal a különböző térszerkezetük, és hogy a triád aminosavai a szekvenciákban más sorrendben helyezkednek el.

Valójában több, mint 50(!), egymással rokonságban nem álló szerinproteáz-családot ismerünk.

A homológia alosajtosai: ortológok és paralógok



Az **ortológ** gének/fehérjék története alapvetően a fajok történetét tükrözi. Ezzel szemben a **paralógok génduplikációval** jönnek létre. A példában a disznó (*Sus scrofa*) fehérjék és azok tengerimalac (*Cavia porcellus*) megfelelői ortológok, míg a

mioglobin - α -hemoglobin,
mioglobin - β -hemoglobin és a
 α -hemoglobin - β -hemoglobin

párok paralógok, valamelyest eltérő funkcióval

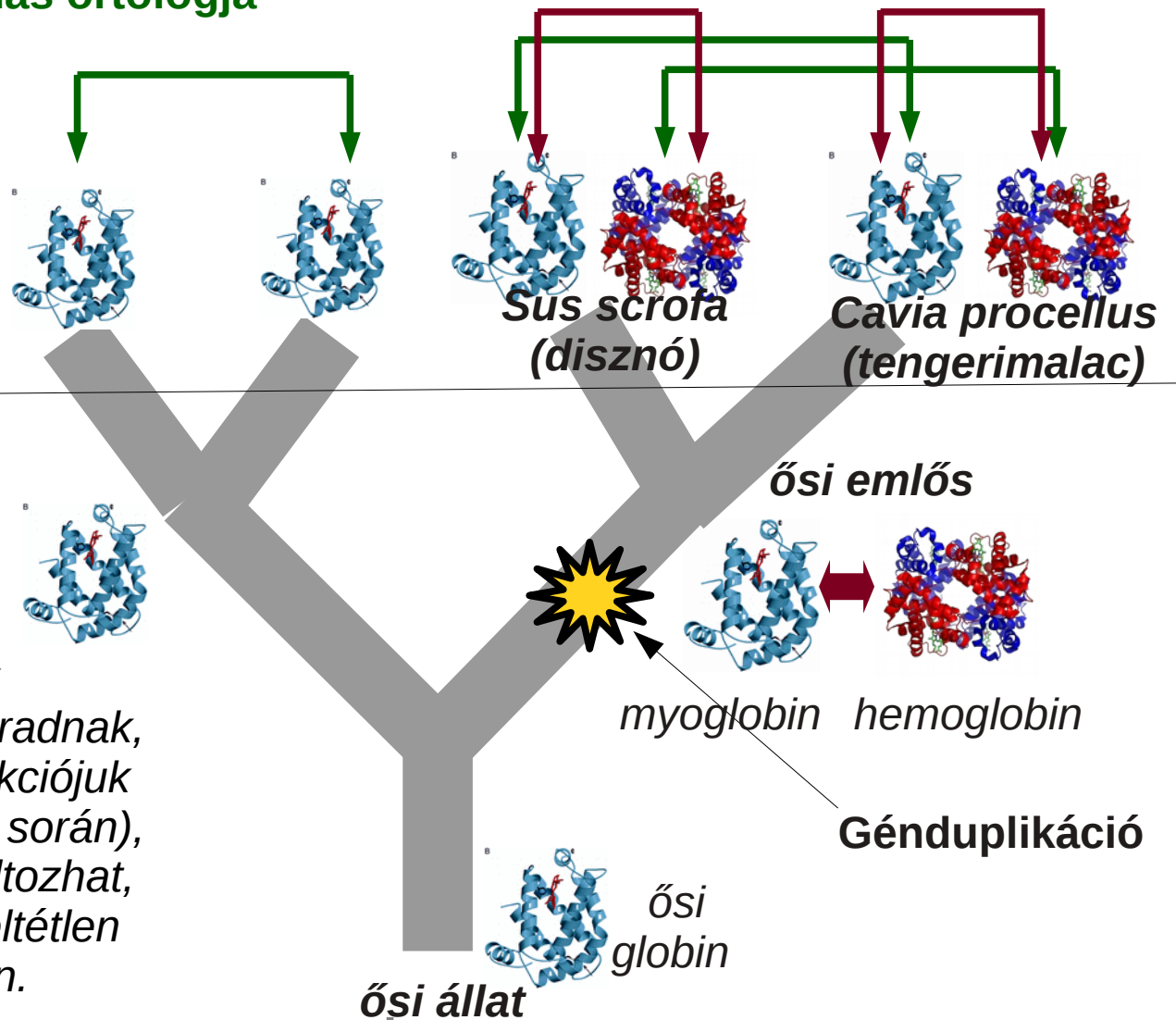


Ortológia és paralógia



- minden globin egymás között homológ
- minden mioglobin egymás ortológja
- minden α/β hemoglobin egymás ortológja
- a hemoglobinok és a mioglobinok paralógok
- a hemoglobin 2 paralóg alegységet tartalmaz (α és β)

Ma élő élőlények



Az ortológ gének a genomokban egymásnak megfelelő pozícióban maradnak, és sokszor hasonló a funkciójuk ("megmarad" az evolúció során), míg a paralógoké megváltozhat, de ezek egyikére sincs feltétlen kényszer általánosságban.

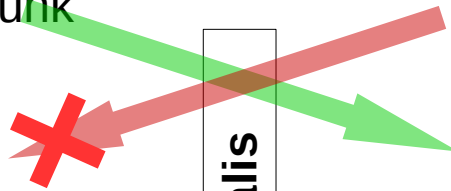
Idő/evolúció

Génduplikáció nélküli leszármazási vonal

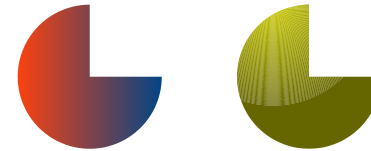
Lokális és globális hasonlóság a bioinformatikában

Az evolúciós rokonság jelének tekintjük (valószínűtlen, hogy egymástól függetlenül ennyire hasonló dolgok alakuljanak ki - globuláris fehérjékre igaz)

Hasonló térszerkezetet várunk



A szekvenciák közötti hasonlóság nem feltétlenül könnyen detektálható (divergencia)



Általában a teljes szekvenciát tekintjük



Szerkezeteknél általában a domének szintjén értelmezzük

szekvencia

hasonlóság

3D szerkezet

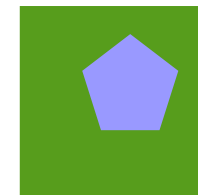
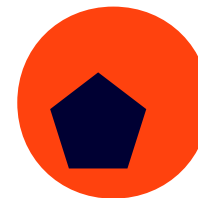
Általában doméneket/motívumokat vizsgálunk



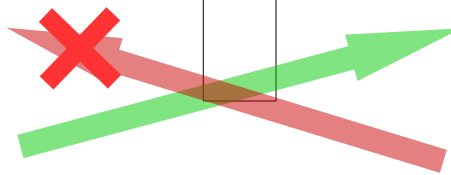
Doméneknél kisebb egységekre értjük általában



Lokális



Jelezhet hasonló lokális szerkezetet



A szekvenciában nem feltétlenül folytonos szegmens (pl. aktív centrum)

A lokális hasonlóság sokszor hasonló funkcióra utal (pl. aktív centrum, partnerkötőhely)
Nem minden esetben feltételez evolúciós rokonságot (konvergencia)

Tipikus bioinformatikai kérdések

(szubjektív lista → rokon területekre való utalásokkal)

Adott szekvenciához milyen funkció / biológiai jelentőség tartozik?

- Hány és milyen fehérjét kódol adott genom/genomi szakasz? (**génpredikció, genomannotáció**)
- Adott fehérje milyen szerkezettel/aktivitással rendelkezhet? (**szerkezet/funkció predikció**)
- Milyen fizikai kötőpartnerei lehetnek?
- Milyen más génekkel szabályozódhat együtt?

Két szekvencia / genom között mely különbségek felelősek egyes funkcionális eltérésekért?

- Milyen genetikai háttér milyen betegségekre hajlamosít? (**GWAS**)
- Mivel érdemes kezelni adott betegséget? (**személyre szabott gyógyítás**)
- Miért működik egy fehérje máshogyan, mint egy másik hasonló?

Két sejt génexpressziós/epigenetikai/splicing stb. mintázata között mely különbségeknek van biológiai jelentősége?

- Milyen funkcióval bírnak az együtt szabályozott fehérjék/DNS-szekvenciák? (**enrichment analysis**)
- Mely változásoknak van a legnagyobb jelentősége? (**adatbányászat**)
Mely változások a kiváltó okok és melyek a következmények?

Adott gén/fehérje működése hogyan befolyásolható?

- Milyen szabályozó mechanizmusok megléte valószínűsíthető adott genetikai környezetben? (**genomannotáció, → rendszerbiológia**)
- Milyen módon befolyásolható a fehérjeműködés a szerkezet ismeretében? (**→ gyógyszertervezés, biotechnológia**)

Hogyan tervezek adott szerkezettel/funkcióval bíró szekvenciát?

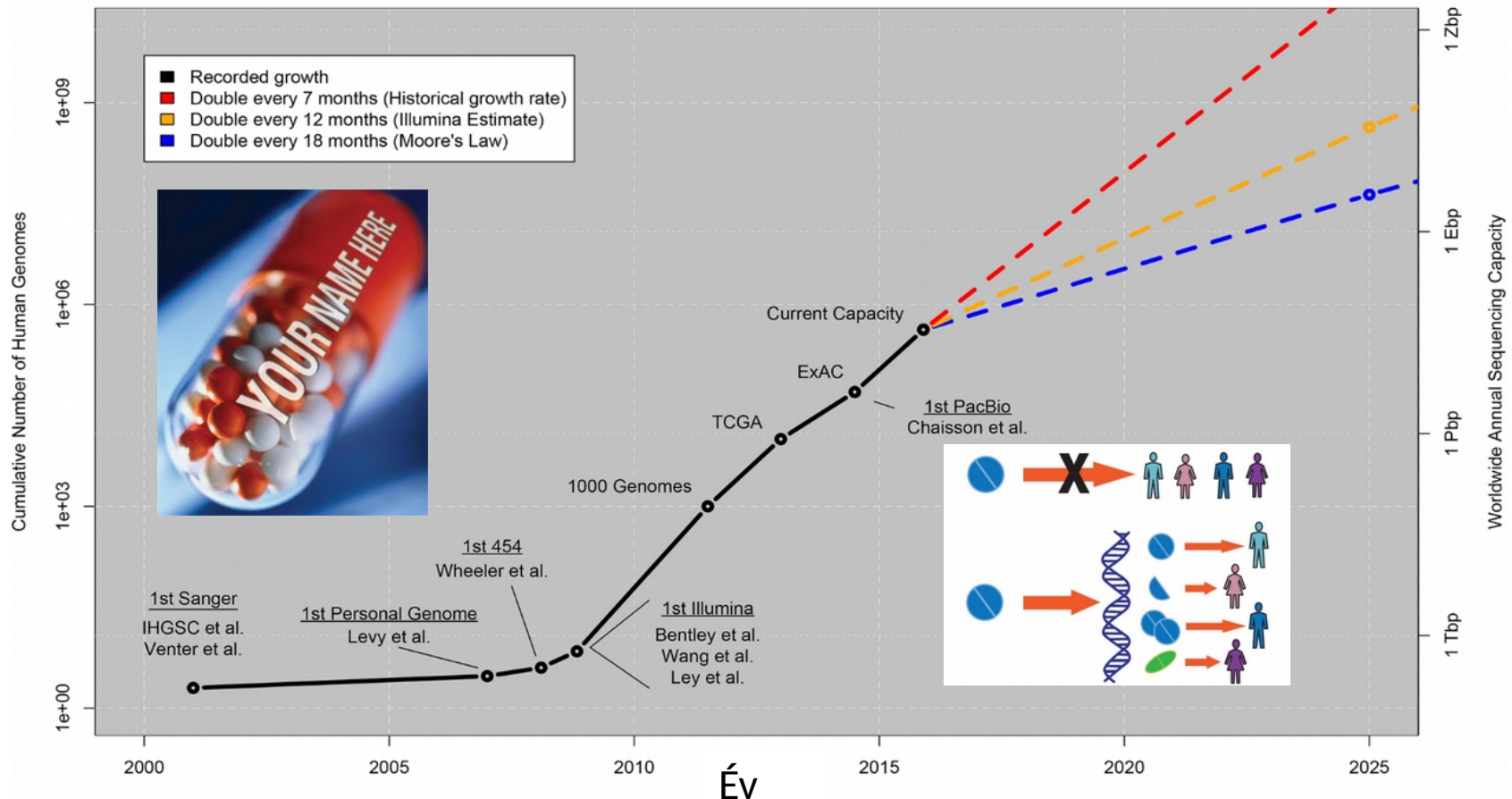
- Hogyan állítsak össze egy általam kívánt szabályozással bíró rendszert? (**→ szintetikus biológia**)
- Hogyan tervezek adott szerkezetű / funkciójú fehérjét?

Bioinformatikai kihívások a XXI. század elején

Genomszekvenálástól a személyre szabott terápiáig

Growth of DNA Sequencing

Szekvenált emberi genomok száma



Bioinformatikai kihívások a XXI. század elején

Szekvenálási technikák fejlődése

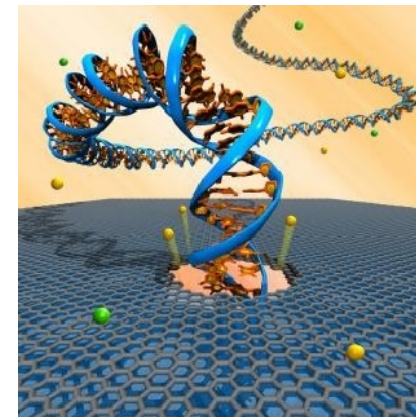


Minion



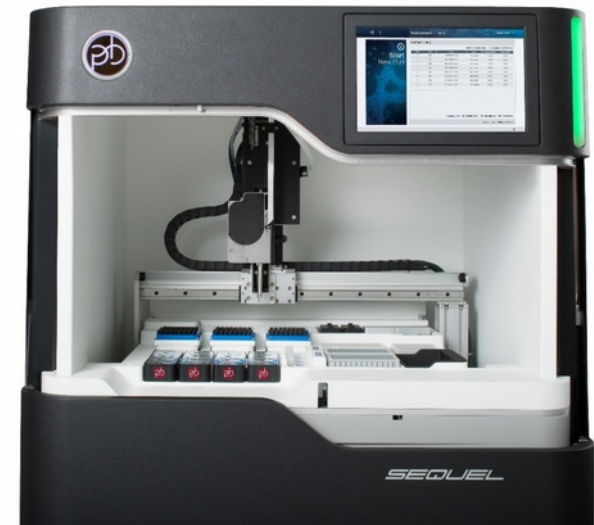
MinION Mk1: portable, real time biological analyses

MinION



DNS-szekvenálás:

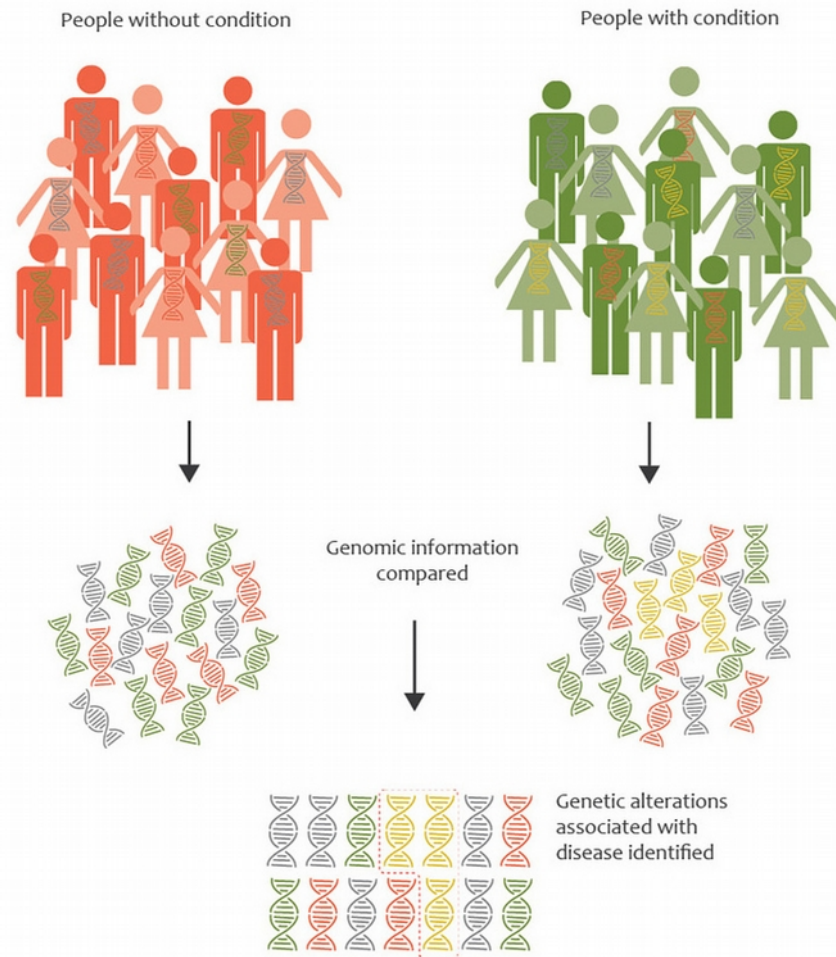
- Ár ↓, volumen ↑
- Egyedi molekulák (SMRT)
- Hordozhatóság



Bioinformatikai kihívások a XXI. század elején

How researchers compare genomic information to identify genetic alterations

Klinikailag releváns információ kinyerése: genomszintű asszociációs vizsgálatok



Bioinformatikai kihívások a XXI. század elején

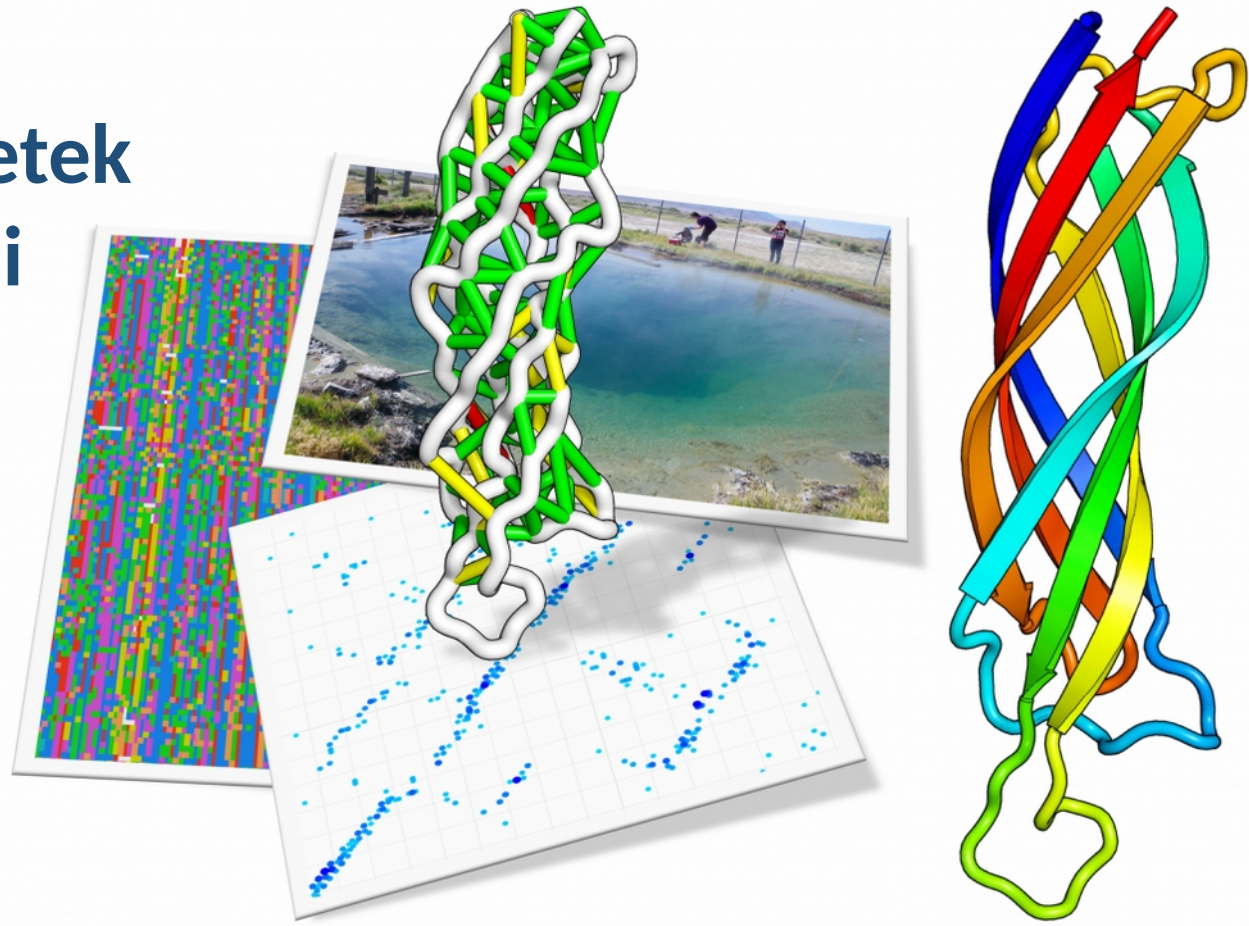
1D -> 3D információ

kinyerése:

fehérjeszerkezetek

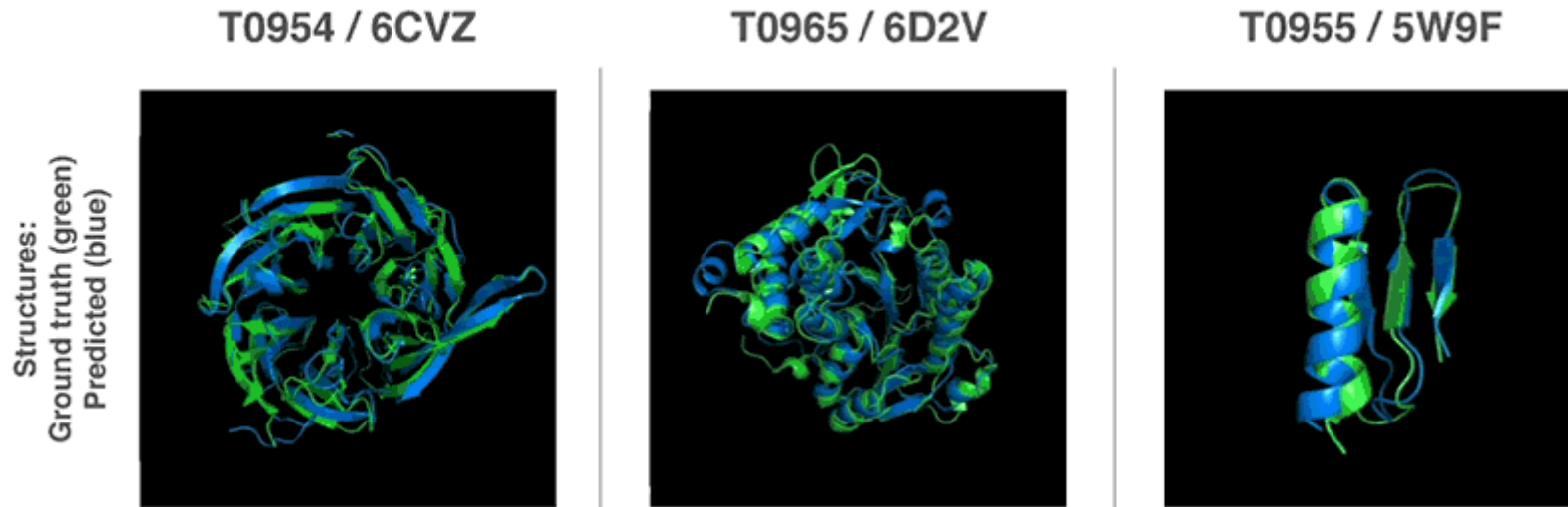
metagenomikai

adatokból



Bioinformatikai kihívások a XXI. század elején

1D -> 3D információ predikciója mesterséges intelligenciával



Bioinformatika, 2018/19. tavaszi félév

dátum	előadó	téma
febr 6	Gáspári Zoltán	Bevezetés
febr 13	Gáspári Zoltán	Proteomikai adatbázisok
febr 20	Gáspári Zoltán	Szerkezeti bioinfo
febr 27	Tantos Ágnes	rendezetlen fehérjék I
márc 6	Tantos Ágnes	rendezetlen fehérjék II
márc 13		ZH I
márc 20		tavaszi szünet
márc 27	Ligeti Balázs	NGS I: technikák
ápr 3	Ligeti Balázs	NGS II: adatfeldolgozás
ápr 10	Békési Angéla	NGS III: alkalmazások
ápr 17	Grolmusz Vince	Hálózatok I
ápr 24	Grolmusz Vince	Hálózatok II
május 1	ünnep	
május 8		ZH II
május 15		pót ZH

A tárgy teljesítéséhez mindkét Zh legalább elégséges jegyre megírása szükséges. Az egyik ZH-t lehet pótolni.

Tárgyfelelős:

Vértessy Beáta, tanszékvezető egyetemi tanár, vertessy@mail.bme.hu

Tantárgy weboldala:

<http://www.ch.bme.hu/oktatas/targyak/BMEVEMBM103>