2 Genomics for Bioenergy Production

Robert J. Henry
University of Queensland

CONTENTS

2.1 Introduction ............................................................................................................................ 21
2.2 Applications of Genomics in the Development of Energy Crops ...........................................22
2.3 Evolutionary Relationships in Higher Plants and Their Genomes ...........................................22
2.4 Genome Sequencing ...............................................................................................................22
2.5 Analysis of Genome Variation .................................................................................................23
2.6 Transcriptome Analysis .........................................................................................................23
2.7 Proteomics of Bioenergy Crops .............................................................................................23
2.8 Metabolomics in Relation to Bioenergy Production ...............................................................24
2.9 Target Traits for Bioenergy Plant Improvement ......................................................................24
2.10 Genetic Modification to Add Novel Bioenergy Traits ............................................................24
2.11 Promoters and Control of Expression ....................................................................................25
2.12 Model Bioenergy Crops .........................................................................................................25
2.13 Genomics of Specific Bioenergy Species .............................................................................25
  2.13.1 Sorghum ..........................................................................................................................25
  2.13.2 Sugarcane (and Close Relatives) ......................................................................................26
  2.13.3 Maize ............................................................................................................................26
  2.13.4 Other Grasses ..................................................................................................................26
  2.13.5 Poplar ............................................................................................................................26
  2.13.6 Eucalyptus .......................................................................................................................27
2.14 Future Prospects for Use of Genomics in Bioenergy Production from Plants .......................27
References ........................................................................................................................................27

2.1 INTRODUCTION

Genomics has become a key tool for the analysis of plants and their performance for different end uses (Henry 2010a and b). Genomics allows for the simultaneous examination of the genes within a given organism, rather than the single-gene approach traditionally used. This paradigm has provided a tool that has greatly increased our knowledge of biological systems. Applications of genomics to food and feed uses of plants have previously dominated research efforts, but the application to energy use is now beginning. Plants have always been the basis of human food and have been burnt for heating, but they have only more recently been considered as potential sources of energy more widely. Humans have come to rely on energy from fossil plants in the form of oil or coal. The direct use of plants to satisfy energy requirements has become a key alternative as these fossil resources become limiting and their consumption threatens to cause global warming along with all of its well-known consequences. Humans domesticated plants by selecting genotypes that perform well under cultivation but may not be well equipped to survive in the wild (Purugganan
and Fuller 2009). Crop plants were also mostly domesticated by selection of the genome for food use. The selection of plants for energy production will require isolating and selecting for previously ignored traits in already domesticated species or selecting previously undomesticated but promising species (Simmons et al. 2008; Henry 2010b). Domestication of plants for food has produced many domesticated plants with nonstructural carbohydrates (e.g., sugars and starches) that are sources of energy for animals and humans. These carbohydrates have been exploited in the first generation of bioenergy production from domesticated plants. Second-generation bioenergy crops are widely recognized as those that are available to be utilized for their more abundant structural (cell wall) carbohydrate with the potential to utilize much more of the carbon in these plants for energy. The ancient process of plant domestication was not always a deliberate one, but the opportunity now exists to domesticate with specific objectives and using the tools of modern science to achieve in a few generations what previously required hundreds. Genomics provides tools for a comprehensive analysis of the plant and the potential for selection of genotypes better suited to any given human use. Thus, genomics is a powerful tool for use in the accelerated domestication and improvement of plants as energy crops. Genomics may also be applied to the engineering of plant and nonplant organisms for the conversion of plant biomass to bioenergy. The application of genomics at these two levels may be complementary or potentially even synergistic in achieving the goal of energy- and cost-efficient bioenergy production.

2.2 APPLICATIONS OF GENOMICS IN THE DEVELOPMENT OF ENERGY CROPS

Genomics can be applied at many different levels in the development of plant species as bioenergy crops:

- In identifying higher plants for which genomes show potential as bioenergy crops,
- In identifying genes for desirable bioenergy traits,
- In screening and selecting superior bioenergy genotypes, and
- In supporting efforts to modify plant genomes, making them better bioenergy crops.

2.3 EVOLUTIONARY RELATIONSHIPS IN HIGHER PLANTS AND THEIR GENOMES

Evolutionary or phylogenetic approaches (Henry 2005) can be applied to the search for suitable species or traits for bioenergy production. The composition of a given plant’s biomass is a major determinant of the suitability of that plant for use in specific bioenergy production processes. Cell wall composition is a major factor, and this varies by plant group. A molecular phylogenetic approach will ultimately allow the plants with the appropriate sets of genes for the desired composition to be found efficiently.

2.4 GENOME SEQUENCING

The last few years have seen the emergence of radically improved technology for DNA sequencing (Schuste 2008), making large-scale plant genome analysis much more feasible. Genome sequencing of potential bioenergy crop plants provides a platform for analysis of the genetic potential of these species and targets for their genetic improvement.

The sequence of the sorghum genome (Paterson et al. 2009) provides a reference genome for not only sorghum but also the many closely related grass species that are potential bioenergy crops (e.g., sugarcane and Miscanthus). Despite the polyploidy complexity of the sugarcane genome, significant efforts are now being made to obtain a reference genomic sequence for this species because of
its existing importance as an energy crop. The eucalypt, the most widely planted forest species in tropical and subtropical parts of the world, is being sequenced because of its energy potential. This, combined with the available poplar genome sequence, will provide a foundation for genomics approaches in woody biomass crops.

2.5 ANALYSIS OF GENOME VARIATION

Genotyping an entire genome allows association genetics to be applied to the discovery of the genetic basis of important traits. Analysis of variation in plant genomes can now be considered an analysis of all of the single-feature polymorphisms. This includes all of the insets and deletions (indels) and the single nucleotide polymorphisms (SNPs). Analytically indels can usually be detected or assayed as SNPs. SNP discovery and analysis in plants has recently been reviewed (Henry and Edwards 2009). SNP discovery even in complex plant genomes is now facilitated by advancing DNA sequencing technologies (Bundock et al. 2009; Trick et al. 2009). Efficient targeted mutagenesis techniques have been developed for the discovery of naturally occurring or induced mutations (Cross et al. 2008). Analysis of variation in known SNPs in a plant population is now possible with very high-throughput techniques (Edwards et al. 2009; Masouleh et al. 2009) and is being advanced by the application of nanotechnology (Pattemore et al. 2008). Association genetics has often considered only a few candidate genes. The identification of the genetic basis of gelatinization temperature in plant starches is a good example of this because this trait may be important in determining the energy required to gelatinize starch for conversion to sugars and subsequently into a biofuel. Starch biosynthesis genes were the obvious candidates, and analysis revealed that SNPs altering key amino acid residues in the soluble starch synthases of rice could individually be responsible for an 8°C reduction in gelatinization temperature (Waters et al. 2007). This alters the structure of the amyllopectin, allowing water to penetrate the starch granule more easily. This approach would complement the expression of thermostable amylases in plants to digest the starch during processing to biofuels (Wolt and Kraman 2007). Genomics tools are making the consideration of all options possible, and these developments are important when candidate genes are not readily identifiable for the trait of interest.

2.6 TRANSCRIPTOME ANALYSIS

The same tools that have revolutionized DNA sequencing have the potential to allow detailed analyses of genes expressed in different plants or tissues at different times. Transcriptome analysis allows the identification of genes controlling key traits for selection in plant domestication and improvement for human use. This approach involves the identification of candidate genes by determining their differential expression in association with the trait of interest. Candidate genes are then usually screened to determine function and to confirm association with the trait. Microarray analysis has been widely applied to transcript analysis in gene discovery. Early complementary DNA (cDNA)-based arrays were replaced by oligonucleotide arrays as the availability of more gene sequence data for more species made this possible. Advances in DNA sequencing technology offer techniques for comprehensive transcript profiling by sequencing, even in species without well-characterized genomes.

2.7 PROTEOMICS OF BIOENERGY CROPS

In some cases, the plant proteome will be the key area for analysis, although for bioenergy traits associated with carbohydrate composition this is probably less likely. However, proteins (for animal feed or other higher value applications) may be an important co-product in the energy crop. Proteomics is likely to make important contributions to understanding of the regulation of genes determining bioenergy traits.
2.8 METABOLIC NETWORKS IN RELATION TO BIOENERGY PRODUCTION

Metabolism will be important in understanding and improving bioenergy crops, especially in establishing maximal metabolic efficiency. Metabolomics will also guide metabolic engineering to ensure carbon is stored in the most accessible form for use in bioenergy production. Metabolomics may also be important if high-value co-product molecules are being produced in the energy crop. For example, the production of plastics in plants has been approached by developing the ability to produce compounds such as polyhydroxybutyrate (Somleva et al. 2008). The production of high-value co-products may be necessary to ensure that the processing of the biofuel crop is economically attractive.

2.9 TARGET TRAITS FOR BIOENERGY PLANT IMPROVEMENT

Identifying traits for improvement in potential bioenergy crops remains a major challenge. A general list of these desirable traits is provided in Table 2.1. Many of these are complex traits that are important targets for dissection using genomic approaches. Additional species-specific traits and traits suiting specific bioenergy applications need to be added to this list to target specific species at desired energy outcomes.

The development of second generation bioenergy crops that utilize plant cell walls will be greatly advanced by an increased understanding of the genomics of cell wall biosynthesis and degradation (Fincher 2009a, 2009b).

2.10 GENETIC MODIFICATION TO ADD NOVEL BIOENERGY TRAITS

Transgenic plants with novel attributes may have superior performance as bioenergy crops. The expression of enzymes in plants to aid their processing into fuel molecules is a good example of this. Expression of carbohydrate degrading enzymes (Taylor et al. 2008) that target structural or nonstructural carbohydrates has the potential to greatly increase the efficiency of biofuel production and reduce the cost of conversion of biomass to biofuel. Genes improving lignin degradation may also be useful in increasing the efficiency of biofuel production (Liang et al. 2008). Genetic engineering of plants for biofuel production will be aided by genomic knowledge, especially of the

| TABLE 2.1 |
| Traits That May Be Considered in Bioenergy Crop Development |
| High biomass accumulation |
| High harvest index |
| High fraction of biofuel in harvested biomass |
| Nutrients partition to nonharvested parts |
| Able to be grown on marginal lands |
| Harvested material able to be stored in the field |
| High bulk density |
| High water use efficiency |
| High nitrogen use efficiency |
| Low potential as a weed |
| High co-product potential |
| Optimal biomass composition |
| Large scale of potential production |
| Low cost of harvest |
| High suitability for genetic improvement |
species being transformed. Metabolic engineering requires knowledge of the existing pathways within the plant, which are best defined at the genomic level. The use of techniques such as zinc-finger nucleases may allow specific modification of plants to produce superior biofuel crops by the addition of new genes (Shukla et al. 2009) or modification of existing ones (Townsend et al. 2009). Thus, genomics has accelerated the discovery of genes that might be manipulated to produce superior bioenergy crops.

2.11 PROMOTERS AND CONTROL OF EXPRESSION

Knowledge of regulatory systems at the genome level will be important for breeding and selecting superior crop varieties and especially for the precise control of important transgenes. The manipulation of regulatory genes may be an important path to significant gene improvements in energy crops. The production of transgenic bioenergy crops with useful transgenes will require the use of gene promoters that direct transgene expression, at appropriate levels, in the required tissue and at the necessary time during plant development. The species specificity of these processes (Furtado et al. 2008) will require detailed analysis and understanding of regulatory processes controlling gene expression in target species. Exploiting our growing knowledge (Held et al. 2008) of the role of small RNA molecules in the regulation of plant performance and composition may also be important in the development of bioenergy crops.

2.12 MODEL BIOENERGY CROPS

Several crops for which genomics tools are available are model systems for analyses of the processes required to adapt plants to be better energy sources. Although not an energy crop, Arabidopsis (a general plant model) and rice (a grass species model) models are very useful for research on energy crops. The rice genome is now very well documented and characterized. Sequencing of many genotypes and of related species in the genus is increasing genomic knowledge of rice. However, the sorghum plant is a much better model for other C₄ bioenergy species such as sugarcane and maize than are C₃ plants like rice. Maize has been proposed as a useful model genome for energy crops in the grasses (Lawrence and Walbort 2007).

2.13 GENOMICS OF SPECIFIC BIOENERGY SPECIES

Plant genomes are generally large compared with other organisms (Table 2.2). However, recent advances in DNA sequencing technology have greatly increased the rate at which plant genomes are being characterized.

2.13.1 SORGHUM

Sorghum is a very efficient C₄ plant with good biomass potential and tolerance to hot and dry environments. Sorghum was domesticated as a food crop and is used today mainly as a source of animal feed. The development of sorghum as an energy crop will require selection for increased plant biomass rather than grain yield. This could lead to a significant change in plant architecture. The genome sequence of sorghum has been reported by Paterson et al. (2009). Sweet sorghum has been identified as a source of biomass for first-generation biofuel production. However, the development of high-biomass sorghum genotypes may provide an important source of biomass for biofuel production in the environments to which sorghum is adapted. Sorghum can be produced in locations without the water that might be required for other energy crops such as maize or sugarcane. The availability of the genome sequence and other genomic resources in sorghum should facilitate the rapid development of these improved types.
2.13.2 **Sugarcane (and Close Relatives)**

Sugarcane is a very efficient C₄ plant grown in tropical and subtropical regions (Henry 2010). Sugarcane (a *Saccharum* hybrid) is probably the leading industrial crop internationally with significant amounts being used for ethanol production and electricity generation. Sugarcane is currently a first-generation biofuel crop with the sugar being used to produce ethanol on a large scale. Sugarcane was first domesticated for the high sugar content of the canes, but development of sugarcane as an energy crop may result in the selection of plants for biomass yield rather than sugar content, resulting in a reduced sugar content.

Close relatives of sugarcane such as *Miscanthus* (Yamada and Henry 2011) may also have potential as energy crops. Hybrids between sugarcane and other species within the *Saccharum* complex (e.g., *Miscanthus* and *Erianthus* species) may provide important options for the development of energy crops. Molecular tools will be essential to the efficient introgression of genes from these other species.

2.13.3 **Maize**

Maize is a major food crop with most maize being used to feed farmed animals. The development of maize as an energy crop or as a dual-purpose food and energy crop will be greatly advanced by applications of genomics. The protein residue remaining after converting the grain carbohydrates to ethanol has become a key animal feed ingredient. Genomics approaches may be used to improve the nutritional value of this protein. Maize may be a useful model genome for grass improvement for bioenergy (Carpita and McCann 2008).

2.13.4 **Other Grasses**

Most grasses have been domesticated and used for food and pasture production. The many grass species (more than 10,000 worldwide) may include species overlooked for these uses but that have potential as energy crops. Species such as switchgrass (*Panicum virgatum*), a native grass from North America, are being intensely evaluated as potential bioenergy crops (McLaughlin et al. 1999). Analysis of grasses that might suit bioenergy production in specific environments may identify many new options for domestication.

2.13.5 **Poplar**

The poplar is a fast-growing woody biomass option for many temperate and cooler climates. The poplar genome has been sequenced by Tuskan et al. (2006). Willows (*Salix* species), a closely related
species, are also potential bioenergy crops. The poplar genome has provided a model genome for research on wood plant species and their use for bioenergy and other applications.

2.13.6 Eucalyptus

Eucalypts are the most planted hardwood tress in tropical and subtropical parts of the world (Henry 2011). Members of this group of more than 700 species are adapted to a wide range of environments. Eucalypts have high biomass accumulation rates and a tolerance of marginal growing conditions, making them good candidates for the production of wood biomass. Many genomics tools for eucalypts, including genetic maps and gene and genome sequences have been developed in recent research (Henry 2011).

2.14 FUTURE PROSPECTS FOR USE OF GENOMICS IN BIOENERGY PRODUCTION FROM PLANTS

The future of energy crops should see the application of genomics to the discovery and manipulation of genes to create optimally designed energy plants. The efficiency of photosynthesis may be enhanced by selecting or engineering plants with an optimal metabolism for specific environments. For example, C₄ pathways of photosynthesis may improve the efficiency of carbon fixation, especially at high light intensities in warm and dry environments. Plant architecture may be optimized to capture solar energy and use it to fix and store the carbon at a high density in tissues that can then be easily harvested and converted to useful bioenergy molecules. Biomass composition may be selected to best suit available biofuel conversion technologies. For example, altered lignin content and modified carbohydrate composition are major options. Also, the proportion of cellulose and noncellulosic polysaccharides may be altered. Most importantly, the linkages between different carbohydrate polymers and between carbohydrate and lignin may be reduced or altered. Genomics will provide the tools to characterize the genetic control of these key biofuel traits. Enzymes required in any conversion process could all be produced in the plant to reduce costs and improve energy efficiency.

Future biomass crops need to be designed for sustainable production with all nutrients being recycled to the soil. Nitrogen needs to be supplied by the crop itself or by other crops in the rotation. Plants also need to be designed to maximize the value of co-products remaining after energy extraction. This may be an essential feature of the economics of growing bioenergy crops. Some of these co-products may be industrial or pharmaceutical products. However, high-value animal feeds and human food co-products have the advantage of reducing the competition between food and energy production. Protein residues remaining after conversion of plant carbohydrates to energy may be significant co-product options. Genomics approaches may allow for analyses of bioenergy and co-product potential and allow selection or breeding for both characteristics.

Genomics also has the potential to make a major contribution to future bioenergy production on the basis of the novel production of bioenergy from algae. Algae also have potential as a direct source of high-value fuel molecules (e.g., alkanes). Thus, algae may also represent an alternative source of biomass.

REFERENCES


Edme SJ, Comstock JC, Miller JD, Tai PYP (2005) Determination of DNA content and genome size in sugar-
Fincher GB (2009a) Exploring the evolution of (1,3,1,4)-beta-D-glucans in plant cell walls: Comparative
genomics can help! Curr Opin Plant Biol 12:140–147
Fincher GB (2009b) Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the
679–693
RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall bio-
synthesis in barley. Proc Natl Acad Sci USA 105:20534–20539
accelerating plant biotechnology. Plant Biotechnol J 7:311
Publishers, Oxon, United Kingdom
Henry RJ (2010a) An overview of advances in genomics in the new millennium. In: Kole C, Abbott AG (eds),
NH and CRC Press, Boca Raton, FL
Henry RJ (2010) Basic information on the sugarcane plant. In: Henry RJ, Kole C (eds), Genetics, Genomics
Forest Trees. Springer, New York, in press
Lawrence CJ, Walbot V (2007) Translational genomics for bioenergy production from grasses: Maize as the
model species. Plant Cell 19:2091–2094
Liang HY, Frost CJ, Wei XP, Brown NR, Carlson JE, Tien M (2008) Improved sugar release from lignocel-
lulosic material by introducing a tyrosine-rich cell wall peptide gene in poplar. Clean Soil Air Water
36:662–668
Masouleh A, Waters D, Reinke R, Henry RJ (2009) A high-throughput assay for rapid and simultaneous analy-
sis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF
McLaughlin S, Bouton JD, Bransby D, Conger B, Ocumpaugh W, Parrish D, Talkiefero C, Vogel K, and
Crops and New Uses. ASHS Press, Alexandria, VA pp 282–299
457:551–556
Henry RJ (ed), Plant Genotyping II: SNP Technology. CABI Publishers, Wallingford, United Kingdom,
pp 133–153
Shukla VK, Doyon Y, Miller JC, deKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan
9:242
Somleva MN, Snell KD, Beaulieu JJ, Peoples OP, Garrison BR, Patterson NA (2008) Production of polyhy-
droxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop.
Plant Biotechnol J 6:663–678
Townsend JA, Wright DA, Winfrey RJ, Fu F, Maeder ML, Joung JK, Voytas DF (2009) High-frequency modi-
Trick M, Long Y, Meng L, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the poly-

