
Einleitende Referate

Keynote presentation

Neue molekularbiologische und bioinformatische Methoden in der Unkrautforschung

New molecular biology and bioinformatics methods in weed research

Antje Krause

Fachhochschule Bingen, Berlinstr. 109, D-55411 Bingen am Rhein
a.krause@fh-bingen.de



DOI 10.5073/jka.2014.443.001

Zusammenfassung

Der zunehmende Einsatz kostengünstiger und zeitsparender Hochdurchsatz-Sequenzieretechniken (NGS, next-generation sequencing) ermöglicht die Bearbeitung völlig neuer Fragestellungen sowohl in der Medizin als auch in den Agrarwissenschaften. Die Verfügbarkeit von immer mehr genetischen Daten verschiedener Spezies und der Einsatz neuer molekularbiologischer und bioinformatischer Methoden zu ihrer Verarbeitung verschiebt damit den Blick von einzelnen Genen und Genprodukten hin zum Verständnis von Reaktionen, Interaktionen und ganzen biologischen Systemen.

Damit sind auch ihre Einsatzmöglichkeiten in der Unkrautforschung vielfältig. Untersuchungen des Genoms und Transkriptom eines Unkrauts können z.B. bei der Identifikation von Genprodukten, die als Zielorte (target-sites) für Herbizide geeignet sind, eingesetzt werden. Genexpressionsanalysen tragen zu einem besseren Verständnis von Herbizidresistenzen (sowohl Zielort- als auch Nicht-Zielortresistenzen) bei. Ferner kann die Sequenzierung des Metagenoms Aufschlüsse über mögliche Interaktionen mit Pflanzen-Mikroorganismen am untersuchten Standort geben.

Stichwörter: Biologische Datenbanken, Bioinformatik, Epigenetik, Exom, Expressionsdatenanalyse, Genom, Hochdurchsatz-Sequenzierung, Transkriptom

Abstract

The application of time and cost efficient high-throughput sequencing technologies called NGS, next-generation sequencing enables researchers to tackle completely new problems in medicine as well as in agriculture. The availability of increasing genetics data from various species along with the use of new molecular biology and bioinformatics methods for data handling shifts the focus from single genes or gene products to the understanding of reactions, interactions and entire biological systems.

In weed research the application of these methods is straightforward. The analysis of the genome and transcriptome of a weed can, for instance, lead to the identification of suitable gene products as target-sites for herbicides. Gene expression analyses result in a deeper understanding of herbicide resistance (target-site as well as non target-site resistance). In addition, sequencing the metagenome provides insights into potential interactions with microorganisms in the habitat of interest.

Keywords: Biological databases, bioinformatics, epigenetics, exome, expression analysis, genome, next-generation sequencing, transcriptome

Einleitung

Seit der kompletten Genom-Sequenzierung des pflanzlichen Modellorganismus *Arabidopsis thaliana* im Jahr 2000 (THE ARABIDOPSIS GENOME INITIATIVE, 2000) und der ersten Nutzpflanze *Oryza sativa* im Jahr 2002 (GOFF *et al.*, 2002) gewinnt die Analyse von genetischen Informationen zunehmend auch in der Pflanzenforschung an Bedeutung. Eine ganze Reihe neuer molekularbiologischer und bioinformatischer Methoden hat diese Entwicklung erst möglich gemacht.

Unter dem Begriff „Next Generation Sequencing“ (NGS) werden insbesondere neue Hochdurchsatz-Sequenzieretechniken zusammengefasst, die es aufgrund einer Parallelisierung bzw. Miniaturisierung der Sequenzierung ermöglichen, DNA mit einem sehr guten Zeit-Kosten-Verhältnis zu sequenzieren. So lagen die Kosten zur Sequenzierung von 1.000.000 Basenpaaren (1 Mbp) im April 2013 bei 0,06\$ im Vergleich zu über 5.000\$ im September 2001 (WETTERSTRAND, 2013). Hierbei wird unterschieden zwischen Sequenziergeräten, die eine Amplifizierung (d.h. Vervielfältigung, üblicherweise mit PCR (Polymerase Chain Reaction)) der zu sequenzierenden DNA voraussetzen (Amplified Single Molecule Sequencing, Second oder Next Generation Sequencing) und solchen, die einzelne DNA-Stränge ohne vorherige Amplifizierung sequenzieren (Single Molecule Sequencing, Third oder Next Next Generation Sequencing). Eine Übersicht über die aktuellen Technologien findet sich in Tabelle 1. Mit diesen Sequenziergeräten können in sehr kurzer Zeit (wenige Stunden) sehr große Mengen (mehrere Terabyte) an DNA- (Genom) bzw. mRNA-Daten (Transkriptom) für weitere Analysen in digitaler Form in den Computer gebracht werden. Diese Rohdaten werden dann dort mit Methoden der Bioinformatik weiter verarbeitet.

Tab. 1 Hochdurchsatz-Sequenzieretechniken; NGS: Next Generation Sequencing, Second Generation Sequencing, Amplified Single Molecule Sequencing; NNGS: Next Next Generation Sequencing, Third Generation Sequencing, Single Molecule Sequencing.

Tab. 1 High-throughput sequencing technologies.

Hersteller	Webseite	Gene-ration	Gerät(e)	Readlänge (in bp)	Reads pro Lauf
454 Sequencing, Roche	http://www.454.com/	NGS	GS Junior GS FLX+	400 600-1.000	100.000 1 Mio.
Illumina®	http://www.illumina.com/	NGS	HiSeq 2500 GAIIx	36-150 35-150	3 Mrd. 320 Mio.
Applied Biosystems®, Life Technologies	http://www.lifetechnologies.com/	NGS	5500xl SOLiD	75	3 Mio.
Ion Torrent™, Life Technologies	http://www.lifetechnologies.com/	NGS	Ion PGM™ Ion Proton™	35-400 200	2-3 Mio. 60-80 Mio.
Pacific Biosciences®	http://www.pacificbiosciences.com/	NNGS	RS II	Max. >20.000	70.000
Oxford Nanopore Technologies	https://www.nanoporetech.com/ Geräte noch nicht auf dem Markt	NNGS	GridION™ MinION™	48.000? 	2.000?

Da die meisten NGS-Sequenziergeräte nur sehr kurze Sequenzfragmente (Reads) produzieren und diese fehlerbehaftet sein können, ist die erste Herausforderung das Zusammenfügen dieser Fragmente zu größeren Sequenzabschnitten. Jede Position der zu rekonstruierenden Sequenz wird dabei idealerweise von mehreren Reads abgedeckt (Coverage), sodass die Reads allein aufgrund ihrer Ähnlichkeiten, d.h. ihres Überlapps (Overlap), untereinander zu einem größeren Sequenzabschnitt (Contig, Contiguous sequence) zusammengefügt werden können. Bei diesem „Assembly“ genannten Prozess kann grob zwischen zwei Problemstellungen unterschieden werden: Beim Mapping werden die Reads in Übereinstimmung mit einem vorgegebenen bekannten „Template“ gebracht. Dies kann ein bereits sequenziertes Genom, Transkriptom oder Exom derselben oder einer möglichst nah verwandten Spezies sein. Bei der *de-novo*-Assemblierung ist dagegen nichts über die Zielsequenz bekannt. Für einen aktuellen Überblick über Assemblierungsmethoden siehe NAGARAJAN und POP (2013).

Erst nach der Assemblierung kann die eigentliche Analyse der Sequenzen am Computer erfolgen. Tabelle 2 listet die wichtigsten Datensammlungen mit Ergebnissen aus öffentlichen Pflanzengenomprojekten auf.

Tab. 2 Pflanzengenetische Datensammlungen.

Tab. 2 *Plant genetics databases.*

Name	Webseite	Literatur
The Arabidopsis Information Resource (TAIR)	http://www.arabidopsis.org/	(LAMESCH <i>et al.</i> , 2012)
Ensembl plants	http://plants.ensembl.org/	(FLICEK <i>et al.</i> , 2013)
MIPS Plants DB	http://mips.helmholtz-muenchen.de/plant/genomes.jsp	(NUSSBAUMER <i>et al.</i> , 2013)
Gramene	http://www.gramene.org/	(YOUENS-CLARK <i>et al.</i> , 2011)
Phytozome	http://phytozome.net/	(GOODSTEIN <i>et al.</i> , 2012)
<i>Brachypodium distachyon</i>	http://www.brachypodium.org/	(THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010)
PlantGDB	http://www.plantgdb.org/	(DUVICK <i>et al.</i> , 2008)
PlantCYC / Plant Metabolic Network (PMN)	http://www.plantcyc.org/	(ZHANG <i>et al.</i> , 2010)
1001 Genomes - A Catalog of <i>Arabidopsis thaliana</i> Genetic Variation	http://www.1001genomes.org/	(CAO <i>et al.</i> , 2011)

Anwendungsgebiete

Im Folgenden sollen die einzelnen Methoden zur Analyse von Genom, Metagenom, Exom, Transkriptom, Metabolom und Epigenom näher betrachtet werden.

Genom

Das Genom umfasst die Gesamtheit der genetischen, vererbaren Information eines Organismus. Die Sequenzierung des gesamten Genoms ermöglicht die Untersuchung von Variationen (Single Nucleotide Polymorphisms (SNP) and Structural Variants) innerhalb einer Spezies (GAN *et al.*, 2011; EDWARDS *et al.*, 2013) und Unterschieden zwischen verschiedenen Spezies. Problematisch bei der Rekonstruktion der kompletten Genomsequenz sind die durch die Sequenzieretechnik bedingten sehr kurzen Reads und das Auftreten repetitiver Sequenzen, die bei der Assemblierung nicht eindeutig zugeordnet werden können.

Neben einer Reihe von inzwischen komplett sequenzierten Nutzpflanzen wie *Oryza sativa* (GOFF *et al.*, 2002), *Hordeum vulgare* (THE INTERNATIONAL BARLEY GENOME SEQUENCING CONSORTIUM *et al.*, 2012), *Zea mays* (WEI *et al.*, 2009), *Triticum aestivum* (BRENCHLEY *et al.*, 2012) und *Glycine max* (SCHMUTZ *et al.*, 2010) gilt *Brachypodium distachyon* (THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010) aufgrund seines kleinen Genoms inzwischen als geeigneter Modellorganismus für Gräser.

Da bisher nur wenige Pflanzengenome komplett sequenziert sind, werden hierfür in erster Linie *de-novo*-Assemblierungstools verwendet (SCHATZ *et al.*, 2012).

Metagenom

Das Metagenom bildet die Gesamtheit der genomischen Information der Mikroorganismen einer bestimmten Lebensgemeinschaft (Biozönose) oder eines Biotops. Da die Herkunftsspezies der genomischen Sequenzen hierbei nicht bekannt sind, müssen durch den Vergleich mit bereits bekannten Sequenzen zunächst die in der Probe (meist Umweltprobe aus Wasser oder Boden) enthaltenen Spezies bestimmt werden. Danach kann dann ihr Einfluss auf und ihre Interaktion mit z. B. Nutzpflanzen genauer untersucht werden.

Geeignete Programme hierfür sind MEGAN (HUSON *et al.*, 2007), mothur (SCHLOSS *et al.*, 2009) und CLOTU (KUMAR *et al.*, 2011).

Exom

Das Exom besteht aus den potenziell für Proteine kodierenden Abschnitten (Exons) der Gene eines Genoms. Um das Exom sequenzieren zu können, müssen zunächst die Exons aus dem Genom selektiert werden (sequence capture). Dies setzt voraus, dass eine Referenz-Genomsequenz der zu untersuchenden Spezies bereits bekannt ist. Da das Exom bei höheren Organismen nur einen kleinen Teil des Genoms ausmacht, ist es sehr viel kostengünstiger, das Exom anstelle weiterer Genome dieser Spezies zu sequenzieren, um Variationen in den kodierenden Bereichen zu untersuchen (MASCHER *et al.*, 2013).

Transkriptom

Das Transkriptom repräsentiert die Gesamtheit der zu einem bestimmten Zeitpunkt unter bestimmten Bedingungen in einem Gewebe exprimierten Gene (d.h. die von der DNA in mRNA transkribierten Gene). Unter Genexpression versteht man dabei die Ausprägung des Genotyps zum Phänotyp eines Organismus oder eines Gewebes. Da die mRNA nicht direkt sequenziert werden kann, muss sie zunächst in cDNA (complementary DNA) mit Hilfe des Enzyms Reverse Transkriptase umgeschrieben werden und die cDNA kann dann anschließend sequenziert werden. Diese Technik wird als RNA-Seq bezeichnet (WANG *et al.*, 2009). Das Transkriptom liefert zusammen mit dem Genom und dem Exom Informationen darüber, welche Exons durch alternatives Spleißen (alternative splicing) eines Gens zu einem bestimmten Zeitpunkt in welche mRNA transkribiert werden (USADEL und FERNIE, 2013). Neben den auf Homologiesuche basierenden weit verbreiteten Tools BLAST, PSI-BLAST (ALTSCHUL *et al.*, 1997), BLAT (KENT, 2002) und InterProScan (QUEVILLON *et al.*, 2005) werden zur Annotation der sequenzierten Transkripte zunehmend auch Ontologien mit einem standardisierten Wortschatz verwendet (LOHSE *et al.*, 2012).

Für einen Überblick über die Transkriptomanalyse siehe GÓNGORA-CASTILLO und BUELL (2013), geeignete Programme für die Assemblierung eines Transkriptoms sind Trinity (GRABHERR *et al.*, 2011) und Oases (SCHULZ *et al.*, 2012).

Neben der qualitativen Information über die in einer Probe enthaltenen Transkripte hinaus lassen sich mit RNA-Seq auch quantitative Daten gewinnen, d.h. Genexpressionsdaten, für die bisher Microarray-Experimente durchgeführt wurden. Auch wenn inzwischen ein großes Repertoire an statistischen Methoden zur Auswertung von Microarray-Experimenten existiert, haben sie den Nachteil, dass nur bekannte Transkripte nachgewiesen und verschiedene Experimente nur sehr aufwändig miteinander verglichen werden können (WANG *et al.*, 2009). Durch die Weiterentwicklung geeigneter Tools zur Auswertung von RNA-Seq-Experimenten werden Microarray-Experimente daher in diesem Bereich zunehmend verdrängt.

Zum exakten qualitativen und quantitativen Nachweis genetischen Materials bzw. der Expression einzelner Gene ist die Quantitative Real-Time-PCR (qRT-PCR oder qPCR) geeignet. Hierbei werden während der Durchführung der PCR Fluoreszenzsignale der amplifizierten Sequenzen gemessen. Voraussetzung für einen quantitativen Nachweis ist das Vorhandensein eines Referenzgens, dessen Expression gleichzeitig gemessen wird und zur Normalisierung der Expression des Zielgens dient. Als Referenzgen wird üblicherweise ein „housekeeping gene“ verwendet, dessen Expression unabhängig von äußeren Einflüssen in allen Zellen und Geweben konstant sein sollte. Da auch hier die Expression über das Vorhandensein von mRNA gemessen wird, für die Durchführung der PCR jedoch DNA benötigt wird, muss die mRNA zunächst mit Hilfe des Enzyms Reverse Transkriptase in cDNA umgeschrieben werden. Das gesamte Verfahren wird dann als RT-PCR (reverse transcription PCR) bezeichnet (HOLZAPFEL und WICKERT, 2007; DUHOUX und DÉLYE, 2013).

Metabolom

Die Metabolomik umfasst alle charakteristischen Stoffwechsel-Eigenschaften einer Zelle, eines Gewebes oder eines Organismus und geht damit über die Analyse von Genom, Transkriptom und Proteom hinaus hin zum Verständnis des Gesamtsystems (Systembiologie) einschließlich der

beteiligten Metaboliten (MACEL *et al.*, 2010). Da bisher weder Hochdurchsatz-Technologien zur Charakterisierung und Quantifizierung von Metaboliten noch Referenzdatenbanken existieren, ist die Metabolomik das bisher am wenigsten fortgeschrittene ‚omik‘-Fachgebiet (HEGEMAN, 2010; DÉLYE, 2013). Zukünftige Entwicklungen werden dann auch eine gemeinsame Betrachtung von Metagenom- und Metabolomdaten (environmental metabolomics) möglich machen (BRUNETTI *et al.*, 2013).

Zur Analyse von Veränderungen im Stoffwechsel von Pflanzen unter unterschiedlichen Bedingungen können jedoch bereits Expressions- und Transkriptomdaten z. B. mit dem Tool MAPMAN auf bekannte Stoffwechselwege projiziert werden (THIMM *et al.*, 2004; USADEL *et al.*, 2009).

Epigenom

Unter Epigenetik werden alle genetischen Faktoren zusammengefasst, die auf Veränderungen an den Chromosomen (z. B. DNA-Methylierung, Histon-Modifikation) beruhen, ohne dass eine Veränderung der DNA-Sequenz zugrunde liegt. Einzelne Abschnitte oder ganze Chromosomen können so in ihrer Aktivität beeinflusst werden. Diese Veränderungen können durch Umwelteinflüsse (z. B. große Trockenheit) verursacht sein und an die Folgegeneration vererbt werden (KINOSHITA und JACOBSEN, 2012). Neueste Entwicklungen (Bisulfit-Sequenzierung, BS-Seq) ermöglichen erste Einblicke in das Methylierungsmuster kompletter Genome (DINH *et al.*, 2012).

Ausblick

Die Entwicklung neuer Sequenzieretechniken ermöglicht die Bearbeitung völlig neuer Fragestellungen sowohl in der Medizin als auch den Agrarwissenschaften. Dadurch sind jedoch auch die Anforderungen an leistungsfähige Soft- und Hardware stark gestiegen und die Menge an Daten ist der aktuellen Entwicklung von Methoden (SCHLESKY *et al.*, 2012) und gebrauchsfähiger Software weit voraus. Damit ist momentan nicht mehr die Erzeugung der Daten der einschränkende Faktor sondern ihre anschließende Speicherung, Verarbeitung, Analyse und Interpretation am Computer. Um die vielfältigen Möglichkeiten der verfügbaren computergestützten Methoden optimal nutzen zu können, ist auf der einen Seite eine Entwicklung zu benutzerfreundlicherer Software (GOFF *et al.*, 2011; SMITH, 2013) und auf der anderen Seite die Bereitschaft von Medizinern, Biologen und Agrarwissenschaftlern nötig, diese Methoden als weitere Werkzeuge in ihren Arbeitsalltag zu integrieren (SERVICE, 2013).

Literatur

- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER und D. J. LIPMAN, 1997: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**(17), 3389-3402.
- BRENCHLEY, R., M. SPANNAGL, M. PFEIFER, G. L. BARKER, R. D'AMORE, A. M. ALLEN, N. MCKENZIE, M. KRAMER, A. KERHORNOU, D. BOLSER, S. KAY, D. WAITE, M. TRICK, I. BANCROFT, Y. GU, N. HUO, M.-C. LUO, S. SEHGAL, B. GILL, S. KIANIAN, O. ANDERSON, P. KERSEY, J. DVORAK, W. R. MCCOMBIE, A. HALL, K. F. X. MAYER, K. J. EDWARDS, M. W. BEVAN und N. HALL, 2012: Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**(7426), 705-710.
- BRUNETTI, C., R. M. GEORGE, M. TATTINI, K. FIELD und M. P. DAVEY, 2013: Metabolomics in plant environmental physiology. *J. Exp. Bot.* **64**(13), 4011-4020.
- CAO, J., K. SCHNEEBERGER, S. OSSOWSKI, T. GÜNTHER, S. BENDER, J. FITZ, D. KOENIG, C. LANZ, O. STEGLE, C. LIPPERT, X. WANG, F. OTT, J. MÜLLER, C. ALONSO-BLANCO, K. BORGWARDT, K. J. SCHMID und D. WEIGEL, 2011: Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* **43**(10), 956-963.
- DÉLYE, C., 2013: Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: a major challenge for weed science in the forthcoming decade. *Pest Manag. Sci.* **69**(2), 176-187.
- DINH, H. Q., M. DUBIN, F. J. SEDLAZECK, N. LETTNER, O. MITTELSTEN SCHEID und A. VON HAESELER, 2012: Advanced methylome analysis after bisulfite deep sequencing: an example in *Arabidopsis*. *PLoS One.* **7**(7), e41528.
- DUHOUX, A., und C. DÉLYE, 2013: Reference genes to study herbicide stress response in *Lolium sp.*: up-regulation of P450 genes in plants resistant to acetolactate-synthase inhibitors. *PLoS One.* **8**(5), e63576.
- DUVICK, J., A. FU, U. MUPPIRALA, M. SABHARWAL, M. D. WILKERSON, C. J. LAWRENCE, C. LUSHBOUGH und V. BRENDDEL, 2008: PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* **36**(Database issue), D959-D965.
- EDWARDS, D., J. BATLEY, und R. J. SNOWDON, 2013: Accessing complex crop genomes with next-generation sequencing. *Theor. Appl. Genet.* **126**(1), 1-11.

- FLICEK, P., I. AHMED, M. R. AMODE, D. BARRELL, K. BEAL, S. BRENT, D. CARVALHO-SILVA, P. CLAPHAM, G. COATES, S. FAIRLEY, S. FITZGERALD, L. GIL, C. GARCÍA-GIRÓN, L. GORDON, T. HOURLIER, S. HUNT, T. JUETTSMANN, A. K. KÁHÁRI, S. KEENAN, M. KOMOROWSKA, E. KULESHA, I. LONGDEN, T. MAUREL, W. M. McLAREN, M. MUFFATO, R. NAG, B. OVERDUIN, M. PIGNATELLI, B. PRITCHARD, E. PRITCHARD, H. S. RIAT, G. R. S. RITCHIE, M. RUFFIER, M. SCHUSTER, D. SHEPPARD, D. SOBRAL, K. TAYLOR, A. THORMANN, S. TREVANION, S. WHITE, S. P. WILDER, B. L. AKEN, E. BIRNEY, F. CUNNINGHAM, I. DUNHAM, J. HARROW, J. HERRERO, T. J. P. HUBBARD, N. JOHNSON, R. KINSELLA, A. PARKER, G. SPUDICH, A. YATES, A. ZADISSA and S. M. J. SEARLE, 2013: Ensembl 2013. *Nucleic Acids Res.* **41**(Database issue), D48-D55.
- GAN, X., O. STEGLE, J. BEHR, J. G. STEFFEN, P. DREWE, K. L. HILDEBRAND, R. LYGNSOE, S. J. SCHULTHEISS, E. J. OSBORNE, V. T. SREEDHARAN, A. KAHLES, R. BOHNERT, G. JEAN, P. DERWENT, P. KERSEY, E. J. BELFIELD, N. P. HARBERD, E. KEMEN, C. TOOMAJIAN, P. X. KOVER, R. M. CLARK, G. RÄTSCH and R. MOTT, 2011: Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**(7365), 419-423.
- GOFF, S. A., M. VAUGHN, S. MCKAY, E. LYONS, A. E. STAPLETON, D. GESSLER, N. MATASCI, L. WANG, M. HANLON, A. LENARDS, A. MUIR, N. MERCHANT, S. LOWRY, S. MOCK, M. HELMKE, A. KUBACH, M. NARRO, N. HOPKINS, D. MICKLOS, U. HILGERT, M. GONZALES, C. JORDAN, E. SKIDMORE, R. DOOLEY, J. CAZES, R. McLAY, Z. LU, S. PASTERNAK, L. KOESTERKE, W. H. PIEL, R. GRENE, C. NOUTSOS, K. GENDLER, X. FENG, C. TANG, M. LENT, S.-J. KIM, K. KVILEKVAL, B. S. MANJUNATH, V. TANNEN, A. STAMATAKIS, M. SANDERSON, S. M. WELCH, K. A. CRANSTON, P. SOLTIS, D. SOLTIS, B. O'MEARA, C. ANE, T. BRUTNELL, D. J. KLEIBENSTEIN, J. W. WHITE, J. LEBBENS-MACK, M. J. DONOGHUE, E. P. SPALDING, T. J. VISION, C. R. MYERS, D. LOWENTHAL, B. J. ENQUIST, B. BOYLE, A. AKOGLU, G. ANDREWS, S. RAM, D. WARE, L. STEIN and D. STANZIONE, 2011: The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Front. Plant Sci.* **2**, 34.
- GOFF, S. A., D. RICKE, T.-H. LAN, G. PRESTING, R. WANG, M. DUNN, J. GLAZEBROOK, A. SESSIONS, P. OELLER, H. VARMA, D. HADLEY, D. HUTCHISON, C. MARTIN, F. KATAGIRI, B. M. LANGE, T. MOUGHAMER, Y. XIA, P. BUDWORTH, J. ZHONG, T. MIGUEL, U. PASZKOWSKI, S. ZHANG, M. COLBERT, W.-I. SUN, L. CHEN, B. COOPER, S. PARK, T. C. WOOD, L. MAO, P. QUAIL, R. WING, R. DEAN, Y. YU, A. ZHARKIKH, R. SHEN, S. SAHASRABUDHE, A. THOMAS, R. CANNINGS, A. GUTIN, D. PRUSS, J. REID, S. TAVTIGIAN, J. MITCHELL, G. ELDRIDGE, T. SCHOLL, R. M. MILLER, S. BHATNAGAR, N. ADEY, T. RUBANO, N. TUSNEEM, R. ROBINSON, J. FELDHAUS, T. MACALMA, A. OLIPHANT and S. BRIGGS, 2002: A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296**(5565), 92-100.
- GÓNGORA-CASTILLO, E. and C. R. BUELL, 2013: Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Nat. Prod. Rep.* **30**(4), 490-500.
- GOODSTEIN, D. M., S. SHU, R. HOWSON, R. NEUPANE, R. D. HAYES, J. FAZO, T. MITROS, W. DIRKS, U. HELSTEN, N. PUTNAM and D. ROKHSAR, 2012: Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**(Database issue), D1178-D1186.
- GRABHERR, M. G., B. J. HAAS, M. YASSOUR, J. Z. LEVIN, D. A. THOMPSON, I. AMIT, X. ADICONIS, L. FAN, R. RAYCHOWDHURY, Q. ZENG, Z. CHEN, E. MAUCELI, N. HACOHEH, A. GNIIRKE, N. RHIND, F. DI PALMA, B. W. BIRREN, C. NUSBAUM, K. LINDBLAD-TOH, N. FRIEDMAN and A. REGEV, 2011: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**(7), 644-652.
- HEGEMAN, A. D., 2010: Plant metabolomics - meeting the analytical challenges of comprehensive metabolite analysis. *Brief. Funct. Genomics.* **9**(2), 139-148.
- HOLZAPFEL, B. und L. WICKERT, 2007: Die quantitative Real-Time-PCR (qRT-PCR). *Methoden und Anwendungsgebiete. Biol. Unserer Zeit.* **37**(2), 120-126.
- HUSON, D. H., A. F. AUCH, J. QI and S. C. SCHUSTER, 2007: MEGAN analysis of metagenomic data. *Genome Res.* **17**(3), 377-386.
- KENT, W. J., 2002: BLAT - the BLAST-like alignment tool. *Genome Res.* **12**(4), 656-664.
- KINOSHITA, T. and S. E. JACOBSEN, 2012: Opening the door to epigenetics in PCP. *Plant Cell Physiol.* **53**(5), 763-765.
- KUMAR, S., T. CARLSEN, B.-H. MEVIK, P. ENGER, R. BLAALID, K. SHALCHIAN-TABRIZI und H. KAUSERUD, 2011: CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics.* **12**, 182.
- LAMESCH, P., T. Z. BERARDINI, D. LI, D. SWARBRECK, C. WILKS, R. SASIDHARAN, R. MULLER, K. DREHER, D. L. ALEXANDER, M. GARCIA-HERNANDEZ, A. S. KARTHIKEYAN, C. H. LEE, W. D. NELSON, L. PLOETZ, S. SINGH, A. WENSEL und E. HUALA, 2012: The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**(Database issue), D1202-D1210.
- LOHSE, M. und B. USADEL, 2012: Klassifikation pflanzlicher Genome im Schnelldurchlauf. *BIOspektrum* **18**(3), 277-279.
- LOHSE, M., A. M. BOLGER, A. NAGEL, A. R. FERNIE, J. E. LUNN, M. STITT und B. USADEL, 2012: RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **40**(Web Server issue), W622-W627.
- MACEL, M., N. M. VAN DAM und J. J. KEURENTJES, 2010: Metabolomics: the chemistry between ecology and genetics. *Mol. Ecol. Resour.* **10**(4), 583-593.
- MASCHER, M., T. A. RICHMOND, D. J. GERHARDT, A. HIMMELBACH, L. CLISSOLD, D. SAMPATH, S. AYLING, B. STEURNAGEL, M. PFEIFER, M. D'ASCENZO, E. D. AKHUNOV, P. E. HEDLEY, A. M. GONZALES, P. L. MORRELL, B. KILIAN, F. R. BLATTNER, U. SCHOLZ, K. F. X. MAYER, A. J. FLAVELL, G. J. MUEHLBAUER, R. WAUGH, J. A. JEDDELOH und N. STEIN, 2013: Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* [Epub ahead of publication].
- NAGARAJAN, N. und M. POP, 2013: Sequence assembly demystified. *Nat. Rev. Genet.* **14**(3), 157-167.
- NUSSBAUMER, T., M. M. MARTIS, S. K. ROESSNER, M. PFEIFER, K. C. BADER, S. SHARMA, H. GUNDLACH und M. SPANNAGL, 2013: MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* **41**(Database issue), D1144-D1151.
- QUEVILLON, E., V. SILVENTOINEN, S. PILLAI, N. HARTE, N., MULDER, R. APWEILER und R. LOPEZ, 2005: InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**(Web Server issue), W116-W120.
- SCHATZ, M. C., J. WITKOWSKI und W. R. MCCOMBIE, 2012: Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* **13**(4), 243.
- SCHLIESKY, S., U. GOWIK, A. P. WEBER und A. BRÄUTIGAM, 2012: RNA-Seq Assembly - Are We There Yet? *Front. Plant Sci.* **3**, 220.

- SCHLOSS, P. D., S. L. WESTCOTT, T. RYABIN, J. R. HALL, M. HARTMANN, E. B. HOLLISTER, R. A. LESNIEWSKI, B. B. OAKLEY, D. H. PARKS, C. J. ROBINSON, J. W. SAHL, B. STRES, G. G. THALLINGER, D. J. VAN HORN und C. F. WEBER, 2009: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**(23), 7537-7541.
- SCHMUTZ, J., S. B. CANNON, J. SCHLUETER, J. MA, T. MITROS, W. NELSON, D. L. HYTEN, Q. SONG, J. J. THELEN, J. CHENG, D. XU, U. HELSTEN, G. D. MAY, Y. YU, T. SAKURAI, T. UMEZAWA, M. K. BHATTACHARYYA, D. SANDHU, B. VALLIYODAN, E. LINDQUIST, M. PETO, D. GRANT, S. SHU, D. GOODSTEIN, K. BARRY, M. FUTRELL-GRIGGS, B. ABERNATHY, J. DU, Z. TIAN, L. ZHU, N. GILL, T. JOSHI, M. LIBAULT, A. SETHURAMAN, X.-C. ZHANG, K. SHINOZAKI, H. T. NGUYEN, R. A. WING, P. CREGAN, J. SPECHT, J. GRIMWOOD, D. ROKHSAR, G. STACEY, R. C. SHOEMAKER und S. A. JACKSON, 2010: Genome sequence of the palaeopolyploid soybean. *Nature* **463**(7278), 178-183.
- SCHULZ, M. H., D. R. ZERBINO, M. VINGRON und E. BIRNEY, 2012: Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**(8), 1086-1092.
- SERVICE, R. F., 2013: Biology's Dry Future. *Science* **342**(6155), 186-189.
- SMITH, D. R., 2013: The battle for user-friendly bioinformatics. *Front. Genet.* **4**, 187.
- THE ARABIDOPSIS GENOME INITIATIVE, 2000: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**(6814), 796-815.
- THE INTERNATIONAL BARLEY GENOME SEQUENCING CONSORTIUM, K. F. MAYER, R. WAUGH, J. W. BROWN, A. SCHULMAN, P. LANGRIDGE, M. PLATZER, G. B. FINCHER, G. J. MUEHLBAUER, K. SATO, T. J. CLOSE, R. P. WISE und N. STEIN, 2012: A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**(7426), 711-716.
- THE INTERNATIONAL BRACHYPODIUM INITIATIVE, 2010: Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**(7282), 763-768.
- THIMM, O., O. BLÄSING, Y. GIBON, A. NAGEL, S. MEYER, P. KRÜGER, J. SELBIG, L. A. MÜLLER, S. Y. RHEE und M. STITT, 2004: MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**(6), 914-939.
- USADEL, B., F. POREE, A. NAGEL, M. LOHSE, A. CZEDIK-EYSENBERG und M. STITT, 2009: A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant Cell Environ.* **32**(9), 1211-1229.
- USADEL, B. und A. R. FERNIE, 2013: The plant transcriptome - from integrating observations to models. *Front. Plant Sci.* **4**, 48.
- WANG, Z., M. GERSTEIN und M. SNYDER, 2009: RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**(1), 57-63.
- WEI, F., J. ZHANG, S. ZHOU, R. HE, M. SCHAEFFER, K. COLLURA, D. KUDRNA, B. P. FAGA, M. WISSOTSKI, W. GOLSER, S. M. ROCK, T. A. GRAVES, R. S. FULTON, E. COE, P. S. SCHNABLE, D. C. SCHWARTZ, D. WARE, S. W. CLIFTON, R. K. WILSON und R. A. WING, 2009: The physical and genetic framework of the maize B73 genome. *PLoS Genet.* **5**(11), e1000715.
- WETTERSTRAND, K. A., 2013: DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: <http://www.genome.gov/sequencingcosts>. Accessed 29.09.2013.
- YOUENS-CLARK, K., E. BUCKLER, T. CASSTEVENS, C. CHEN, G. DECLERCK, P. DERWENT, P. DHARMAWARDHANA, P. JAISWAL, P. KERSEY, A. S. KARTHIKEYAN, J. LU, S. R. MCCOUCH, L. REN, W. SPOONER, J. C. STEIN, J. THOMASON, S. WEI und D. WARE, 2011: Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* **39**(Database issue), D1085-D1094.
- ZHANG, P., K. DREHER, A. KARTHIKEYAN, A. CHI, A. PUJAR, R. CASPI, P. KARP, V. KIRKUP, M. LATENDRESSE, C. LEE, L. A. MUELLER, R. MULLER und S. Y. RHEE, 2010: Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol.* **153**(4), 1479-1491.