



SECUENCIACIÓN DE GENOMAS ANTIGUOS Y MODERNOS

Gabriel Dorado¹, Víctor Vásquez², Isabel Rey³, Fernando Luque⁴, Inmaculada Jiménez⁵, Arturo Morales⁶, Manuel Gálvez⁷, Jesús Sáiz⁸, Adela Sánchez⁸, Pilar Hernández⁹

¹Autor para correspondencia, Dep. Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Universidad de Córdoba, 14071 Córdoba (Spain), CE: <bb1dopeg@uco.es>; ²Centro de Investigaciones Arqueobiológicas y Paleoeológicas Andinas, ARQUEOBIOS, Apartado Postal 595, Trujillo (Perú); ³Colección de Tejidos y ADN, Museo Natural de Ciencias Naturales, 28006 Madrid; ⁴Servicio Sanidad Exterior, Dependencia de Sanidad, Subdelegación del Gobierno en Huelva, C/. Sanlúcar de Barrameda 7, 21001 Huelva; ⁵I.E.S. Abyla, Polígono Virgen de África, s/n, 51001 Ceuta; ⁶Dep Biología, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Cantoblanco (Madrid); ⁷Dep. Radiología y Medicina Física, Unidad de Física Médica, Facultad de Medicina, Avda. Menéndez Pidal s/n, Universidad de Córdoba, 14071 Córdoba; ⁸Dep. Farmacología, Toxicología y Medicina Legal y Forense, Facultad de Medicina, Avda. Menéndez Pidal, s/n, Universidad de Córdoba, 14071 Córdoba; ⁹Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Alameda del Obispo s/n, 14080 Córdoba

La primera generación de técnicas para efectuar la secuenciación del ADN empezó en 1975 con la metodología de Sanger y Coulson, la "más y menos" (del inglés, "plus and minus"), la cual necesitaba clonar cada lectura inicial para producir un ADN de cadena simple. En 1977, Maxam y Gilbert publicaron la metodología de secuenciación de ADN mediante degradación química. Este método estaba basado en la modificación química y posterior rotura del ADN, y empezó a ser el método de secuenciación más utilizado porque permitía utilizar un ADN purificado sin necesidad de clonarlo. El mismo año Sanger publicó el método de secuenciación de ADN por síntesis química, que estableció un nuevo estándar para los próximos 30 años (Sanger et al, 1977, 1992; Wikipedia, 2008a).

El método de Sanger permitía leer 25 bases (b) y más tarde, 80 b, utilizando los terminadores dideoxi. El método fue

optimizado con la utilización de dideoxinucleótidos fluorescentes en vez de productos tóxicos y radioisótopos, detección automatizada, mayor rendimiento y precisión, permitiendo leer 1.000 b, a lo que hemos contribuido significativamente (Lario et al, 1997). Estos avances representaron una revolución fascinante, porque permitieron descifrar inicialmente genes y eventualmente los genomas completos (Schuster, 2008), aunque con un alto costo para el segundo caso.

Actualmente el método de Sanger es todavía utilizado en la mayoría de los laboratorios del mundo. Sin embargo, esta tecnología no es muy adecuada para las muestras arqueológicas, porque esta obstaculizada por la escasa cantidad y degradación del ADN. Esta es la razón por lo cual la primera generación de secuenciación de ADN no funciona bien con genomas del ADN antiguo (ADNa; del inglés, "aDNA"). Sin embargo, este escenario ha cambiado

con el desarrollo de la segunda generación de plataformas de secuenciación.

La segunda generación de plataformas de secuenciación de ADN (también llamada de siguiente generación; del inglés, "next-generation") fue desarrollada hace tres años, aumentando la efectividad de la secuenciación del ADN algunos ordenes de magnitud (Bonetta, 2006). Así, permitió generar lecturas de gigabases en una solo experimento (del inglés, "run"). Cuatro plataformas de segunda generación han sido comercializadas por el momento:

1) El instrumento 454 (454 Life Sciences), está basado en emulsión, secuenciación por síntesis (SBS; del inglés, "sequencing-by-synthesis") y pirosecuenciación. Este desarrollo fue publicado en 2005 (Margulies, 2005), comprado por Roche Diagnostics en 2007 y vendido como el "Genome Sequencer 20 System" y el "Genome Sequencer FLX System" (Roche Applied Sciences) <<https://www.roche-applied-science.com/sis/sequencing/index.jsp>>. La tecnología 454 empezó leyendo 100 b, después de 16 meses podía leer 250 b, y ahora más de 400 b (Schuster, 2008).

2) El protocolo de secuenciación de 'polonias' multiplexadas (del inglés, "multiplex polony sequencing protocol") es parecido al método mencionado anteriormente, pero es mas barato porque usa instrumentos y reactivos estándar. Las bibliotecas genómicas obtenidas por la técnica de perdigonazos (del inglés, "shotgun genomic libraries") son amplificadas en microesferas mediante PCR por emulsión. Después, son utilizadas como sustratos para hacer la secuenciación con las reacciones fluorescentes de ligación nonamerica sobre un portaobjetos de microscopio, generando

millones de lecturas de 26 pares de bases (pb) (Porreca et al, 2008), de modo que cualquier laboratorio pueden desarrollar este método.

3) El "Genome Analyzer System" (Solexa) combina la química SBS con terminadores y tecnología de grupos (del inglés, "cluster"). La compañía fue adquirida por Illumina en 2007, produciendo el "Genome Analyzer Sequencing System" <<http://www.illumina.com/pages.ilmn?ID=204>>. Este tipo de tecnología genera diez veces más lecturas que la 454, pero con solamente 35 b o menos en longitud.

El "SOLiD System" (Applied Biosystems) usa una química basada en ligasa (Chi, 2008) y fue producido en 2007 <<http://solid.appliedbiosystems.com>>.

La segunda generación de plataformas de secuenciación de ADN difiere de los métodos tradicionales de secuenciación en dos aspectos. Primero, en vez de hacer una secuenciación de clones de ADN de algunos individuos (p. ej., 96 secuenciaciones de sustratos en un secuenciador capilar Sanger), cientos de miles (sistema 454) o miles de millones (Solexa y SOLiD) de moléculas de ADN son secuenciadas en paralelo, usando volúmenes de reacción menores (Schuster, 2008). Segundo, la secuencias obtenidas son generalmente mucho mas cortas (25-50 nucleótidos para las tecnologías de 'polonias', Solexa y SOLiD, aunque pueden alcanzar 200-400 nucleótidos para el sistema 454) que las generados por secuenciación tradicional (Gravely, 2008). No obstante, el costo de los nuevos instrumentos es mayor (aproximadamente, unos 500.000 dólares) que los que usan el método de Sanger (de 10.000 a 100.000 dólares), que también pueden realizarse con instrumentación manual mas barata usando radioisótopos o fluoróforos.

Como ejemplo práctico de estos avances, la primera secuenciación del genoma humano (*Homo sapiens sapiens*) ha requerido cientos de máquinas trabajando 24 horas al día, durante 13 años, con un costo de más de 300 millones de dólares. El proyecto empezó en 1990, generando un borrador de trabajo del genoma en el año 2000, y uno más completo en 2003, con más análisis que todavía se están publicando (HGP, 2008; HUGO, 2008; Wikipedia, 2008b). Dicho logro ha sido de hecho comparado en tiempo y costo al del proyecto Apolo, el cual logro poner un hombre sobre la luna. Más tarde, el genoma diploide de una sola persona (J. Craig Venter) fue leído mediante secuenciación de genomas completos por perdigonazos (del inglés, "whole-genome shotgun sequencing"), necesitando 10 años y 70 millones de dólares, usando la tecnología optimizada de Sanger (Levy et al, 2007). Por su parte, el genoma de Watson fue secuenciado en sólo dos meses y un costo de un millón de dólares, usando la máquina "454 Life Sciences" (Chi, 2008; Wheeler et al, 2008). Otro ejemplo es la secuenciación del genoma del "ornitorrinco" (*Ornithorhynchus anatinus*), revelando marcas únicas de su evolución, con genes que aparecen en reptiles, o aves y otros de mamíferos. Esta mezcla fascinante de características en el genoma del "ornitorrinco" proporciona pistas sobre el rol y la evolución de los genomas de los mamíferos (Warren et al, 2008).

Dado que el método de Sanger tiene un precio prohibitivo para los proyectos de ADN nuclear (nuDNA), muchos laboratorios sólo utilizan ahora la segunda generación de secuenciadores, combinando los beneficios de las lecturas relativamente largas del sistema 454 con un bajo costo de operación del sistema Solexa o SOLiD, o aplicando el protocolo de 'colonias'.

La tercera generación (también llamada "next-next-generation") de la secuenciación de ADN ha sido producida este año, con químicas revolucionarias de una sola molécula:

1) El secuenciador HeliScope de molécula única (del inglés, "HeliScope Single Molecule Sequencer"), de Helicos BioSciences <<http://www.helicosbio.com>>, fue anunciado este año. Ofrece lecturas muy precisas de 25 a 45 bases para miles de millones (millardos) de cadenas en un solo experimento (produciendo más que 2 Gb de datos de secuenciación por día), y hasta un millardo de bases por hora en el futuro <http://www.helicosbio.com/Portals/0/Videos/tSM_S-How_It_Works.flv> Ello es debido al uso de la verdadera secuenciación de molécula única (del inglés, "true Single Molecule Sequencing" (tSMS), para leer hebras individuales de ADN (Blow, 2008; Harris, 2008).

2) "VisiGen Biotechnologies" <<http://visigenbio.com>> no ha sido producido todavía, pero promete micromatrices masivamente paralelas (del inglés, "microarrays") de nanomáquinas, con una tasa de secuenciación de un Mb/s/máquina (más de 86 Gb de secuencia de datos por día) <http://visigenbio.com/flash/stream/visigen_movie_6mb.swf>, leyendo también moléculas simples.

Estos avances permitirán reducir el precio de la secuenciación de uno a dos órdenes de magnitud, lo cual propiciará el desarrollo del "la genómica personal": hacer la secuenciación de todo el genoma humano de cualquier persona en menos de un día, por 1.000 dólares o menos (Mardis, 2006; Milos, 2008; VonBubnoff, 2008; Schuster, 2008). La reducción del precio de la secuenciación

y la mejora de los microprocesadores de los ordenadores actuales y las herramientas bioinformáticas pueden ser explotados para evitar el posible embotellamiento analítico que tantos datos podrían producir (Díaz et al, 2008a,b; Schuster, 2008). De modo que se espera que, para finales de este año, se podrá completar la secuenciación del genoma de más de 1.000 bacterias y arqueobacterias, y 100 eucariotas (Liolios et al, 2008), y para un número aún más grande de orgánulos, virus, plásmidos, viroides y virusoides.

No debe sorprender que todas estas tecnologías estén teniendo un impacto importante en la arqueología y otras ciencias relacionadas, permitiendo secuenciar más de 25.000 pb de ADNnu de un hueso de oso (*Ursus spelaeus*) de una cueva del Pleistoceno de 40.000 años de antigüedad (Noonan et al, 2005), y 13 millones de pb (Mpb) de ADNnu de un hueso de mamut (*Mammuthus primigenius*) con 38.000 años de edad (Green et al, 2006; Noonan et al, 2006). Sin embargo, esto no se realiza fácilmente. Primero, la cantidad de ADN normalmente obtenida a partir de muestras antiguas como el Neandertal es menos de 5%, comparado con muestras modernas. En otras palabras, los proyectos de genoma de ADNn necesitan 20 veces más de secuenciación que un proyecto de ADN moderno. Segundo, el ADNn está normalmente roto y químicamente alterado (dañado). Por último, pero no menos importante, muchas veces, el ADNn está contaminado con ADN moderno, lo cual es particularmente relevante para el ADN humano. Para superar los obstáculos de los proyectos genómicos de ADNn, puede requerir la secuenciación múltiple de una región determinada, re-secuenciación, así como el uso de las nuevas plataformas de PCR en emulsión (emPCR) (Blow et al, 2008). La recompensa es apasionante, porque permite la

secuenciación de genomas, y por tanto el análisis a nivel molecular de la flora y fauna de los últimos 100.000 años (Schuster, 2008).

Por otro lado, aparte de los ensamblajes de genomas de novo, es ahora económicamente factible la re-secuenciación de genomas de diferentes muestras, organismos o individuos, utilizando una secuencia de referencia previa para guiar el nuevo ensamblaje. Esta estrategia requiere mucha menos cobertura (de 8 hasta 12 veces), que el ensamblaje de genomas de novo (de 25 hasta 70 veces) y fue utilizado para hacer la secuenciación de genomas mitocondriales de pelos humanos antiguos (Gilbert, 2007, 2008), ofreciendo así la posibilidad de hacer estudios de poblaciones (Schuster, 2008). El ADN mitocondrial (ADNmt) ofrece ventajas comparado con el ADNnu para los estudios del ADNn: hay miles de genomas mitocondriales por célula, indica la herencia materna y tiene un ritmo de mutación acelerado.

Por otra parte, la tercera generación de métodos de secuenciación es tan poderosa que permite hacer estudios no solamente de genómica estructural, sino también de genómica funcional y consenso de secuencias (Wold y Myers, 2008), incluyendo: i) ChIP-Seq, que está basado en la inmunoprecipitación de cromatina (ChIP), para mapear in vivo las secuencias del ADN ocupadas por proteínas de unión al ADN; ii) Sec-ARNm (del inglés, "mRNA-Seq"), para estudiar la expresión de genes (Graveley, 2008); y iii) Sec-Metil (del inglés, "Methyl-Seq"), para analizar los patrones de metilación. Estos procedimientos se pueden aplicar también al ADN antiguo, siempre que ADN, ADN-proteína o ARNm pueda ser aislado de tales muestras.

En resumen, todo esto demuestra que no estamos en la era post-genómica

como algunos habían pensado, sino empezando la era genómica. No solamente por las nuevas y revolucionarias tecnologías, sino también porque sólo un minúsculo número de genomas han sido secuenciados por el momento, cuando se compara con la enorme cantidad de entidades biológicas que ocupan el planeta: de los virusoides, viroides y virus hasta procariontas y eucariontas. De modo que se deben esperar sorprendentes noticias en la actual era genómica de secuenciación de ácidos nucleicos, porque la carrera ha empezado, de nuevo, y ha venido para quedarse durante mucho, mucho, tiempo (Dorado, 2008).

Agradecimientos:

Financiado por proyectos AGL2006-12550-C02-01 y AGL2006-12550-C02-02 del Ministerio de Educación y Ciencia, Proyecto 041/C/2007 de la Consejería de Agricultura y Pesca de la Junta de Andalucía, y Grupo PAI AGR-248 de la Junta de Andalucía (España).

Referencias Bibliográficas

- Blow N (2008): DNA sequencing: generation next-next. *Nature Methods* 5: 267-272.
- Bonetta L (2006): Genome sequencing in the fast lane. *Nature Methods* 3: 141-146.
- Chi KR (2008): The year of sequencing. *Nature Methods* 5: 11-14.
- Díaz D, Claros MG, Falgueras J, Hernández P, Caballero JA, Dorado G, Gálvez S (2008a): DemAlign/Omega-Jalview: an algorithm/viewer for fast discovering differences in similar sequences. *Accelrys Science & Technology Forum* (Paris, France; <<http://accelrys.com/events/seminars/scienceand-technology-forums>>)
- Díaz D, Falgueras J, Claros MG, Guerrero DD, Hernández P, Caballero JA, Dorado G, Gálvez S (2008b): Integrating bioinformatics workflow tools: Omega-Brigid and Scitegic Pipeline Pilot. *Accelrys Science & Technology Forum* (Paris, France; <<http://accelrys.com/events/seminars/scienceand-technology-forums>>)
- Dorado G (ed) (2008): "Molecular Markers, PCR, Bioinformatics and Ancient DNA – Technology and Applications". *Science Publishers* (New York, NY, USA). In press.
- Gilbert MT, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF, Rasmussen M, Metspalu M, Higham TF, Schwenninger JL, Nathan R, De Hoog CJ, Koch A, Møller LN, Andreasen C, Meldgaard M, Villems R, Bendixen C, Willerslev E (2008): Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320: 1787-1789.
- Gilbert MT, Tomsho LP, Rendulic S, Packard M, Drautz DI, Sher A, Tikhonov A, Dalén L, Kuznetsova T, Kosintsev P, Campos PF, Higham T, Collins MJ, Wilson AS, Shidlovskiy F, Buigues B, Ericson PG, Germonpré M, Götherström A, Jacumin P, Nikolaev V, Nowak-Kemp M, Willerslev E, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC (2007): Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317: 1927-1930.
- Graveley BR (2008): Power sequencing. *Nature* 453: 1197-1198.
- Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Pääbo S (2006): Analysis of one million base pairs of Neanderthal DNA. *Nature* 444: 275-276.
- Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J,
- Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer

- E, Tyurina A, Ward R, Weiss H, Xie Z (2008): Single-molecule DNA sequencing of a viral genome. *Science* 320: 106-109.
- HGP (2008): Human Genome Project. Web: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml and http://www.ornl.gov/sci/techresources/Human_Genome/project/progress.shtml.
- HUGO (2008): Human Genome Organisation. Web: <http://www.hugo-international.org>.
- Lario A, González A, Dorado G (1997): Automated laser-induced fluorescence DNA sequencing: equalizing signal-to-noise ratios significantly enhances overall performance. *Analytical Biochemistry* 247: 30-33.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC (2007): The diploid genome sequence of an individual human. *PLoS Biol* 5: e254.
- Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC (2008): The Genomes On Line
- Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 36(Database issue): D475-D479.
- Mardis ER (2006): Anticipating the 1,000 dollar genome. *Genome Biol* 7: 112.1-112.5.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005): Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
- Milos P (2008): Helicos BioSciences. *Pharmacogenomics* 9: 477-480.
- Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM (2006): Sequencing and analysis of Neanderthal genomic DNA. *Science* 314: 1113-1118.
- Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Dettler JC, Pääbo S, Rubin EM (2005): Genomic sequencing of Pleistocene cave bears. *Science* 309: 597-599.
- Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC (2006): Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311: 392-394.
- Porreca GJ, Shendure J, Church GM (2006): Polony DNA sequencing. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds): "Current Protocols in Molecular Biology". Vols 1-4, Chapter 7: Unit 7.8. Greene & John Wiley (New York).
- Sanger F, Nicklen S, Coulson AR (1977): DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467.

- Sanger F, Nicklen S, Coulson AR (1992): DNA sequencing with chain-terminating inhibitors. *Biotechnology* 24: 104-108.
- Schuster SC (2008): Next-generation sequencing transforms today's biology. *Nature Methods* 5:16-18.
- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005): Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309: 1728-1732.
- VonBubnoff A (2008): Next-generation sequencing: the race is on. *Cell* 132 :721-723.
- Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, López-Otín C, Ordóñez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, Papenfuss AT, Wakefield MJ, Olender T, Lancet D, Huttley GA, Smit AF, Pask A, Temple-Smith P, Batzer MA, Walker JA, Konkel MK, Harris RS, Whittington CM, Wong ES, Gemmell NJ, Buschiazio E, Vargas Jentzsch IM, Merkel A, Schmitz J, Zemann A, Churakov G, Kriegs JO, Brosius J, Murchison EP, Sachidanandam R, Smith C, Hannon GJ, Tsend-Ayush E, McMillan D, Attenborough R, Rens W, Ferguson-Smith M, Lefèvre CM, Sharp JA, Nicholas KR, Ray DA, Kube M, Reinhardt R, Pringle TH, Taylor J, Jones RC, Nixon B, Dacheux JL, Niwa H, Sekita Y, Huang X, Stark A, Kheradpour P, Kellis M, Flicek P, Chen Y, Webber C, Hardison R, Nelson J, Hallsworth-Pepin K, Delehaunty K, Markovic C, Minx P, Feng Y, Kremitzki C, Mitreva M, Glasscock J, Wylie T, Wohldmann P, Thiru P, Nhan MN, Pohl CS, Smith SM, Hou S, Renfree MB, Mardis ER, Wilson RK (2008): Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175-183.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM (2008): The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872-876.
- Wikipedia(2008a): DNAsequencing. Web: http://en.wikipedia.org/wiki/DNA_sequencing.
- Wikipedia (2008b): Human Genome Project. Web: http://en.wikipedia.org/wiki/Human_Genome_Project.
- Wold B, Myers RM (2008): Sequence census methods for functional genomics. *Nature Methods* 5: 19-21.