

Következő generációs szekvenálási technológiák kifejlődése és alkalmazásai

Mihály Zsuzsanna¹ ■ Gyórfy Balázs dr.²

Semmelweis Egyetem, Általános Orvostudományi Kar, ¹I. Gyermekgyógyászati Klinika,

²Gyermekgyógyászati és Nefrológiai Kutatócsoport, Magyar Tudományos Akadémia, Budapest

A következő generációs szekvenálási technológiák megjelenése az elmúlt tíz évben jelentős előrelépést jelentett a gyors és hatékony genomiális DNS-szekvenálás terén. Tanulmányunkban áttekintjük az 1975-ös sangeri láncterminációs szekvenálástól a valós idejű DNS-szekvenálás lehetővé válásáig vezető módszertani vívmányokat. A klonális amplikonokkal dolgozó, sok szálon párhuzamosan futó szekvenálási módszerek a következő generációs szekvenálási technológiák alapjai. Manapság leginkább a funkcionális genomikai alapú kutatásban alkalmazzák ezen szekvenálási technológiákat, amelyek a szignáltranszdukciós útvonalak, ontológiák, a proteomikai, metabolomikai eredményeknek a metaanalízise során nélkülözhetetlen információt adnak. Bár klinikumban rutinszerűen még csak elvétve alkalmaznak következő generációs szekvenátorokat, azonban az onkológiában, kardiológiában és epidemiológiában már van igény a technológia által elérhető extra ismeretekre. Az elterjedés fő gátja az adatelemzési módszerek standardizáltságának hiánya, amely az objektív kiértékelést megnehezíti. *Orv. Hetil.*, 2011, 152, 55–62.

Kulcsszavak: következő generációs szekvenálási technológiák (KGST), DNS-szekvenálás, funkcionális genomika, onkológia, génpolimorfizmusok (SNP)

Next generation sequencing technologies (NGST) development and applications

In the past ten years the development of next generation sequencing technologies brought a new era in the field of quick and efficient DNA sequencing. In our study we give an overview of the methodological achievements from Sanger's chain-termination sequencing in 1975 to those allowing real-time DNA sequencing today. Sequencing methods that utilize clonal amplicons for parallel multistrand sequencing comprise the basics of currently available next generation sequencing techniques. Nowadays next generation sequencing is mainly used for basic research in functional genomics, providing quintessential information in the meta-analyses of data from signal transduction pathways, ontologies, proteomics and metabolomics. Although next generation sequencing is yet sparsely used in clinical practice, cardiology, oncology and epidemiology already show an immense need for the additional knowledge obtained by this new technology. The main barrier of its spread is the lack of standardization of analysis evaluation methods, which obscure objective assessment of the results. *Orv. Hetil.*, 2011, 152, 55–62.

Keywords: next generation sequencing technologies (NGST), DNA sequencing, functional genomics, oncology, SNP

(Beérkezett: 2010. október 7.; elfogadva: 2010. november 12.)

Rövidítések

CNV = (copy number variation) kópiaszám-variációk; ChIP-Seq = (chromatin immunoprecipitation sequencing) kromatin-immunoprecipitációs szekvenálás; DGE = (dynamic gene expression) dinamikus génexpresszió; EGFR = (epidermal growth factor receptor) növekedésfaktor-receptor; GWAS = (genome-

wide association studies) teljesgenom-asszociációs tanulmány; ICGC = (International Cancer Genome Consortium) Nemzetközi Rák Genom Konzorcium; IHGSC = (International Human Genome Sequencing Consortium) Nemzetközi Humán Genom Szekvenálási Konzorcium; KGST = következő generációs szekvenálási technológiák; PARS = (parallel analysis

of RNA structure) RNS struktúrájának párhuzamos analízise; PME = (pulsed multiline excitation) pulzáló többsoros gerjesztés; SMRT-szekvenálás = (single molecule real time) egyes molekulák valós idejű szekvenálása; SNP = (single nucleotide polymorfism) génpolimorfizmus; sRNS = (short RNS) rövid RNS

Jóllehet az elmúlt negyven évben számtalan különböző eljárást dolgoztak ki a DNS nukleotid sorrendjének megismeréséhez, ennek ellenére jelenleg is újabbnál újabb technológiákat fejlesztenek ki az egyre pontosabb, gyorsabb és olcsóbb leolvasás érdekében. A következő generációs szekvenálási technológiák (KGST) alkalmazása először a tudományos alap kutatásban terjedt el, ám manapság a klinikumban is felhasználhatók a nagy teljesítményű szekvenálás segítségével gyorsan kinyerhető és pontos genomikus adatok.

A genomszekvenálás eddig legismertebb mérföldköve az ezredfordulóra befejeződő Humán Genom Projekt volt. 1996 és 2001 között két párhuzamosan, de egymástól nem teljesen függetlenül dolgozó csoport – Nemzetközi Humán Genom Szekvenálási Konzorcium (IHGSC) és a Celera kutatói – a teljes humán genomot feltérképezte [1, 2]. A 2001 februárjában nyilvánosságra hozott nyers emberi genomszekvenancia azonban lényegében csak „egyetlen” verzió, ami nem tartalmazza a homo sapiens variabilitásait. Bár azóta is lázasan folyik az információk pontosítása, az adatbázisok belátható időn belül nem lesznek teljesen készek.

A KGST-k már képesek mindössze néhány óra alatt több százmillió bázist leolvasni, azonban még ez sem elegendően gyors a rutinszerű orvosi felhasználáshoz. Ezért a nagy tudományos áttöréseket támogató X PRIZE alapítvány 10 millió dollárt ajánlott fel az első csoportnak, amelyik képes tíz nap alatt száz genomot megszekvenálni, egyenként kevesebb mint 10 ezer dollár költséggel (<http://genomics.xprize.org/>).

A genomszekvenálás módszerei

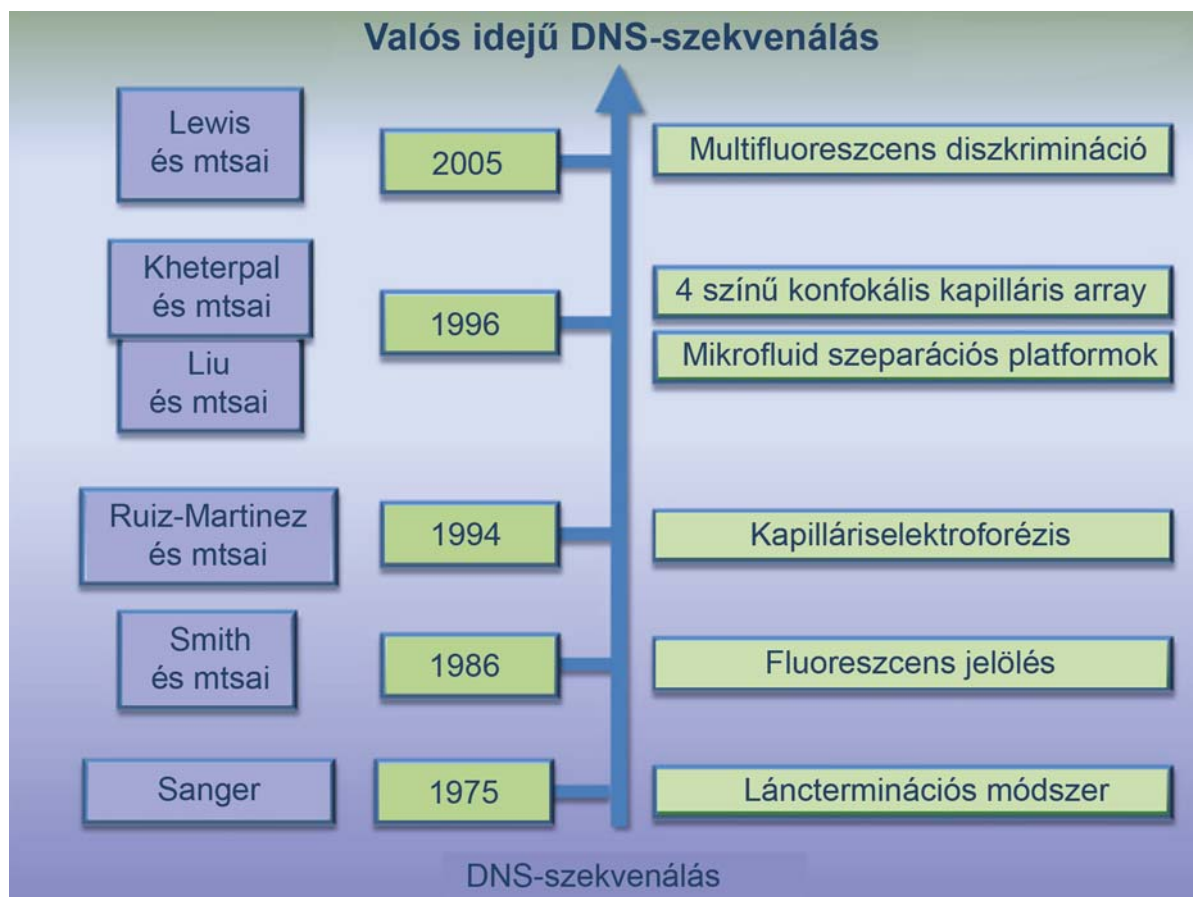
A kezdetektől az új generációs szekvenálásig

Az 1970-es években *Frederick Sanger* fejlesztette ki az első DNS-szekvenálási technikát, a későbbiekben tökéletesített láncterminációs módszert [3], amiért 1980-ban Nobel-díjjal jutalmazták. A Sanger-féle klasszikus láncterminációs módszer radioaktívan jelölt DNS primer, egyszálú DNS templát, DNS polimeráz, valamint a dezoxi- és radioaktívan jelzett didezoxinukleotidok segítségével szekvenált DNS-t. A 4 nukleotidnak (A, C, G, T) megfelelően 4 külön csőben különböző hosszúságú DNS-fragmentek képződtek a szálon leálló szintézis nyomán a didezoxinukleotid-beépülés miatt. A négy nukleotid alapján párhuzamos futtatott lemezekon végzett gélelektroforetikus szeparációt követte az autora-

diográfiai előhívás. Így 24 órával a kísérlet megkezdését követően manuálisan olvashatóvá váltak a DNS-szekvenációk. Ugyan így csak 500 bázist lehet leolvasni egyszerre, ennek ellenére az elvét napjainkban is még mindig használják.

1986-ban, majdnem tíz évvel Sanger után, *Smith és munkatársai* vezették be a DNS-darabok fluoreszcens jelölését [4]. A négy különböző bázishoz más-más színű fluorofort kovalensen kötöttek, így egy közös csőben egyszerre lehetett elvégezni a szintézist, majd a gélelektroforézis után kapott úgynevezett kromatogramból számítógép segítségével nyerték ki a szekvenációkat. Az 1994-ben bevezetett új, 3' végén módosított dezoxinukleotid-trifoszfát tette lehetővé a bázisspecifikus terminációt és a 3' protektív csoport hatékony fotolitikus eltávolítását. Így lehetővé vált az újbóli DNS-szintézis [5], ami az alapjává vált a későbbiekben alkalmazott sokciklusos szekvenálásnak. A 90-es évek másik újtása a kapilláris-elektroforézis-technika kifejlesztése, amely során lineáris poliakrilamidot használtak szűrőmátrixként a DNS-szekvenáláshoz [6]. Ennek segítségével igen kis DNS-mennyiségek is rövidebb idő alatt szétválaszthatóvá váltak, és több mint 1000 bázis vált leolvashatóvá 80 perc alatt. 1995-ben az addigiaknál megbízhatóbb olyan fluoreszcens jelölőfestéket fejlesztettek ki, ami az energiatranszfer révén optimálisabb abszorpciós és emissziós értékekkel rendelkezett [7]. Ezt követte még egy újabb automatizált szekvenálásra is alkalmas fluoreszcens festék megjelenése [8]. 1996-ban *Kheterpal és munkatársai* már egy olyan 4 színű konfokális kapilláris array olvasót használtak szekvenálásaikhoz, amely automatikusan végezte a szekvenálást [9]. Majd megindult az adatfeldolgozási folyamatokhoz szükséges szoftverek fejlesztése és tesztelése is [10, 11].

Az új évezred első éveiben mindenki az új nagy teljesítményű szekvenálási eljárások megjelenéséről beszélt, mivel a mikrofluid szeparációs platformok elterjedésével szinte minden adottá vált a következő generációs szekvenátorok kifejlesztéséhez. *Liu és Schmalzing* 1999-ben még csak egycsatornás mikrogyártású géppel dolgoztak, azonban a leolvasott bázispárok tekintetében messze meghaladták elődeiket [12, 13]. *Liu* egy évvel később már egy 16 csatornás készüléket mutatott be [14], de ebben az évben megjelentek 48 csatornás 400 [15] és 640 bázishosszúságban [16] szekvenáló, valamint 32 csatornás 800 bázishosszúságú [17] szakaszokat szekvenáló gépek is. A *Lewis* által kifejlesztett pulzáló többsoros gerjesztés (pulsed multiline excitation, PME) segítségével lehetővé vált a multifluoreszcens diszkrimináció, így a többkomponensű fluoreszcens assay-k színvak módon mérhetővé váltak [18]. E módszer felhasználásával a valós idejű DNS-szekvenálás is lehetővé vált. A valós idejű DNS-szekvenálás technológiájának kifejlődéséhez vezető legfontosabb módszertani vívmányokat az *1. ábra* foglalja össze. Néhány évtized alatt nemcsak az egyszerre leolvasott bázishossz



1. ábra

A sangeri szekvenálástól a valós idejű szekvenálásig vezető módszertani vívmányok összegzése.

A későbbiekben kifejlesztett KGST létrejöttéhez az 1975-ös sangeri láncterminációs módszer [3] után harminc évvel a valós idejű szekvenálás technológiájának megjelenésére volt szükség. Ezt a módszertani fejlődést mutatja be az ábra. A Smith és munkatársai által bevezetett fluoreszcens jelöléssel [4], majd az 1994-ben Ruiz-Martinez és munkatársai nyomán megjelent kapilláriselektroforézis [6] már 20 év alatt 1000 bázispár megszekvenálásának időigényét 80 percre csökkentette. Ezt követően még tíz évre és a 4 színű konfokális kapilláris array-k [9] és a mikrofluid szeparációs platform [13] létrejötte is elengedhetetlen volt, hogy a multifluoreszcens diszkriminációs eljárás [18] segítségével létrejöhessen a valós idejű DNS-szekvenálás 2005-re

nőtt 1000-re a másodpercenkénti 24 bázis leolvasása mellett, hanem a 10 dollárból leolvasható bázisok száma is az 1985-ös egyről 10 000 bázisra emelkedett [19].

Következő generációs szekvenálási technológiák (KGST)

Az ezredfordulót követően a DNS-könyvtárakból klonálisan amplifikált DNS-molekulákat különböző módszerek segítségével szekvenáló (454 Life Science, Illumina és Applied Biosystems) következő generációs szekvenálási technológiák (KGST) jelentek meg. Ezekben közös, hogy sok szálon párhuzamosan folyik a DNS szekvenálása. Összehasonlítva a klonális amplitonokkal dolgozó KGST-t a sangeri módszerrel, az előbbieknél a rövidebb leolvasási hossz és a lassabb szekvenáció ellenére a párhuzamos amplitonfeldolgozásnak köszönhetően a végteljesítmény nagyobb, akár több ezer bázis másodpercenként, valamint az egy bázisra számolt költség is alacsonyabb [20, 21].

A Roche 454 Sequencing szekvenátorai a „sequencing by synthesis” módszerét alkalmazzák, amellyel 500

millió bázisnyi nyers szekvencia képezhető le néhány óra alatt. Az egyszálú DNS-t egy gyöngyre kötik, és PCR-reakcióval milliónyi klónná sokszorozítják. A gyöngyöket ezután egy méhkaptárszerű lapra helyezik, amelyen apró lyukak (PicoTiterPlate) vannak, ahova egyszerre csak egyetlen gyöngy fér be. Itt piroszekvenálás segítségével olvassák le a szekvenciát [22]. A reakció során a polimeráz meghosszabbítja a DNS-szálat a platekhez ciklikusan hozzáadott fluoreszcens jelet adó nukleotidokkal [23]. Végül a ciklus végén lemoszák a fluoreszcens nukleotidot, és a következő ciklusban egy újabb nukleotidot épít be a polimeráz. A sikeres nukleotidbekötődést fotonemisszió jelzi, amit egy CCD-kamera detektál. Végül az adatfeldolgozást követően válnak megismerhetővé a szekvenciák (www.454.com).

Az Illumina a „pair-end” szekvenálási módszert alkalmazza [24]. Először DNS-könyvtárakat hoznak létre úgynevezett híd-amplifikáció révén. Az egyszálú DNS-darabok végére kötött adapterveződések segítségével oligonukleotid horgonyokhoz hibridizálva immobilizálják a DNS-fragmentumokat. A lekötött egyszálú DNS másik szálát primerek segítségével megszüntetizálják és a

szabad végükre kötött adaptervégződésel hídszerűen meghajlítva kihorgonyozzák a nukleotidszálakat. Az amplifikációt követően a reverz szálakat eltávolítják és a csoportokban klonálisan amplifikált nukleotidszálon végzik a szekvenálást fluoreszcensen jelölt nukleotidok segítségével. Minden egyes nukleotid kötődése után detektálják a fluoreszcens jelet, végül lemossák a festéket, mielőtt egy újabb ciklusba kezdenek (www.illumina.com).

Az ABI „mate-paired library” a teljesgenom-szekvenálás mellett még célzott reszekvenálásra, génexpressziómérésre és kromatin-immunoprecipitációs szekvenálásra (ChIP-Seq) is alkalmazható. A DNS-könyvtárak tartalmát klonálisan amplifikálják gyöngyökre, majd PCR-reakcióval megsokszorozzák, és a gyöngyöket kovalens kötéssel egy tárgylemezhez kötik. A szekvenálást két ismert bázisból álló próbákkal végzik, így a 4 különböző fluoreszcens festékkel minden kötés első és második bázisa megismerhető. A komplementerpróbák hibridizálódnak a leolvasandó szekvenciához, majd végül a fluoreszcens jel mérése révén határozzák meg a szekvenciát (www.appliedbiosystems.com).

Ezen legelterjedtebb KGST-k már Magyarországon is megjelentek. Például Solid rendszer van a Bay Zoltán Alkalmazott Kutatási Közalapítványnál (www.baygen.hu), 454-es szekvenálás pedig a Semmelweis Egyetemen (www.usn.hu).

Második generációs szekvenálási technológiák

Az utóbbi 1-2 évben jelent meg a KGST-k második generációja, mint például a Helicos, Pacific Biosciences, Nanopore vagy a NABsys szekvenátorai. Ezek a cégek az amplifikációs lépést átugorva „single-molecule sequencing” technológiákat alkalmazva az egyes molekulákat határozzák meg.

A Helicos szekvenátora 100 millió leolvasást végez el kísérletenként direkt a DNS-en. Az egyes molekulákat „paired-ends” módszerrel [25] szekvenálja meg, eredményei kvantitatívak. A mintákhoz a szekvenálás előtt egy univerzális kötőveget illesztenek a fragmentek 5' végéhez és egy poliA farkat a 3' végéhez. Majd minden fragmentet a HeliScope Flow Cell felszínre rögzítés után a fluoreszcens jelölés alapján CCD-kamerával lokalizálnak. A meghatározás során minden ciklusban ismert fluoreszcensen jelzett nukleotidot kötnek a darabokhoz, így egyszerre több darab szekvenálását végzik párhuzamosan valós időben (www.helicosbio.com).

A Pacific Biosciences az SMRT (single molecule real time) DNS-szekvenálási technológiát fejlesztette ki, amely a „sequencing-by-synthesis” elvén alapul [26]. A szekvenáláshoz mindkét végén foszforilált csoportot hordozó nukleotidokat használnak, amelyek egyik végének foszforilált csoportja 4 nukleotidnak megfelelően különböző színű lehet. Ezeket a nukleotidokat építi be a DNS-polimeráz az éppen megkettőződő DNS-szál-

ban. Minden beépülést a színes foszforilált vég leválásával detektálható nukleotidspecifikus fényjelenség kíséri (www.pacificbiosciences.com).

A Helicos és a Pacific Biosciences második generációs szekvenátorai a sangeri leolvasási elven alapulnak. A KGST-hez képest a teljesítményt maximalizálták, míg a költségeket és a munkaidőt minimálisra szorították, azonban felmerül a kérdés, vajon a továbbiakban ezt a módszert lehet-e még tovább optimalizálni. A Nanopore (www.nanoporetech.com) és a NABsys (www.nabsys.com) nemrégiben bevezetett technológiai újítása, a nanopórus alkalmazása [27, 28] merőben eltér az eddigi módszerektől. Ennek az új szemléletű technológiai megoldásnak az alkalmazása a KGST teljesítményét nagyban felülmúlja, azonban kereskedelmi forgalomban még nem kaphatók ilyen technológiát alkalmazó szekvenátorok.

A különböző szekvenálási technológiák és gépek összehasonlításának megkönnyítéséhez az 1. táblázatban összefoglaltuk a különböző, jelenleg kapható gépek legfőbb tulajdonságait. A cégek által feltüntetett adatok felhasználásával látható a különböző termékek teljesítménye, átlagos alkalmazott mintahossz, valamint egy genom megszekvenálásának költsége (2010) az általuk forgalmazott kettek felhasználásával. A cégek által megadott szekvenálási teljesítmény sok esetben optimális és több futtatás esetén érhető csak el, így némi nehézséget jelent az objektív összehasonlításuk.

A piacon megjelentek más molekuláris biológiai műszereket és kettek forgalmazó cégek (Qiagen, Bio Scientific, Epicentre Biotechnologies), amelyek a KGST-szekvenátorokhoz kettek árulnak. (Ezeket azonban az összefoglaló táblázatunkban nem tüntettük fel.) Ezek a kettek akár több szekvenátorra is validáltak lehetnek (www.bioscientific.com; www.epibio.com), de egyes esetekben csak a szekvenálás egyes részlépéséhez (például DNS-könyvtár létrehozása) alkalmazhatóak (www.qiagen.com).

Bioinformatikai kiértékelés

KGST-k alkalmazása során kritikus tényező a bioinformatika kiértékelése. Elsőként az óriási adattömeg miatt: egy teljes nyers genom körülbelül 100 Gb-nyi adat, ami csak 24 darab DVD-n férne el. Ezt az óriási adattömeget kell feldolgozni, hogy megkapjuk magát a szekvenciát. A feldolgozás során egy referenciagenomhoz kell minden egyes darabot hasonlítani. (Alignment-nek nevezik a szakirodalomban, azonban ez nem más, mint egy óriási blastolás.) Rövidebb leolvasási hosszánál (ABI) ez körülbelül 2 hét, míg nem a teljesgenom-szekvenálás esetén vagy hosszabb leolvasási hossz mellett (Roche) ez már elkészülhet akár egy nap alatt is. Mindez a gépekkel együtt beszerzendő nagy teljesítményű cluster számítógépeken történő feldolgozás esetén igaz – technikailag asztali gépeken is fel lehet az adatokat dolgozni, de ez extrém időigényes lenne.

1. táblázat

Kereskedelmi forgalomban kapható szekvenátorok összehasonlítása.

A szekvenátorokat forgalmazó cégek és elérhetőségük, valamint az általuk forgalmazott termékek adatait hasonlítja össze a táblázat. A szekvenátorok teljesítményre, átlagos mintahosszra és az egygenomnyi DNS megszekvenálásának költségeire (ár) vonatkozó adatai (2010) a cégek hivatalos honlapján elérhető információk

Vállalat	Webcím	Termék	Teljesítmény	Átlagos mintahossz	Ár
454 Life Science Corp.	www.454.com	GS FLX Titanium Series	400–600 megabázis/futás	400 bázis	n. a.
		GS Junior	35 millióbázis/futás	400 bázis	n. a.
Illumina	www.illumina.com	HiSeq2000	150–200 gigabázis/futás	2×100 bázis	10 000 dollár
		Genome Analyzer IIx	85–95 gigabázis/futás	2×150 bázis	n. a.
		Genome Analyzer IIe	34–38 gigabázis/futás	2×150 bázis	n. a.
Applied Biosystems Inc.	www.appliedbiosystems.com	SOLiD™ 4 System	100 gigabázis/futás	50 bázis	6000 dollár
		SOLiD™ 4hq System	300 gigabázis/futás	75 bázis	3000 dollár
		SOLiD™ PI System	50 gigabázis/futás	75 bázis	8000 dollár
Helicos BioSciences Corp.	www.helicosbio.com	HeliScope™ Single Molecule Sequencer	420–700 megabázis/futás/csat.	25–55 bázis	n. a.
Pacific Biosciences	www.pacificbiosciences.com	PacBio RS	n. a.	75 000 molekula egyidejű olvasása	n. a.

n. a.: nem ismert adat; csat.: csatorna

Amikor megvan maga a szekvencia, akkor következhet a részletes kiértékelés, amely géneexpresszió mérését (transzkriptomszekvenálás esetén), SNP-k azonosítását, poszttranszkripcionális nukleotidvariációk detektálását, fúziós gének azonosítását, új fajok genomjának meghatározását, az RNS másodlagos szerkezetének meghatározását stb. tartalmazhatja. A bioinformatikai kiértékelésre már ma is számtalan program áll rendelkezésre, amelyek legrészletesebb összefoglalása elérhető a SeqWiki-n (<http://seqanswers.com/wiki/Special:BrowseData>). Itt jelenleg összesen 340 alkalmazás szerepel, amelyek döntő többsége Roche 454, Solid és az Illumina platformok alignmentjére koncentrál.

A KGST-k bioinformatikája azonban egy rendkívül gyorsan fejlődő terület, ahol a jelenleg használatban levő programok akár már egy év múlva is teljesen elavultak lehetnek. Általánosságban elmondható, hogy ha valaki kutatási projektben akar KGST-t alkalmazni, akkor a legtöbb időt a bioinformatikai kiértékelésre kell szánnia, amelyhez ráadásul megfelelően képzett bioinformatikust is be kell vonnia.

A KGST alkalmazásai

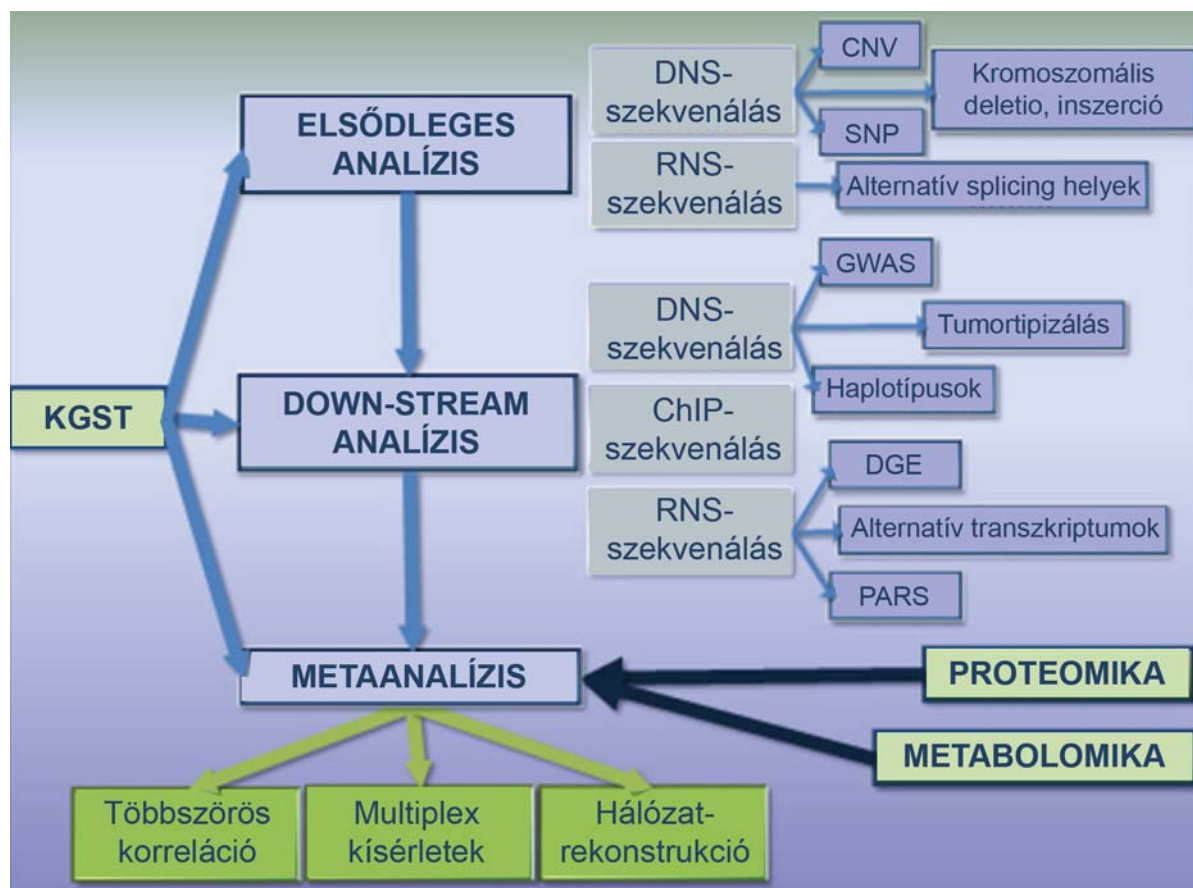
KGST alkalmazása a kutatásban

A nagy teljesítményű szekvenálás révén a funkcionális genomikában egy új fejezet kezdődött. Mivel az emberi genomok nem azonosak, így valójában annyi különböző humán genom létezik, ahány ember [29]. A genomon belüli eltérések (mint például az SNP-k) által létrehozott részleges változások a kódoló régióban vagy a génszabályozó régiókban eredményezhetnek funkcionális eltéréseket. Ezek vizsgálata a KGST-k megjelenése előtt

részleges és nagyon időigényes, valamint cseppet sem költségkímélő volt.

A KGST-k lehetőséget adnak egy elsődleges adatszerzésre nemcsak a genomikában, hanem az epigenetikában, transzkripció faktorok kötődéseinek vizsgálatában és a transzkriptomikában is. Azonban a genomikus szekvenciák ismerete önmagában nem elegendő a funkcionális genomikai kérdések megválaszolásához. A KGST-vel nyert adatokat microarray-s kísérletek eredményeivel kiegészítve egy down-stream analízis végezhető. Az így nyert információk [30, 31] szükségesek ahhoz, hogy a végső lépésben a platformokon átívelő metaanalízis során számos egyéb -omikus tudományág (proteomika, metabolomika) eredményeivel kiegészítve betekintést nyerhessünk a vizsgált szekvenciákban talált variációk okozta funkcionális eltérésekbe [32]. Így nyílik lehetőség a funkcionális genomika felvetette kérdések megválaszolására.

A KGST alkalmazásával nyert adatok elsődleges analízise során az eredmények direkt módon, a leolvasást követően a szekvenciák fizikális összehasonlítása révén nyerhetők ki. DNS-szekvenálás révén például kópiaszámvariációk (CNV), kromozómadeletio, inszercióvizsgálat, génpolimorfizmus- (SNP-) annotáció, de novo SNP-detektálás válik lehetővé. Eközben az RNS-szekvenálás révén alternatív splicinghelyek és ezek kapcsán új transzkriptumok ismerhetők meg. Az adatfeldolgozás következő lépéseként betekintést nyerhetünk a különböző kapcsolatokba és olyan funkcionális összefüggésekbe, mint például az ontológiák [33] vagy a szignáltranszdukciós útvonalak [34]. Azonban ehhez már down-stream analízis szükséges, amelyhez az adatokat a genomikus DNS-szekvenálással [mint például genome-wide association studies (GWAS), haplotípus-



2. ábra

A KGST segítségével nyerhető adatok feldolgozásának lehetőségei.

A KGST által nyerhető adatok feldolgozásának módszereit foglalja össze az ábra. Az elsődleges analízis során, amikor csak a leolvasott szekvenciák fizikális összehasonlítását végezzük, a DNS-szekvenálással kromoszómák inszerciójának és deletiójának, CNV-k és SNP-k vizsgálata végezhető. Az RNS-szekvenálás révén például alternatív splicinghelyek válhatnak ismertté. A down-stream analízis egy nagyobb áttekintést ad a DNS-, az RNS- és a ChIP-szekvenálással nyert információk által [30, 31]. Az előbbi kettő esetében konkrét példákkal élve a DNS-szekvenálás révén GWAS, haplotípusok definiálását és tumortipizálást, míg az RNS-szekvenálás esetén a DGE-mérés végezhető, az alternatív transzkriptumok válhatnak ismertté vagy a PARS révén az RNS másodlagos szerkezetének megismerésének segítségével meghatározhatók azoknak a szakaszoknak a hossza, amelyek kétszálúak, vagy a hárombázisonkénti periodicitás, és ezeknek a translációval való direkt összefüggése [35]. A KGST mellett más tudományágak (például proteomika vagy metabolomika) eredményeit felhasználja a végső metaanalízis. Ennek révén hálózati rekonstrukciót, többszörös korrelációkat vagy multiplex kísérleteket hozhatunk létre [32], melyek a genom funkcionális folyamataiba nyújtanak betekintést

CNV = (copy number variation) kópiaszám-variációk; ChIP-Seq = (chromatin immunoprecipitation sequencing) kromatin-immunoprecipitációs szekvenálás; DGE = (dynamic gene expression) dinamikus génexpresszió; GWAS = (genome-wide association studies) teljesgenom-asszociációs tanulmány; KGST = következő generációs szekvenálási technológiák; PARS = (parallel analysis of RNA structure) az RNS struktúrájának párhuzamos analízise; SNP = (single nucleotid polimorfizmus) génpolimorfizmus

definiálás vagy tumortipizálás eredményei], ChIP-Seq-val vagy RNS-szekvenálással [mint például dinamikus génexpresszió mérés (DGE), alternatív transzkriptumok megismerése vagy az RNS struktúrájának párhuzamos analízise (PARS)] [35] szerezhetjük meg. A down-stream analízis során kinyert adatok alapján azonosíthatók útvonalak, cisz-regulációs modulok és szabályozóhálózatok. Ezekből egy végső metaanalízis során a KGST segítségével nyert információk és más forrásokból (például proteomika, metabolomika) származó eredmények együttes vizsgálatának eredményeként (a hálózatok rekonstrukciója, különböző -omikák eredményeinek többszörös korrelációi, és a multiplex kísérletek összevetése) lehetővé válik a biológiai folyamatok teljesebb, funkcionális leírása [36]. A KGST által

nyert adatok analízisének többlépcsős stratégiáját a 2. ábra foglalja össze.

KGST alkalmazása a klinikumban

Bár a KGST-k rutinszerű klinikai alkalmazása ma még csak vízió, azonban bizonyos klinikai kérdések megválaszolásához már ma is kihasználják a genomszekvenálás adta pluszinformációkat. Indiai személyek Leber-féle öröklődő opticus neuropathiás betegek mitokondriális DNS-ét szekvenálták meg a klinikai kivizsgálás kiegészítéseként [37]. A baktériumok genomszekvenálása révén elérhető pontos baktériumidentifikálás hasznos kiegészítője lehet a klinikai epidemiológiai vizsgálatoknak. Az *Acinetobacter baumannii* esetében a kórokozó

terjedésének elemzése során használták a KGST-t [38]. A multigénés eredetű hypertrophiás cardiomyopathia esetében már ki is fejlesztettek egy 12 génből álló DNS-reszekvenáláson alapuló array-t. A nagy teljesítményű reszekvenáló array segítségével a hypertrophiás cardiomyopathiás betegek diagnosztikájában és prognosztikai beosztásához egy könnyebb és gyorsabb molekuláris tesztet hoztak létre [39]. A foetalis kromoszómaaneuploiditás anyai vérből való noninvazív vizsgálata is kivitelezhető szekvenálással [40, 41].

A tumoros betegségek diagnosztikájában egy egyre gyakrabban alkalmazott diagnosztikai alternatívát nyújthat a DNS-szekvenálás [42]. Már néhány éve a KGST segítségével végezhető reszekvenálások nyomán váltak ismertté olyan gének, amelyek mutációja humán tumorokban nagyon gyakran fordul elő [43, 44]. A malignus melanómák 60%-ában BRAF-mutációt [45], nem kis sejtes tüdőrákban és különböző adenocarcinómákban a növekedésifaktor-receptort (EGFR) kódoló gén mutációját [46], valamint a PIK3CA és az AKT1 gén mutációját különböző típusú tumorokban [47, 48] is leírták már. Egy német kutatócsoport 81 krónikus myeloid leukaemiás beteg anyagából KGST-vel szekvenálta meg a CBL, JAK2, NPL, N-RAS, K-RAS géneket a mutációs hot spot régiókban, valamint az RUNX1 és TET2 gének teljes kódoló régióját. Az összehasonlító molekuláris vizsgálat és a klinikai adatok összevetését követően 82 különböző mutációt találtak, amelyek közül például a TET2 mutációja jobb prognózissal párosult [49].

A genomszekvenálás során megismert információk, mint például a különböző betegségekből származó más-más típusú tumorok genomjainak meghatározásával feltérképezhető tumorspecifikus mutációk elősegíthetik a tumorok etiológiájának megfejtését [50]. A Nemzetközi Rák Genom Konzorcium (International Cancer Genome Consortium, ICGC) 50 különböző tumortípus alapján egy széles spektrumú átfogó tumorgenom-kutatást fog össze. Céljuk több mint 25 000 tumoros genom megszekvenálása révén egy, a genomikai abnormalitásokat összefoglaló katalógus létrehozása, amely elősegítheti a betegségek megértését, kezelését és megelőzését. Az eddigiekben összegyűjtött adatok egy része szabadon elérhető az interneten (www.icgc.org).

Kitekintés

Tanulmányunkban a szekvenálási technológiák kialakulását és elterjedését időrendben vettük végig, valamint részletesen is áttekintettük a KGST-eket alkalmazó szekvenátorokat. A szekvenálás adta lehetőségek szinte korlátlanok, hiszen epidemiológiai vizsgálatokban, betegségek diagnosztikájában vagy akár az evolúció kutatásában is fel lehet őket használni. Ezek közül tanulmányunkban röviden összefoglaltunk néhány, várhatóan klinikai szereppel is bíró eredményt.

Nem hagyhatjuk ki azonban annak a ténynek a megemlítését, hogy a technológia magas ára miatt a gyors elterjedése nem várható. Klinikai szempontból igen nehéz megbecsülni, hogy hány beteg genomját kell megszekvenálni ahhoz, hogy egy, a populációban 1%-os gyakorisággal jelen levő génmutáció prognosztikus hatását vizsgálni tudjuk.

A jövőre vonatkozó optimizmust az is beárnyékolja, hogy a KGST-k elsősorban exonszekvenálásra használhatóak: bár technikailag lehetséges lenne valóban a „teljes” genom szekvenálása, azonban a sebesség gyorsítására és a költségek csökkentése érdekében a legtöbb esetben csak az exonokat (tehát a „transzkriptom”-ot) szekvenálják meg.

Végül még meg kell említenünk a kiértékelés nehézségeit. Hiába lesz meg 50 ezer gén szekvenciája, ha azokban találunk 25 millió mutációt, amelyeknek csak 1%-áról lesz klinikai adat. (Már ez is 250 ezer génpolimorfizmus ismeretét feltételezi.) Hogyan fogunk eljutni a szekvenciáktól a valódi kérdésig, vagyis ahhoz, hogy megértsük a gének működését? Ezen kérdés megválaszolásában az újabb és újabb generációs szekvenálási technológiáknak kulcsszerepük lesz.

Irodalom

- [1] Lander, E. S., Linton, L. M., Birren, B. és mtsai: Initial sequencing and analysis of the human genome. *Nature*, 2001, 409, 860–921.
- [2] Venter, J. C., Adams, M. D., Myers, E. W. és mtsai: The sequence of the human genome. *Science*, 2001, 291, 1304–1351.
- [3] Sanger, F.: Nucleotide sequences in DNA. *Proc. R. Soc. Lond. Biol. Sci.*, 1975, 191, 317–333.
- [4] Smith, L., Sanders, J., Kaiser, R. és mtsai: Fluorescence detection in automated DNA sequence analysis. *Nature*, 1986, 321, 674–679.
- [5] Metzker, M. L., Raghavachari, R., Richards, S. és mtsai: Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates. *Nucleic Acids Res.*, 1994, 22, 4259–4267.
- [6] Ruiz-Martinez, M. C., Berka, J., Belenki, A. és mtsai: DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem.*, 1993, 65, 2851–2858.
- [7] Ju, J., Ruan, C., Fuller, C. és mtsai: Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.*, 1995, 92, 4347–4351.
- [8] Lee, L., Spurgeon, S., Heiner, C. és mtsai: New energy transfer dyes for DNA sequencing. *Nucleic Acids Res.*, 1997, 25, 2816–2822.
- [9] Kheterpal, I., Scherer, J., Clark, S. és mtsai: DNA sequencing using a four-color confocal fluorescence capillary array scanner. *Electrophoresis*, 1996, 17, 1852–1859.
- [10] Ewing, B., Green, P.: Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.*, 1998, 8, 186–194.
- [11] Ewing, B., Hillier, L., Wendl, M. C. és mtsai: Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.*, 1998, 8, 175–185.
- [12] Schmalzing, D., Tsao, N., Koutny, L. és mtsai: Toward real-world sequencing by microdevice electrophoresis. *Genome Res.*, 1999, 9, 853–858.

- [13] Liu, S., Shi, Y., Ja, W. és mtsai: Optimization of high-speed DNA sequencing on microfabricated capillary electrophoresis channels. *Anal. Chem.*, 1999, 71, 566–573.
- [14] Liu, S., Ren, H., Gao, Q. és mtsai: Automated parallel DNA sequencing on multiple channel microchips. *Proc. Natl. Acad. Sci.*, 2000, 97, 5369–5374.
- [15] Simpson, J. W., Ruiz-Martinez, M. C., Mulhern, G. T. és mtsai: Transmission imaging spectrograph and microfabricated channel system for DNA analysis. *Electrophoresis*, 2000, 21, 135–149.
- [16] Backhouse, C., Caamano, M., Oaks, F. és mtsai: DNA sequencing in a monolithic microchannel device. *Electrophoresis*, 2000, 21, 150–156.
- [17] Koutny, L., Schmalzing, D., Salas-Solano, O. és mtsai: Eight hundred-base sequencing in a microfabricated electrophoretic device. *Anal. Chem.*, 2000, 72, 3388–3391.
- [18] Lewis, E. K., Haaland, W. C., Nguyen, F. és mtsai: Color-blind fluorescence detection for four-color DNA sequencing. *Proc. Natl. Acad. Sci.*, 2005, 102, 5346–5351.
- [19] Emrich, C. A., Tian, H., Medintz, I. L. és mtsai: Microfabricated 384-lane capillary array electrophoresis bioanalyzer for ultra-high-throughput genetic analysis. *Anal. Chem.*, 2002, 74, 5076–5083.
- [20] Pettersson, E., Lundberg, J., Ahmadian, A.: Generations of sequencing technologies. *Genomics*, 2009, 93, 105–111.
- [21] Mardis, E. R.: The impact of next-generation sequencing technology on genetics. *Trends Genet.*, 2008, 24, 133–141.
- [22] Margulies, M., Egholm, M., Altman, W. E. és mtsai: Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437, 376–380.
- [23] Ronaghi, M., Uhlen, M., Nyren, P.: A sequencing method based on real-time pyrophosphate. *Science*, 1998, 281, 363–365.
- [24] Korbel, J. O., Urban, A. E., Affourtit, J. P. és mtsai: Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 2007, 318, 420–426.
- [25] Harris, T. D., Buzby, P. R., Babcock, H. és mtsai: Single-molecule DNA sequencing of a viral genome. *Science*, 2008, 320, 106–109.
- [26] Eid, J., Febr, A., Gray, J. és mtsai: Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, 323, 133–138.
- [27] Fologea, D., Gershow, M., Ledden, B. és mtsai: Detecting single stranded DNA with a solid state nanopore. *Nano Lett.*, 2005, 10, 1905–1909.
- [28] Stoddart, D., Maglia, G., Mikhailova, E. és mtsai: Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angew. Chem. Int. Ed. Engl.*, 2010, 49, 556–559.
- [29] Varki, A., Geschwind, D. H., Eichler, E. E.: Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nat. Rev. Genet.*, 2008, 9, 749–763.
- [30] Beyer, A., Bandyopadhyay, S., Ideker, T.: Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.*, 2007, 8, 699–710.
- [31] Hu, Z., Killion, P. J., Iyer, V. R.: Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.*, 2007, 39, 683–687.
- [32] Werner, T.: Regulatory networks: linking microarray data to systems biology. *Mech. Ageing Dev.*, 2007, 128, 168–172.
- [33] Thomas, P. D., Mi, H., Lewis, S.: Ontology annotation: mapping genomic regions to biological function. *Curr. Opin. Chem. Biol.*, 2007, 11, 4–11.
- [34] Werner, T.: Bioinformatics applications for pathway analysis of microarray data. *Curr. Opin. Biotechnol.*, 2008, 19, 50–54.
- [35] Kertesz, M., Wan, Y., Mazor, E. és mtsai: Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 2010, 467, 103–107.
- [36] Werner, T.: Next generation sequencing in functional genomics. *Brief Bioinform.*, 2010, 5, 449–511.
- [37] Kumar, M., Tanwar, M., Saxena, R. és mtsai: Identification of novel mitochondrial mutations in Leber's hereditary optic neuropathy. *Mol. Vis.*, 2010, 16, 782–792.
- [38] Lewis, T., Loman, N. J., Bingle, L. és mtsai: High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *J. Hosp. Infect.*, 2010, 75, 37–41.
- [39] Fokstuen, S., Lyle, R., Munoz, A. és mtsai: A DNA resequencing array for pathogenic mutation detection in hypertrophic cardiomyopathy. *Hum. Mutat.*, 2008, 29, 879–885.
- [40] Fan, H. C., Blumenfeld, Y. J., Chitkara, U. és mtsai: Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl. Acad. Sci. USA*, 2008, 105, 16266–16271.
- [41] Chiu, R. W., Chan, K. C., Gao, Y. és mtsai: Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc. Natl. Acad. Sci. USA*, 2008, 105, 20458–20463.
- [42] Papadopoulos, N., Kinzler, K. W., Vogelstein, B.: The role of companion diagnostics in the development and use of mutation-targeted cancer therapies. *Nat. Biotechnol.*, 2006, 8, 985–995.
- [43] Ding, L., Getz, G., Wheeler, D. A. és mtsai: Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, 2008, 455, 1069–1075.
- [44] Greenman, C., Stephens, P., Smith, R. és mtsai: Patterns of somatic mutation in human cancer genomes. *Nature*, 2007, 446, 153–158.
- [45] Davies, H., Bignell, G. R., Cox, C. és mtsai: Mutations of the BRAF gene in human cancer. *Nature*, 2002, 417, 949–954.
- [46] Sharma, S. V., Bell, D. W., Settleman, J. és mtsai: Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer*, 2007, 7, 169–181.
- [47] Carpten, J. D., Faber, A. L., Horn, C. és mtsai: A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature*, 2007, 448, 439–444.
- [48] Samuels, Y., Wang, Z., Bardelli, A. és mtsai: High frequency of mutations of the PIK3CA gene in human cancers. *Science*, 2004, 304, 554.
- [49] Kohlmann, A., Grossmann, V., Klein, H. U. és mtsai: Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. *J. Clin. Oncol.*, 2010, 28, 3858–3865.
- [50] Pfeifer, G. P., Besaratinia, A.: Mutational spectra of human cancer. *Hum. Genet.*, 2009, 125, 493–506.

(Mihály Zsuzsanna,
Budapest, Bókay J. u. 53., 1083
e-mail: zsmihi@msn.com)