

## NÖVÉNYI GENOMIKA

JÓRI BALÁZS

Eötvös Loránd Tudományegyetem, Növényélettani és Molekuláris Növénybiológia Tanszék,  
1117 Budapest, Pázmány P. sétány 1/c.

Elfogadva: 2004. december 29.

**Kulcsszavak:** genomika, adatbázisok, TAIR, bioinformatika, microarray, *in silico* növény

**Összefoglalás:** A növényi genomok feltárásával a növénytan kutatások is szemléletmód-váltásra kényszerülnek, a genomprojektek eredményei a növénytan bármely területén felhasználhatóak. A szekvenciák univerzális és speciális adatbázisokban halmozódnak fel, mindenki számára hozzáférhető módon. Az adatokat bioinformatikai módszerekkel elemezve a kutatásokat új alapokra lehet helyezni, az egyes gének helyett gének együttes működésének vizsgálata lesz az irányadó. További perspektívát jelent egy mesterséges, *in silico* növényi modellrendszer létrehozása.

### Bevezetés

A molekuláris biológiában alkalmazott technikák rohamos fejlődésével és a génállományok feltárásával foglalkozó nemzetközi projektek előrehaladtával nemcsak a genetika, hanem a klasszikus növénybiológia is szemléletváltásra kényszerül. Az elmúlt években egyre szélesebb teret kap a genomika, az egyes élőlények teljes örökítőállományának, azaz genomjának stuktúrájával és funkciójával foglalkozó tudományág. Az egyes élőlények genomjaira vonatkozó, adatbázisokban felhalmozott információtömeg immár nem jelenthet korlátozó tényezőt. Az adatok azonban önmagukban nem bírnak jelentéssel, azokat értelmezni is kell. Az így kapott információ viszont személyre szabható tartalmú és a növénybiológia számos területén hasznosítható: ma már a genomanalízis nem csupán a rendszertan számára jelenthet perspektívát, hanem újszerű alapját képezi a modern élettani, ökológiai vagy szervezettani kutatásoknak. Dolgozatomban olyan adatbázisokba kívánok betekintést nyújtani és rávilágítani a bennük rejlő lehetőségekre, melyek a növénybiológiában azok számára is érdeklődésre tarthatnak számot, akik eddig idegenkedtek a bioinformatikai megközelítéstől.

### A genomika története

A genomikai korszak alapköveit a klasszikus genetika rakta le. Az 1960-as évek molekuláris biológiai kutatásainak alapján az 1980-as évek végére automatizált nukleinsav (DNS és RNS) és fehérjeszekvenálási módszerek jöttek létre. Már a kezdetekkor, a 1960-as években, MARGARET DAYHOFF úttörő fehérjekutatási munkássága során is nyilvános adatbázisokba gyűjtötték a fehérjeszekvenciákat (ELLIS és ATTWOOD 2001). Később, a számítástechnika fejlődésével az összegzés helye is megváltozott: a strukturális

genomikai eredményeket, a legkülönbözőbb (fehérjét, rRNS-t, tRNS-t kódoló stb.) nukleinsav és peptidszekvenálási projektekből érkező információkat az interneten, többnyire szabadon hozzáférhető adatbázisokban kezdték gyűjteni. A látványos fordulatot a kilencvenes évek második fele hozta el: a tökéletesedő műszerek ontani kezdték az adatokat, melyeket folyamatosan fejlesztett, strukturált rendszerek, a genomi adatbázisok fogadtak.

## A genomprogramok ma

A genomikai éra szempontjából jelenleg a főbb genomszekvenálási projektek bírnak a legnagyobb jelentőséggel. Ezeknek a gyorsasága és száma főként a nagy felbontású szekvenátorok új generációjától függ, melyek újfajta tömegspektroszkópiai módszereket használnak. Növényi vonatkozásban eddig két élőlény teljes genetikai állományát sikerült megszekvenálni: a RÉDEI GYÖRGY által életre hívott *Arabidopsis Information Service* munkájának köszönhető (MEYEROWITZ 2001), 2000-ben publikált, növények között az egyik legkisebb genommal rendelkező lúdfüvet (*Arabidopsis thaliana*) az egyzikkű rizs (*Oryza sativa*) genomja követte. A modellértékű lúdfű és rizs után napjainkban többek között egy pillangós növény, a *Medicago trunculata* teljes genomját szekvenálják meg. Ez a kutatás több olyan gén szerepét és meglétét is tisztázhatja, mely a növényi – mikrobiális kölcsönhatásokban, illetve a szimbiotikus nitrogénfixációban is szerepet játszik. A lucernához hasonló pillangós mellett, a legkisebb genomú fás szárú, a 450 millió bázispár nagyságú eukaliptusszal történt kezdeti próbálkozások után hatalmas erővel folynak a fás szárú rezgő nyár (*Populus tremula*) genomjának kutatására irányuló törekvések. Az ebből származó eredmények rávilágíthatnak a fás- és a lágyszárú növények szerveződési és élettani különbségeinek molekuláris hátterére, miközben maga a növény jelentős gazdasági haszonnal is bír. Szintén gazdasági megfontolásokból folyik a kukorica (*Zea mays*) génállományának feltárása, ezen kívül nagyobb mennyiségű biológiai adat származik a paradicsomból, a szójából és a gyapotból is.

## Genom szerveződés és összehasonlító genomika

Az egymástól különböző növények genomja más-más módon szerveződik, eltérnek méretükben, ploidi szintjükben és a kromoszómák számában is. Elég, ha csak a 120 millió bázispárt tartalmazó *Arabidopsis*, illetve egyes liliumok 50 milliárd bázispárnyi genomjára gondolunk (TERRY et al. 1999): különböző anyagcsereutakkal és egyedfejlődéssel rendelkeznek, de ez nem jelenti még azt, hogy minden növényfajnak szükséges lenne a jövőben a teljes genomját megfejteni. Alkalmassint az egyedi expresszázó génszakaszok, az ún. expressed sequence tag-ek (EST-ek) megszekvenálása is elegendő elsődleges megoldásként a különbözőségek meghatározására (YUAN et al. 2001). A szekvenáló projektek rávilágítanak a különböző fajok kromoszómális elrendezésének szerkezetbeli különbségeire: a genomszerveződési vizsgálatok során eltérő szempontok szerint kutatják többek között a centromerek és a telomerek elhelyezkedését a kromoszómákon (PAUL és FERL 1998). Az ilyen típusú vizsgálatok eredményei is megmutatják, mennyire különbözően szerveződnek genomi szinten például a kétszikkű *Arabidopsis* és a fűfélék. Míg az *Arabidopsis* genomjában a gének elszórtan helyezkednek el, addig a kukoricán és a rizsen végzett kutatások szerint az egyszikkűek génei na-

gyobb csoportokat képeznek, amelyeket nagyméretű, nem kódoló régió tagol. Az összehasonlító genomika lehetővé teszi, hogy növénycsaládokat a genomjaik alapján kapcsoljunk össze, és segítenek megérteni, hogyan fejlődnek és szerveződnek a bonyolult szervezetek génállományai. Jópár növényi genom megszekvenálása után pontosabb betekintést nyerhetünk majd a fajok evolúciójába az egyes gének és a genomok szerveződésének összehasonlítása által.

### **Térképezés és szekvenálás**

A növényi genomok analízisének alapját legtöbbször egy klasszikus módszerekkel elkészített genetikai térkép képezi. Az ilyen genetikai térképek alapján indulhat meg az egyes növényfajok genomjának teljes szekvenálása. A fizikai térképek információi alapján egyedi molekuláris markerek szerkeszthetőek, így a pontosan ismert pozíciók alapján tagolni lehet a genomot enzimatis hasításokkal. Ezeket a kromoszómarészleteket mesterséges élesztőgomba klónokkal (Yeast Artificial Clones – YAC) és mesterséges bakteriális klónokkal (Bacterial Artificial Clones – BAC) illesztik össze: az ilyen eljárások során százezres nagyságú nukleotidszakaszt tartalmazó, külső, növényi eredetű örökítőanyagot juttatnak bakteriális vagy élesztőgomba-telepek genomjának ismert pozíciójába. Ezek a telepek a saját osztódásaik során az exogén eredetű génállomány-részletet is felszaporítják. A különálló klónok nukleotid-állományát feldarabolják nagyjából pár száz nukleotid hosszú kisebb szakaszokra és azokat egyenként szekvenálják. Mivel a megszekvenált kisebb darabok terminális régiói átfednek egymással, az egyes klónokon belüli százezres szekvencia pontos sorrendjének meghatározása számítógépek alkalmazásával megoldható, a klónok sorrendiségét pedig az eredeti térkép alapján szerkesztett könyvtárszerkezet határozza meg. Ezt a hierarchikus, térkép alapú módszert azonban csak teljes genomok meghatározása esetén használják a genomprojektek. Az eddigi genetikai térképek – növények közül a szójabab, paradicsom, zab, árpa, búza és kukorica ismert – többek között az NCBI honlapján is hozzáférhetőek: <http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>.

### **EST – szekvenálás**

Az érdeklődés középpontjában elsősorban az átíródo régiók állnak, a kisebb szekvenáló projektek is főként az EST szekvenciák meghatározásával kezdődnek. Az EST-k gyakorlatilag cDNS részleteknek feleltethetőek meg és 2005. január elején több, mint hatszáz fajból, közel huszonöt millió EST szekvencia áll rendelkezésünkre (1. táblázat). A GenBank honlapján (<http://www.ncbi.nlm.nih.gov/dbEST/>) megtekinthetjük a rendelkezésre álló fajok nevét és a fajokból nyert EST-k számát is. Ezen kívül több profitorientált cég is rendelkezik saját, a GenBanktól független adatbázissal, melyekben több haszonnövényből származtatott EST található meg.

### **A bioinformatika szerepe**

A biológiai tudományok napjainkban jutottak el abba a fázisba, hogy az adatbázisokból származó információt ténylegesen hasznosíthassák. Először nyílik alkalom arra, hogy növényi szervezetek összes génjét és azoknak működését, termelt fehérjét

1. táblázat  
Table 1

EST-k száma a GenBank adatbázisban a 2004. februári és 2004. decemberi állapot szerint  
([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html))  
Number of EST sequences in GenBank. State in February 2004 and December 2004  
([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html))

Faj latin neve	EST-k száma 2004. februárban	EST-k száma 2004. decemberben
<i>Avena sativa</i>	509	7624
<i>Beta vulgaris subsp. vulgaris</i>	21523	22233
<i>Glycine max</i>	345723	342413
<i>Gossypium arboreum</i>	39007	39007
<i>Gossypium hirsutum</i>	13643	23899
<i>Hordeum vulgare</i>	377074	367800
<i>Lotus japonicus</i>	36311	111459
<i>Lycopersicon esculentum</i>	150519	153911
<i>Lycopersicon hirsutum</i>	2504	2504
<i>Lycopersicon pennellii</i>	8346	8346
<i>Marchantia polymorpha</i>	1415	1415
<i>Medicago sativa</i>	879	6495
<i>Medicago truncatula</i>	187763	191131
<i>Mesembryanthemum crystallinum</i>	25803	25803
<i>Pinus taeda</i>	110622	173680
<i>Populus tremula x Populus tremuloides</i>	72060	65981
<i>Secale cereale</i>	9194	9194
<i>Solanum tuberosum</i>	132837	192038
<i>Sorghum bicolor</i>	161766	190949
<i>Sorghum propinquum</i>	21387	21387
<i>Triticum aestivum</i>	549926	587650
<i>Zea mays</i>	391145	417056

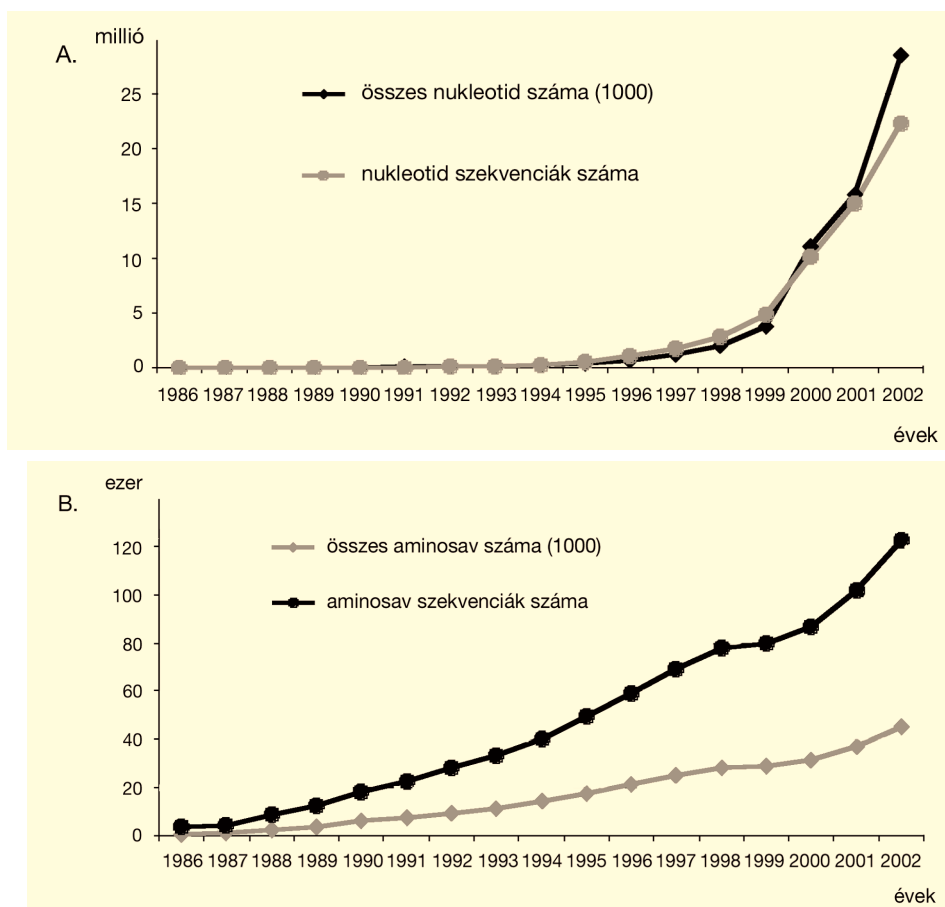
együtt és összefüggésében legyünk képesek vizsgálni. Azonban több évnyi kutatás után is csupán keveset tudunk a magasabb rendű növények genomjának szerveződéséről és evolúciójáról, illetve az egyes gének funkciójáról. A teljes genetikai információval foglalkozó genomika és fehérjeoldali megfelelője, a proteomika, vagy az anyagcsereutakat összességében vizsgáló metabolomika mint új, a többi tudományhoz szervesen kapcsolódó, de ugyanakkor önálló tudományágak próbálnak ezekre a kérdésekre válaszolni. (TERRYN et al. 1999)

A genomprojektek által produkált hihetetlen mennyiségű információtömeget értelmezni is kell. A bioinformatika egyik célja, hogy új, számítógépes módszereket dolgozzon ki a genomokból származó biológiai információ kezelésére és elemzésére. További feladata, hogy ezt képes legyen alkalmazni a biológiai tudományok többi területén is, legyen szó akár növényélettani, -rendszeri vagy akár -szervezettani vonatkozásokról. Az új, molekuláris biológiai kutatások irányai is a bioinformatika két alapelveinek – a mintázat-felismerésnek és predikciónak – megfelelően változnak meg: az egyes gének funkcióját meghatározva, illetve az eddigi funkcionális információk ismeretében új gé-

nek szerepét előrejelezve. Azonban hiába nyújt bőséges információt többek között az *Arabidopsis* teljes genetikai állományából származó adattömeg, magában mégsem ad választ a növényi szerveződés, fejlődés- és élettan minden aspektusára. A bioinformatikai adatbázisok lényege pontosan ebben rejlik: mind általánosságban és összefüggéseket vizsgálva, mind egyes területeket, fajokat, illetve csoportokat nézve, specifikusan is kaphatunk számunkra fontos információt.

### Adatbázisok

A szekvenált nukleinsav-szekvenciák halma jelenleg is évről-évre exponenciálisan nő. Ezzel az irammal a fehérjék szekvenciáinak és térszerkezeteinek megállapítása nem tud lépést tartani (1. ábra); az ezredfordulón naponta 30 újabb, tisztázott szerkezetű fehérje szekvenciája került be az adatbázisokba, míg ez nukleinsav esetén körülbelül napi 20000 szekvenciát jelentett (BENSON et al. 2004, SZILÁGYI 2001).



1. ábra. A GenBank nukleotid- (A.) és a Swiss-Prot fehérjeszekvencia (B.) adatainak növekedési üteme  
 Figure 1. Growth of number of sequences in GenBank (A) and in Swiss-Prot (B) Database.  
 GenBank: number of sequences and total number of nucleotides (1000). Swiss-Prot: number of sequences and total number of amino acids (1000)

A helyzet 2004 végén sem változott jelentősen: nagyjából napi átlag 23 ezer új nukleotid (<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>) – és 80 annotált fehérjeszekvenciával ([http://www.expasy.org/sprot/relnotes/sp\\_news.html](http://www.expasy.org/sprot/relnotes/sp_news.html)) nő az adatállomány. Az adatbázisok eredetük szerint lehetnek elsődlegesek; ezek adataikat közvetlenül, non-profit és profitorientált kutatóintézetek szekvenálási projektjeiből nyelik, illetve másodlagosak és harmadlagosak, melyek az elsődleges adatbázisok alapján származtatják adataikat. Léteznek külön nukleotid- és fehérjeszekvenciákat tartalmazó adatbázisok is, illetve taxonómiailag, makromolekuláris szerkezetre, illetve egyéb szempontok szerint nézve specifikusak is. Az adatbázisokról minden év januárjában a Nucleic Acid Research különszámot jelentet meg, ebben minden aktuális adatbázist felsorol, illetve az új eredményekről, újonnan létrehozott adatbázisokról is hírt ad. Maguknak az adatbázisoknak is létezik adatbázisuk. A DBCAT nevű meta-adatbázist a <http://www.infobiogen.fr/services/dbcat> címen találjuk, a 2005. januári állapot szerint 511 különböző nukleotid, fehérje, strukturális, genetikai térképpel és anyagcserével foglalkozó adatbázist tartalmaz.

Bármilyen későbbi kutatások szempontjából elengedhetetlen, hogy a közreadott szekvenciák standardizált formában is megjelenjenek. Ez lehet:

a. “raw”, azaz nyers formátum, melynél a karakterek sorra egy-egy nukleotidot, illetve aminosavat jelölnek,

b. “FASTA” formátumú, melynél a raw formátumhoz hasonló, de a 70 betűnként tagolt karakterláncot egysoros jellemzés előzi meg, valamint egy “accession number”, hozzáférési azonosító, mely egy szakaszra egyedileg jellemző. Esetenként egy szekvenciához több azonosító is tartozik, ez az adatbázisok redundanciáját okozza,

c. egyedi, adatbázisonként különböző, azokra egyedien jellemző formátum, (lásd “GenBank flat file” formátum elemzését a Nukleotid adatbázisok fejezetben) melyből a szekvencián túl megtudhatjuk annak eredetét, verziószámát, a hozzá kapcsolódó publikációkat, valamint a szekvencia összetételét.

## Nukleotid adatbázisok

Az elsődleges, nukleotidalapú információkat, a megszekvenált szakaszokat az International Nucleotide Sequence Database Collaboration tagjai, amerikai National Center for Biotechnology Information (NCBI) által fenntartott GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>), a pán-európai European Bioinformatics Institute (EBI) adatbázisa, az EMBL (<http://www.ebi.ac.uk/embl/>), és a japán National Institute of Genetics adatbázisa, a DNA Data Base of Japan (DDBJ) (<http://www.ddbj.nig.ac.jp/>) gyűjtik össze (WHEELER et al. 2004). Mivel ezek adataikat naponta egyeztetik, gyakorlatilag ugyanazt az információt találhatjuk meg bennük.

Minden szekvenciát önálló bejegyzésként, úgynevezett GenBank formátumban tárolnak. Egy-egy ilyen file a következő adatokat foglalja magába:

LOCUS : tartalmazza a lókuszt nevet, a szekvencia hosszát és típusát, melyik GenBank részlegbe tartozik és a flat file legutolsó módosításának dátumát.

DEFINITION: a szekvencia rövid leírása, mely megjelöli a szekvencia forrásaként szolgáló organizmust, a gén/fehérje nevet vagy a szekvencia valamely feltételezett vagy bizonyított funkcióját, ha nem kódoló szekvenciáról van szó. Ha a szekvencia kódoló régió, a leírás utalhat a szekvencia komplettálásának aktuális állapotára.

**ACCESSION:** A szekvenciabejegyzés egyedi azonosítója. Rendszerint egy betű – öt számjegy (pl.: U12345), vagy két betű – hat számjegy kombinációjából tevődik össze (pl.: AF123456). Az accession number sosem változik: még ha a szerző később módosítja is a bejegyzést, ez az adat marad az állandó jellemzője egy-egy szekvenciának.

**VERSION:** Nukleotid szekvencia azonosító szám, mely egyedi, specifikus szekvenciát jelöl a GenBank adatbázisban. Ha bármilyen változtatás történik a szekvencia adatban (legyen szó akár egyetlen bázis cseréjéről is), a verziószám növekedni fog. (pl: U12345.1 > U12345.2) A flat file VERSION része tartalmazza az úgynevezett “GenInfo Identifier” (GI) szekvencia azonosító számot. Bármely szekvencia változtatás esetén új GI számot kap a bejegyzés.

**KEYWORDS:** A szekvenciát leíró pár kulcsszó vagy rövid mondat. Ha nem tartalmaz kulcsszavakat a bejegyzésünk, csak egy pontot találunk helyette.

**SOURCE:** A szekvencia forrásaként szolgáló organizmus rövidített neve, néha a molekulatípus követi. Alatta találjuk a flat file ORGANISM részét, mely a forrás taxonómiai besorolását jelöli, egységes szabályrendszer alapján, így később a bejegyzés könnyen kereshetővé válik a GenBank Taxonómiai Adatbázisában (Taxonomy Database <http://www.ncbi.nlm.nih.gov/Taxonomy> ).

**REFERENCE:** A szekvenciát jegyző szerzők azon publikációjára utaló rész, amely a szekvenciával foglalkozik. Természetesen találhatóunk publikálatlan, illetve “in press” státuszú szekvenciákat is.

**AUTHORS:** A szekvenciát jegyző szerzők neve.

**TITLE:** A publikált munka címe, illetve publikálatlan munka esetén átmeneti cím.

**JOURNAL:** Publikált munka esetén a cikket közlő újság MEDLINE formátumú rövidítése.

**MEDLINE:** A publikáció MEDLINE adatbázisbeli UID egyedi azonosítója.

**PUBMED:** A publikáció PUBMED adatbázisbeli PMID egyedi azonosítója.

**FEATURES:** A gének és a géntermékek felépítése, a szekvencia tagolása (például kódoló – nem kódoló régiókra bontva), illetve a biológiai szignifikanciák jelölését találhatjuk itt.

**BASE COUNT:** A, C, G, T bázisok száma a szekvenciában.

**ORIGIN:** Maga a szekvencia, FASTA-hoz hasonló formátumban.

## Fehérje-adatbázisok

Az elsődleges nukleotid-adatbázisokhoz hasonlóan épülnek fel az elsődleges fehérjeszekvencia-adatbázisok is. Három nagyobb elsődleges fehérjeszekvencia-adatbázist ismerünk.

1. Protein Information Resource – PIR (<http://pir.georgetown.edu/> ). Az 1988 óta, a PIR-International konzorcium (NBRF, Japanese International Protein Information Database, Munich Information Center for Protein Sequences) által fenntartott adatbázis négy tagozatra oszlik. Az elsőben (PIR1) a kísérletesen visszaigazolt, annotált szekvenciákat, a másodikban (PIR2) a részlegesen igazolt, előzetes szekvenciákat, a harmadikban (PIR3) a nem ellenőrzött szakaszok adatait, míg a negyedik (PIR4) szekcióban a mesterseges, természetesen nem kifejeződő szakaszok adatait találjuk. 2004 szeptemberében 36300 PIR szupercsaládot, 145300 családot, 7310 domént, 1300 motívumot, 280 poszt-transzlációs módosítást és több mint 50, más biológiai adatbázis felé mutató linkgyűjteményt tartalmazott a több, mint 1230000 szekvencián kívül.

2. Az 1986-ban alapított, jelenleg az European Bioinformatics Institute és a Swiss Institute of Bioinformatics által üzemeltetett Swiss-Prot 2005. január elején 167089 magas szinten annotált fehérjeszekvenciát tartalmazott. (<http://us.expasy.org/sprot/>) A Swiss-Prot célja a minőségi szolgáltatás: csakis többször visszaigazolt, ismert fehérjeszekvenciákat vesz fel adatbázisába, és azokat több, más adatbázisok felé irányuló kapcsolódási ponttal látja el. Egyébiránt található itt információt a másod-, illetve harmadlagos szerkezetéről, doménszerveződéséről, poszttranszlációs módosítási lehetőségekről is. Jóllehet az adatbázis kevesebb szekvenciát, de egyben minimális redundanciát tartalmaz. A Swiss-Prot-hoz tartozik a TrEMBL adatbázis, mely az EMBL nukleotid-adatbázisból számítógépesen származtatott adatokat tartalmazza, és a visszaigazolt szekvenciákat áthelyezi a Swiss-Prot adatbázisba.

3. A Protein Data Bankból (PDB, <http://www.rcsb.org/pdb>) kivont szekvenciákat tartalmazza az NRL3D (<http://pir.georgetown.edu/pirwww/dbinfo/nrl3d.html>). Itt csak a PDB-ben ténylegesen szerkezettel rendelkező aminosavakat találjuk, alapos egyéb jellemzőkkel ellátva. Az eredeti PDB-ben térszerkezeti információkkal ellátott fehérjéket találhatóunk, de ez erősen redundáns, mivel egy fehérje más-más térszerkezettel rendelkezhet, a különböző körülmények függvényében.

2002-ben a Swiss-Prot, PIR és TrEMBL adatbankokban megtalálható információk egyesítésével alkották meg az UniProt fehérje adatbázist (Universal Protein Resource, <http://www.uniprot.org>). Ez az adatbázis három részlegre osztható, külön célok szerint optimalizálva. Az UniProt Knowledgebase (UniProt) a fehérjék széleskörű, szabatos információinak központi gyűjtőhelye. Ebben a részben találunk információkat a funkcióról, fehérjecsaládbeli klasszifikációról és keresztreferenciákat. Az UniProt Non-redundant Reference (UniRef) a gyorsabb keresést teszi lehetővé azért, hogy a rokon szekvenciákat egyetlen bejegyzéssé kombinálja, csökkentendő a redundanciát. Az UniProt Archive (UniParc) az összes fehérjeszekvencia történetének közzétehető nem redundáns tárháza.

### **Speciális növényi adatbázis: TAIR - The Arabidopsis Information Resource**

Az egyes szakterületeken dolgozó kutatóknak nem feltétlenül van szükségük az elsődleges adatbázisokban felhalmozott összes információra, olykor a specifikus adatok közti, szűkített keresés hamarabb eredményre vezet. Ezért több, az elsődleges adatbázisok adatain alapuló specifikus adatbázist is kialakítottak az adott szakterületeken dolgozó kutatócsoportok (2. táblázat). A speciális adatbázisok felépítése és működése egy konkrét példa, a TAIR Arabidopsis Információs Adatbázis alapján jól szemléltethető (RHEE et al. 2003, GARCIA-HERNANDEZ et al. 2002). Az *Arabidopsis thaliana* az elmúlt húsz év során vált igazán kedvelt modellnövényé, gyors életciklusa, könnyű genetikai módosíthatósága, napjainkban pedig elsősorban a teljes méretében megszekvenált genomja miatt. A Brassicaceae család tagja olyan haszonnövényekkel rokon, mint a karfiol, fejes káposzta és a brokkoli, így több, mint 11000 kutató használja közel 5000 akadémiai és profitorientált kutatólaborban világszerte a legkülönbözőbb kutatási területeken, mezőgazdasági, energetikai vagy környezetvédelmi kérdések gyakorlati megválaszolása céljából.



Nagyobb növényi adatbázisok (a Nucleic Acid Research nyomán)  
Main databases of plants: name, short Hungarian description, URL  
(based on Nucleic Acid Research)

Adatbázis neve	Rövid leírása	Internetes elérhetőség
ARAMEMNON	<i>Arabidopsis thaliana</i> membránfehérjék és transzporterek	<a href="http://aramemnon.botanik.uni-koeln.de/">http://aramemnon.botanik.uni-koeln.de/</a>
AthaMap	A vélhető transzkripció faktorok kötődési helyeinek térképe az <i>Arabidopsis</i> genomban	<a href="http://www.athamap.de/">http://www.athamap.de/</a>
BGI-RISe	Pekingi genomikai intézet rizs információs rendszere	<a href="http://rise.genomics.org.cn/">http://rise.genomics.org.cn/</a>
CADRE	Központi <i>Aspergillus</i> adattár	<a href="http://www.cadre.man.ac.uk/">http://www.cadre.man.ac.uk/</a>
CATMA	Komplett <i>Arabidopsis</i> transzkriptom microarray: génszekvencia elemek	<a href="http://www.catma.org">http://www.catma.org</a>
COGEME	Fitopatogén és petesporás gombák EST adatbázisa	<a href="http://cogeme.ex.ac.uk">http://cogeme.ex.ac.uk</a>
CropNet	Haszonnövények géntérképjei	<a href="http://ukcrop.net/">http://ukcrop.net/</a>
DoOP	Gerincesek és növények közötti ortológ promóterek adatbázisa	<a href="http://doop.abc.hu/">http://doop.abc.hu/</a>
FLAGdb/FST	<i>Arabidopsis thaliana</i> T-DNA transzformánsok	<a href="http://genoplante-info.infobiogen.fr/">http://genoplante-info.infobiogen.fr/</a>
FLAGdb++	Növényi genomok integrált adatbázisa	<a href="http://genoplante-info.infobiogen.fr/FLAGdb/">http://genoplante-info.infobiogen.fr/FLAGdb/</a>
GénoPlante-Info	A GénoPlante konzorcium növényi genomikai adatai	<a href="http://genoplante-info.infobiogen.fr/">http://genoplante-info.infobiogen.fr/</a>
GrainGenes	A búza, árpa, rozs, zab és egyéb gabonafélék molekuláris és fenotipikus információi	<a href="http://wheat.pw.usda.gov">http://wheat.pw.usda.gov</a> or <a href="http://www.graingenes.org">http://www.graingenes.org</a>
INE	Integrált rizs genom böngésző	<a href="http://rgp.dna.affrc.go.jp/giot/INE.html">http://rgp.dna.affrc.go.jp/giot/INE.html</a>
IRIS	Nemzetközi rizs információs rendszer	<a href="http://www.iris.irri.org/">http://www.iris.irri.org/</a>
KaryotypeDB	Növény- és állatfajok kariotípusa és kromoszomális információja	<a href="http://www.nenno.it/karyotypedb/">http://www.nenno.it/karyotypedb/</a>
LOX-DB	Növényi, gomba, gerinctelen és emlős lipoxigenázok	<a href="http://www.dkfz-heidelberg.de/spec/lox-db/">http://www.dkfz-heidelberg.de/spec/lox-db/</a>
MagnaportheDB	<i>Magnaporthe grisea</i> integrált fizikai-genetikai térképe	<a href="http://www.fungalgenomics.ncsu.edu/Projects/mgdatabase/int.htm">http://www.fungalgenomics.ncsu.edu/Projects/mgdatabase/int.htm</a>
MaizeGDB	Kukorica genomikai adatbázis, (a MaizeDB és ZmDB utódja)	<a href="http://www.maizegdb.org/">http://www.maizegdb.org/</a>
MAtDB	MIPS <i>Arabidopsis thaliana</i> adatbázis	<a href="http://mips.gsf.de/proj/thal/db">http://mips.gsf.de/proj/thal/db</a>
Mendel	Növényi EST és STS szekvenciák adatbázisa és géncsalád-információk	<a href="http://www.mendel.ac.uk/">http://www.mendel.ac.uk/</a>
MGI	<i>Medicago</i> genomi adatbázis: EST-k, génexpressziós és proteomikai adatok	<a href="http://xgi.ncgr.org/mgi">http://xgi.ncgr.org/mgi</a>
MNCDB	MIPS <i>Neurospora crassa</i> adatbázis	<a href="http://mips.gsf.de/proj/neurospora/">http://mips.gsf.de/proj/neurospora/</a>

2. táblázat folytatása  
contd Table 2

Adatbázis neve	Rövid leírása	Internetes elérhetőség
MOsDB	MIPS <i>Oryza sativa</i> adatbázis	<a href="http://mips.gsf.de/proj/rice">http://mips.gsf.de/proj/rice</a>
MPIMP	Mitokondriális fehérje import mechanizmusok a növényekben	<a href="http://millar3.biochem.uwa.edu.au/~lister/index.html">http://millar3.biochem.uwa.edu.au/~lister/index.html</a>
MtDB	<i>Medicago trunculata</i> genom	<a href="http://www.medicago.org/MtDB">http://www.medicago.org/MtDB</a>
Oryzabase	Rizs genomika és genetika	<a href="http://www.shigen.nig.ac.jp/rice/oryzabase/">http://www.shigen.nig.ac.jp/rice/oryzabase/</a>
PGC Database	<i>Phytophthora infestans</i> és <i>P. sojae</i> EST adatbázis	<a href="https://xgi.ncgr.org/pgc">https://xgi.ncgr.org/pgc</a>
PHYTOPROT	Növényi fehérjecsaldók	<a href="http://genoplante-info.infobiogen.fr/phytoprot">http://genoplante-info.infobiogen.fr/phytoprot</a>
PLACE	Cisz-regulátor DNS elemek a növényekben	<a href="http://www.dna.affrc.go.jp/htdocs/PLACE">http://www.dna.affrc.go.jp/htdocs/PLACE</a>
Plant snoRNA DB	snoRNS gének a gővényekben	<a href="http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home">http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snoRNA/home</a>
PlantCARE	Növényi promoterek és cisz-regulátor elemek	<a href="http://intra.psb.ugent.be:8080/PlantCARE/">http://intra.psb.ugent.be:8080/PlantCARE/</a>
PlantGDB	Az aktivan átírózó növényi genomi szekvenciák adatbázisa	<a href="http://www.plantgdb.org/">http://www.plantgdb.org/</a>
PLANTncRNAs	Nem-kódoló RNS-ek a növényekben	<a href="http://www.prl.msu.edu/PLANTncRNAs">http://www.prl.msu.edu/PLANTncRNAs</a>
PlantProm	Növényi RNS polimeráz II promoter szekvenciák	<a href="http://mendel.cs.rhul.ac.uk/">http://mendel.cs.rhul.ac.uk/</a>
PLMItRNA	Növényi mitokondriális tRNSek	<a href="http://bighost.area.ba.cnr.it/PLMItRNA/">http://bighost.area.ba.cnr.it/PLMItRNA/</a>
PPNEMA	rRNSek növényparazita nematodákban	<a href="http://bighost.area.ba.cnr.it/PPNEMA/">http://bighost.area.ba.cnr.it/PPNEMA/</a>
Rice PIPELINE	Rizs adatbázisok egyesítő eszköze	<a href="http://cdna01.dna.affrc.go.jp/PIPE">http://cdna01.dna.affrc.go.jp/PIPE</a>
RiceGAAS	Rizs genom automatikus annotáló rendszer	<a href="http://ricegaas.dna.affrc.go.jp/">http://ricegaas.dna.affrc.go.jp/</a>
RPD	Rizs proteom adatbázis	<a href="http://gene64.dna.affrc.go.jp/RPD/">http://gene64.dna.affrc.go.jp/RPD/</a>
SeedGenes	Az <i>Arabidopsis</i> fejlődésében résztvevő alapvető gének	<a href="http://www.seedgenes.org/">http://www.seedgenes.org/</a>
SGMD	Szójabab genomikai és microarray adatbázis	<a href="http://psi081.ba.ars.usda.gov/SGMD/default.htm">http://psi081.ba.ars.usda.gov/SGMD/default.htm</a>
Sputnik	Növényi EST csoportok és funkcionális annotációk	<a href="http://mips.gsf.de/proj/sputnik">http://mips.gsf.de/proj/sputnik</a>
TAIR	Az <i>Arabidopsis</i> információs forrás	<a href="http://www.arabidopsis.org/">http://www.arabidopsis.org/</a>
TIGR plant repeat database	A növényi genom repetitív szakaszainak osztályozása	<a href="http://www.tigr.org/tdb/e2k1/plant.repeats">http://www.tigr.org/tdb/e2k1/plant.repeats</a>
TropGENE DB	A trópusi haszonnövények (cukornád, banán és kókusz) genetikai és genomikai adatbázisa	<a href="http://tropgenedb.cirad.fr/">http://tropgenedb.cirad.fr/</a>

Az *Arabidopsis* teljes genom analízisének köszönhetően a kutatásokból származó adatmennyiség exponenciális méretekben nőtt meg az évek során. A 2005. januári állapot szerint 888000 nukleotid- és 34000 aminosav szekvenciát, 490170 klónt, 3700 genetikai markert, 406800 polimorfizmust és 24400 publikációt tartalmaz az adatbázis. Az adatok többsége eddig is rendelkezésre állt, azonban ezek az adatok különálló helyeken voltak csak elérhetőek. Az egységes, mindent átfogó, összefüggő, központi adatbázis hiányzott, ennek létrehozását az *Arabidopsis*-szal foglalkozó kutatói közösség is szorgalmazta (<http://www.arabidopsis.org/info/carnegieworkshop.jsp>). Az Arabidopsis Információforrás (TAIR – The Arabidopsis Information Resource) ezt a hiányzó szerepet kísérte meg betölteni.

A TAIR-t az NSF – National Science Foundation alapította azért, hogy összegyűjtse, felügyelje, és közreadja az *Arabidopsis* kutatások során felhalmozott információkat. Kiindulásul a korábbi Arabidopsis thaliana DataBase (AtDB)-t használták, melynek célja a szekvenálási projektek során a részfeladatok koordinálása, és az *Arabidopsis* genetikai és fizikai térképeinek integrálása volt. A TAIR ehhez képest új adattípusokat és analitikai eszközöket is kínál, ezáltal helyezve középpontba az Arabidopsis biológiájának összes aspektusát szemben a régebbi, csupán szekvenciákra összpontosító adatbázis perspektívájával. Így ma a TAIR egy többféle adattípusból álló, kereshető és összefüggő adatbázis. Az adatokat az interaktív MapViewer-rel lehet megtekinteni és különböző *in silico* eszközökkel elemezni. Az így kapott eredmény standardizált formátumban menthető el. Ezen kívül híreket és az Arabidopsis Genome Initiative-vel (AGI) kapcsolatos információkat, *Arabidopsis* laboratóriumi protokollokat, és hasznos internetes linkeket is találhatunk még a TAIR oldalain.

A TAIR oldalainak látogatottsága (<http://www.arabidopsis.org/usage>) az 1999 novemberében mért kezdeti 20000 hozzáférés/hó-ról egy éven belül a kilencszeresére nőtt, 2005 elején pedig az egymillió havi kattintást is eléri.

A TAIR adatbázisban tárolt főbb adattípusok a következők: gének, markerek, polimorfizmusok, illesztések (genetikai térképen elfoglalt pozíciók), szekvenciák, klónok, mutáns vonalak, fehérjék, expresszálandó termékek, kutatók és referenciák. Ezek egymással integrált és összefüggő adatbázisként épülnek fel. Az adatok mind a TAIR szöveg-alapú keresőjével, mind grafikus megjelenítéssel elérhetőek.

Új típusú adatokra is nyitott a TAIR adatbázisa, ezeket folyamatosan töltik fel, párhuzamosan az adatbázis bővítésével, hiszen magát is alkalmassá kell tenni ezen új adatok befogadására. Az ilyen újdonságokat a vezető hírek között találjuk meg, egyes esetekben pedig az *Arabidopsis* hírcsoportban (<http://www.arabidopsis.org/news/newsgroup.html>).

### **Az *Arabidopsis* genom és a tair adatai**

A TAIR tartalmazza az *Arabidopsis* teljes genomját, mely öt centromer régió által tagoltan 125 millió bázispár nagyságú.

A jelenlegi állapotról, a még tisztázatlan, hézagos klónkönyvtári részekről is naprakész információt kaphatunk (<http://www.arabidopsis.org/info/agicomplete.jsp>). A megszekvenált genom, melyet kezdetben az AGI szekvenáló partnerek annotáltak, jelenleg

strukturát és funkciót tekintve általános újra-annotálási folyamaton megy keresztül. Ezt a TIGR (<http://www.tigr.org/tdb/e2k1/ath1>) és a MIPS (<http://mips.gsf.de/proj/thal/db>) és a TAIR közösen végzi: a klónokból származó szekvenciákat illesztik egymás mellé.

A TAIR-ban megtalálható génekről szóló információ 2004 elejére elérte a 31 ezer lókuszt (ennek több, mint 90%-a fizikai lókuszt) és 37 ezer génmodellt (<http://www.arabidopsis.org/jsp/tairjsp/pubDbStats.jsp>). A TAIR fizikai lókuszt alatt azt a kromoszómárészletet érti, mely az Arabidopsis genomon belül egyetlen transzkripciósi egységnek feleltethető meg, és egyben egy gén minimális reprezentációját is képezi. A génmodellt magában a CDS-ben (kódoló szekvenciában) vagy transzkriptálódó szekvenciában definiálja. Azok a gének, melyek azonos helyzetet foglalnak el a genomon, de strukturális annotációjuk különböző, külön génmodelleket képeznek, de ugyanazzal a fizikai lókuszzal asszociáltak.

A fizikai lókusztok – azaz a genomszekvenálási projektekből származó gének – az elnevezésüket a kromoszómán elfoglalt helyzetük alapján kapják. A gének nevei a következő rendszer alapján épülnek fel: a szervezet nevét (AT) kromoszómális helyzet követi (1-5), majd egy G betű jelöli a gént, utána pedig egy egyedi azonosító szám jön (pl. AT1G36160). Ez az új nevezéktan lett az általánosan elfogadott és váltotta fel a régi nomenklatúrát, mely a BAC-klónok alapján készült. A génadat a nevével kívül tartalmazza az egyéb alternatív elnevezéseket, a szekvenciát, a leírást, a fizikai/genetikai térképen elfoglalt helyet, a lehetséges kapcsolt markereket, a tulajdonságokat és a publikációkat. Jelentős erőfeszítések zajlanak az álnevekből eredő, génadatok közti redundancia felszámolására. A géneket lehet keresni a nevük alapján (legyen ez szimbolikus, teljes, ORF vagy géntermék név), GenBank azonosító vagy leírás alapján. A keresést szűkíteni is lehet a legkülönbözőbb paraméterek (kromoszóma vagy térképen elfoglalt helyzet, vagy akár a kísérletes visszaigazoltság) alapján. A keresés eredményét a lókuszt neve és a génmodell alapján rendezve kapjuk meg. Ezek további adatbázisok felé mutatnak, így olyan információkhoz kapcsolnak, mint a gének részletes leírása, a lehetséges polimorfizmusok, markerek, géntérképen elfoglalt pozíció, géntermékek és szakirodalom.

## Protein és géncsalád adatok

A TAIR-hoz ma már részletes fehérje és géncsalád adatbázis is tartozik. A fehérje-adat tartalmazza az aminosav-szekvenciát, a várható fizikai-kémiai tulajdonságokat (mint például izoelektromos pont vagy molekulatömeg), a sejtbeli előfordulás helyét, a fehérjét funkcionális részeik alapján családokba osztó InterPro adatbázis (InterPro, <http://www.ebi.ac.uk/interpro/>) szerinti doménjeit, vagy a strukturák alapján osztályozó SCOP adatbázis (Structural Classification of Proteins, <http://scop.mrc-lmb.cam.ac.uk/scop/>) szuperfamilijait. Az adatbázisba az összes előre jelzett fehérje beletartozik. Az adatbázis tervezi olyan további információk beillesztését, melyek a poszt-transzlációs módosítási és a fehérje komplex-képzési hajlamra vonatkoznak. A fehérje-adatbázisban a <http://www.arabidopsis.org/tools/bulk/protein/index.html> cím alatt lehet keresni.

A géncsaládokra vonatkozó információ statikus weboldalon található meg, (6378 gént osztottak 860 fehérjecsaládba 2005. januári adatok alapján) ez irodalmi adatokra és kutatási eredményekre támaszkodik (<http://www.arabidopsis.org/info/genefamily/genefamily.html>). A géncsaládok felosztása mindenhol a fehérjehasonlóságon alapul, specifikus paraméterek szerint optimalizált szoftverekkel.

## **AraCyc: Anyagsere-folyamatok interaktív keresője**

A TAIR talán egyik leghasznosabb részét képezi az *Arabidopsis* anyagsere-folyamatainak grafikus megjelenítése, az AraCyc (MUELLER et al. 2003). Ebben a szekcióban egymás mellett „madártávlatból” láthatjuk a növény főbb anyagsere-folyamatainak ábrázolását. Ez az eszköz lehetővé teszi, hogy a ciklusok bármely pontjára kattintva, a kijelölt, a szűkebb folyamatról és az abban résztvevő enzimekről, kiindulási vegyületekről és a létrejött anyagsere-termékekről bővebb információt kaphassunk. Azonban az eszköz nem csupán az elnevezést és a molekulaképletet ismerteti, hanem kapcsolódási pontot is tartalmaz a protein, illetve nukleotid szekvenciákat tartalmazó szekciók felé, minden egyéb résztvevő gén, fehérje, kofaktor, reakció feltüntetésével. Ebben a szekcióban is lehetőség van név és azonosító alapján történő keresésre.

## **További TAIR eszközök**

A MapViewer (<http://www.arabidopsis.org/servlets/mapper>) nevű eszköz a különböző genetikai, fizikai és szekvencia-alapú térképeket jeleníti meg grafikusan, keresési lehetőségekkel együtt. A SeqViewerrel (<http://www.arabidopsis.org/servlets/sv>) egy adott szekvencia helyzetét tekinthetjük meg a genomon belül, a kromoszómákon belül a nagyító eszközzel közelíthetünk a legalsóbb, a nukleotidok szintjéhez. A felhasználó akár klón, marker vagy a gén neve alapján is kereshet. A TAIR további hasznos eszközei között megtalálhatjuk a többi adatbázisból is ismert FASTA (<http://www.arabidopsis.org/cgi-bin/fasta/nph-TAIRfasta.pl>), illetve a Basic Local Alignment Search Tool, azaz röviden BLAST (<http://www.arabidopsis.org/Blast/>) keresőprogramokat, melyekkel a szekvenciák hasonlósági keresését lehet végezni. (További információkat a FASTA programról a <http://www.ebi.ac.uk/2can/tutorials/nucleotide/fasta.html>, míg a BLAST program működéséről a <http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html> honlapokon találhatunk.) A PatMatch (<http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl>) programmal 30-nál rövidebb nukleotid és aminosav szekvenciák motívumait kereshetjük meg. Ez a konzervatív jellegű szakaszok felismerésénél hasznos eszköz, mellyel például olyan cisz regulátor elemek, vagy protein módosító helyek kereshetőek, melyeket a BLAST vagy a FASTA nehezen talál meg.

## **Hazai speciális adatbázis: a Lucerna Genetikai Csoport (Medicago Genetics Group)**

Kisebb méretű, de hazai vonatkozású a gödöllői Mezőgazdasági Biotechnológiai Központ által fejlesztett, a lucerna genommal foglalkozó *Medicago trunculata* adatbázis (<http://medicago.abc.hu>). Az adatbázis célja olyan webszerver felállítása volt, amely teljes genetikai térképet nyújt különböző (morfológiai, izoenzim, RFLP és RAPD) markerek segítségével. Az ilyen térkép hasznosnak bizonyul, ha különböző, mutációk okozta morfológiai jellegek vagy szimbiontákkal kapcsolatos (pl. Nod ) gének kromoszómán elfoglalt helyét kell meghatározni. A génekhez kapcsolódó molekuláris markerek segítségével, melyeket az adatbázis tartalmaz, a lucernára jellemző élettani folyamatok mole-

kuláris biológiai kutatása, illetve a későbbiekben térkép alapú szekvenálása válik lehetővé. Az adatbázis jelenleg tartalmazza a genetikai térképeket, egy BLAST keresőprogramot és kiterjedt linkgyűjteményt.

Szintén ehhez az adatbázis-projekthez tartozik az SSRDB, a Simple Sequence Repeat Database (<http://bioinformatics.abc.hu/ssr/>) is, amely még fejlesztés alatt áll, de már az eddigi eredmények elérhetőek nyilvánosan is. Az SSRDB olyan rövid, 1–6 ismétlődő bázisból álló mikroszatellit szekvenciákat tartalmaz, melyeket több eukarióta genomon megtalálhatunk, taxonokra jellemző variációkban.

## Egyéb bioinformatikai alkalmazások

A bioinformatikai adatbázisok web-alapú szekvencia-kezelő és analízáló szoftverein kívül természetesen léteznek olyan programok is, melyet a felhasználó a saját munkáállomására is telepíthet. Ezek a szoftverek és programcsomagok működhetnek mind WINDOWS operációs rendszert használó környezetben, mind LINUX operációs rendszert futtató gépeken, vagy akár mindkettőn. Az alábbiakban bemutatnánk pár ilyen szoftvert.

PHYLIP: A PHYLogeny Inference Package elnevezésű szoftver filogenetikai alkalmazás, evolúciós kapcsolatok feltárásánál használhatjuk ki előnyeit Windowst, Linuxot és Macintosht futtató gépeken egyaránt. Bővebb információt és ingyenes letöltést a <http://evolution.genetics.washington.edu/phylip.html> honlapon találhatunk.

BIOEDIT: Windows környezetre írt ingyenes szekvenciaelemző és szerkesztő program, egyszerűbb szekvencia-rendezésekre is alkalmas. Vizuális környezetét, könnyű kezelhetőségét hamar meg lehet kedvelni (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).

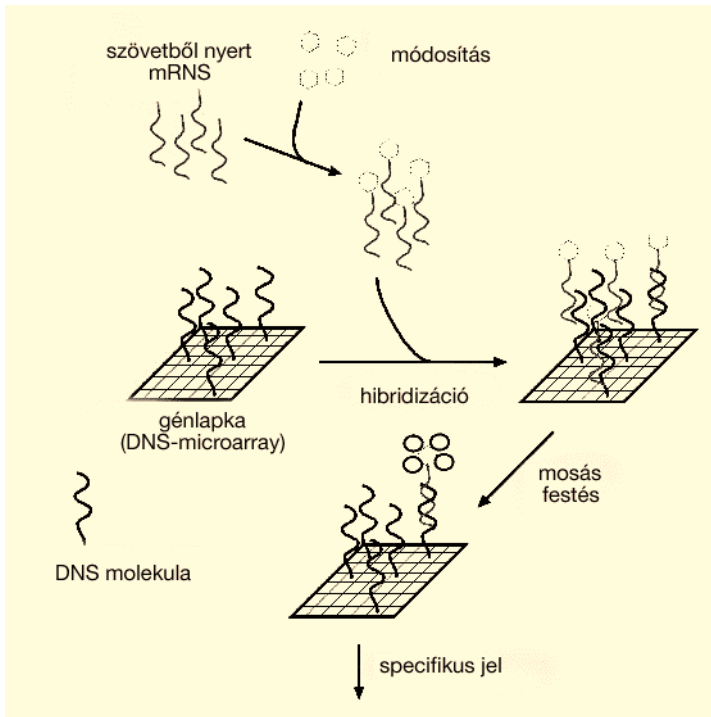
GENEDOC: A BIOEDIT-hez hasonló, de annál egyszerűbb, korlátozottabb lehetőségekkel bíró kis méretű program (<http://www.psc.edu/biomed/genedoc/>).

EMBOSS: European Molecular Biology Open Software Suite, nyílt forráskódú, tehát programozó ismeretekkel bírók számára egyedileg fejleszthető ingyenes programcsomag UNIX rendszerű gépekre. Több, mint száz alkalmazást találhatunk benne: segítségével elemezhetünk szekvenciákat, kereshetünk helyi adatbázisunkban, akár protein motívumokat is, sőt, kisebb genomok kodonhasználati analízisét is gyorsan elvégzi. Hátránya, hogy használata magabiztos UNIX programozói ismereteket igényel (<http://www.hgmp.mrc.ac.uk/Software/EMBOSS/overview.html>).

## Funkcionális genomika

A modern kutatásokban kikerülhetetlen a molekuláris biológiai megközelítés. Az 1980-as évek génvadászati korszakával szemben ma, a teljes genomok ismeretének korában a szekvencia adatok nem korlátoznak. Az adatbázisok ma is kiindulópontjait jelentik az egyes, különálló gének egyedi vizsgálatának. Legyen szó az egyes gének meglétének igazolásához használt PCR-technikákról (MULLIS et al. 1986), vagy génexpressziós vizsgálatokról (pl. Real-time PCR, illetve Northern-hibridizációs módszerek), az adatbázisból származó információk képezik a kísérletek alapját, tehát a molekuláris biológiai kísérletek megtervezésénél kikerülhetetlen az adatbázisok használata. Ma már azonban, az egyes gének vizsgálatával szemben, egyre elterjedtebb

a génműködések együttállásnak, azaz több ezer gén egyidejű expressziós mintázatának vizsgálata, a génlapkás, más néven microarray-technika (2. ábra).



2. ábra. A microarray (génlapkás) technika

Figure 2. Microarray technology: modified mRNA from tissue sample is hybridised with DNA microarray. The expression pattern-specific signal is measured after dyeing and washing the chip

Ennél a technikánál a szövetből nyert, általában mRNS mintát némi módosítás után hibridizáltatják egy pár négyzetcentiméteres darabon elhelyezett több ezer, ismert szekvenciájú DNS-szakasszal. Miután a szövetspecifikus mRNS hibridizált a génlapkával, mosás és festés után az adott fiziológiai állapotban „bekapcsolt gének” által kirajzolt mintázat, specifikus jel megjeleníthető. A genom projektekből származó információs tömeg a bioinformatika és a génlapka-technológia segítségével így alakítható át gyakorlati tudássá.

### A posztgenomi éra – rendszerbiológia

A teljes genetikai állományok feltárása utáni posztgenomi érában azonban nemcsak az mRNS-ek, fehérjék, struktúrális biopolimerek vagy metabolitok expressziós szintjének nagyszámú vizsgálata szükségeltetik, hanem a molekuláris változások időbeli felbontásának növelése is fontos feladat. A szöveti vizsgálatokon túl szintén perspektívát jelent az anyagcsere-folyamatok helyének sejtek szintjén történő meghatározása. Ez nagy mennyiségű adatot generáló, úgynevezett high-throughput mérési eljárások fejlesztésével lehetséges, melyek a mikroszkóp- és robottechnológiát sem nélkülözik. Az

ilyen nagy felbontású eszközök és eljárások megtervezése, valamint a pontos és összefüggő adatok gyűjteményének előmozdítása érdekében a jövő növénybiológusainak más tudományterület képviselőivel, vegyészekkel, informatikusokkal és tervezőmérnökökkel kell szorosan együttműködniük, újabb interdiszciplináris tudományterületeket megalapozva.

Mindezek után lehetőség nyílik a növényi fejlődésbiológiai és anyagcsere-folyamatok elméleti modellezésére is, így létrehozva egy *in silico*, tehát számítógépes alapon működő biológiai modellrendszert (RAIKHEL és CORRUIZ 2003). Ebben az *in silico* növényben különböző hatásokra bekövetkező, a teljes szervezetet érintő válaszreakciók összességét lehet vizsgálni. Így a rendszerbiológia remek eszközül szolgálhat a konkrét laboratóriumi munka lefolytatása nélküli kísérletek elvégzésére. A részfolyamatok, illetve különálló növényi komponensek funkciójának megértése helyett idővel előtérbe kerül a teljes élőlény összes folyamatának egyszerre történő vizsgálata. Jóllehet az *in silico* növények csak részben adhatnak feleletet az utóbbi évtizedekben jelentkező környezeti ártalmak következtében felmerült kérdésekre, vagy olyan problémákra, mint az újszerű energiaforrások keresése vagy az élelmezési gondok megoldása, de jelentős mértékben hozzájárulnak a nagymennyiségű biológiai információ gyakorlati alkalmazásához.

Tavaly ünnepeltük a DNS felfedezésének ötvenedik évfordulóját. Ötven év alatt a szemléletmód is megváltozott: gének egyenkénti vizsgálata helyett ma már a teljes genomok kerültek a kutatások homlokterébe. A posztgenomi érában a jövő kutatóinak olyan feladattal kell majd megküzdeniük, ami ma még elképzelhetetlen: a biológiai kirakós játék egyes darabjaiból dinamikusan működő, interaktív rendszert kell alkotni, így a klasszikus kísérletes növénybiológiát kiegészítheti, és az idők folyamán több vonatkozásban pedig teljes mértékben felválthatja egy *in silico* növényi modellrendszer.

#### IRODALOM – REFERENCES

- BENSON D. A., KARSCH-MIZRACHI I., LIPMAN D. J., OSTELL J., WHEELER D. L. 2004: GenBank: update. *Nucleic Acids Res.*, 32. Database issue: D23–26.
- ELLIS L. B., ATTWOOD T. K. 2001: Molecular biology databases: today and tomorrow. *Drug Discov. Today* 6: 509-513.
- GARCIA-HERNANDEZ M., BERARDINI T. Z., CHEN G., CRIST D., DOYLE A., HUALA E., KNEE E., LAMBRECHT M., MILLER N., MUELLER L. A., MUNDODI S., REISER L., RHEE S. Y., SCHOLL R., TACKLIND J., WEEMS D. C., WU Y., XU I., YOO D., YOON J., ZHANG P. 2002: TAIR: a resource for integrated *Arabidopsis* data. *Funct. Integr. Genomics* 2: 239–253.
- MEYEROWITZ E. M. 2001: Prehistory and history of *Arabidopsis* research. *Plant Physiol.* 125: 15-19.
- MUELLER L. A., ZHANG P., RHEE S. Y. 2003: AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132: 453–460.
- MULLIS K., FALOONA F., SCHARF S., SAIKI R., HORN G., ERLICH H. 1986: Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51: 263–273.
- PAUL A. L., FERL R. J. 1998: Higher order chromatin structures in maize and *Arabidopsis*. *Plant Cell* 10: 1349–1359.
- RAIKHEL N. V., CORUZZI G. M. 2003: Achieving the in silico plant. Systems biology and the future of plant biological research. *Plant Physiol.* 132: 404–409.
- RHEE S. Y., BEAVIS W., BERARDINI T. Z., CHEN G., DIXON D., DOYLE A., GARCIA-HERNANDEZ M., HUALA E., LANDER G., MONTOYA M., MILLER N., MUELLER L. A., MUNDODI S., REISER L., TACKLIND J., WEEMS D. C., WU Y., XU I., YOO D., YOON J., ZHANG P. 2003: The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* 31: 224–228.



- SZILÁGYI A. 2001: Bioinformatika, <http://www.enzim.hu/~szia/bioinformatika> .
- TERRY N., ROUZE P., VAN MONTAGU M. 1999: Plant genomics. *FEBS Lett.* 452: 3–6.
- WHEELER D. L., CHURCH D. M., EDGAR R., FEDERHEN S., HELMBERG W., MADDEN T. L., PONTIUS J. U., SCHULER G. D., SCHRIML L. M., SEQUEIRA E., SUZEK T. O., TATUSOVA T. A., WAGNER L. 2004: Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res.* 32. Database issue: D35–40.
- YUAN Q., QUACKENBUSH J., SULTANA R., PERTEA M., SALZBERG S.L., BUELL C.R.: Rice bioinformatics 2001: Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiol.* 125: 1166–1174.

## PLANT GENOMES

B. Jóri

Department of Plant Physiology and Molecular Plant Biology, Eötvös University,  
Budapest, Pázmány P. lane 1/C, H-1445, Hungary

Accepted: 29 December 2004

**Keywords:** genome, data bases, TAIR, bioinformatics, microarray, *in silico* plant

After sequencing the plant genomes plant biologists need to change their aspects and approach. Wide range of biological data and sequences are accessible in general and specific internet databases. Mining will be brought into the focus using bioinformatics and *in silico* methods. In monitoring gene expression micro arrays will become an enhanced role. Creating the *in silico* plant means a further prospect in modeling the plant life cycle and metabolism.