

Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda

A Research Roadmap Resulting from the Biomass to Biofuels Workshop Sponsored by the U.S. Department of Energy

December 7–9, 2005, Rockville, Maryland

DOE/SC-0095, Publication Date: June 2006

Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program
Office of Energy Efficiency and Renewable Energy, Office of the Biomass Program

[DOE Genomics:GTL](#)

[GTL Biofuels](#)

[Home Page This Document](#)

Chapter PDFs

- [Executive Summary](#) (257 kb)
- [Introduction](#) (1524 kb)
- [Technical Strategy: Development of a Viable Cellulosic Biomass to Biofuel Industry](#) (263 kb)
- System Biology to Overcome Barrier to Cellulosic Ethanol
 - [Lignocellulosic Biomass Characteristics](#) (794 kb) ← **Current File**
 - **Feedstocks for Biofuels** (834 kb)
 - [Deconstructing Feedstocks to Sugars](#) (632 kb)
 - [Sugar Fermentation to Ethanol](#) (1367 kb)
- [Crosscutting 21st Century Science, Technology, and Infrastructure for a New Generation of Biofuel Research](#) (744 kb)
- [Bioprocess Systems Engineering and Economic Analysis](#) (66 kb)
- [Appendix A. Provisions for Biofuels and Biobased Products in the Energy Policy Act of 2005](#) (54 kb)
- [Appendix B. Workshop Participants and Appendix C. Workshop Participant Biosketches](#) (529 kb)

John Houghton
Office of Science
Office of Biological and
Environmental Research
301.903.8288
John.Houghton@
science.doe.gov

Sharlene Weatherwax
Office of Science
Office of Biological and
Environmental Research
301.903.6165
Sharlene.Weatherwax@
science.doe.gov

John Ferrell
Office of Energy Efficiency
and Renewable Energy
Office of the Biomass
Program
202.586.6745
John.Ferrell@
hq.doe.gov

Feedstocks for Biofuels

One critical foundation for developing bioenergy crops and their processing technologies is ameliorating cell-wall recalcitrance to breakdown. Understanding cell walls is essential for optimizing their synthesis and the processes used to deconstruct them to sugars for conversion to ethanol (as discussed in the previous chapter, Lignocellulosic Biomass Characteristics, p. 39). A prerequisite for a competitive biofuel industry is the development of crops that have both desirable cell-wall traits and high biomass productivity under sustainable low-input conditions. Major agricultural crops grown today for food, feed, and fiber in the United States have not been bred for biofuels. Thus, many carefully selected traits in food and feed crops, such as a high ratio of seed to straw production (harvest index), are disadvantageous in biofuel production. A suite of new crops and new varieties of existing crops specifically bred for biofuels and adapted to a range of different soil types and climatic conditions is required.

During the past century, improvement of agricultural crops was supported by federal investment in many aspects of basic plant science, agronomy, plant breeding, pathology, agricultural engineering, and soil science. However, many topics particularly important in biofuel production have not been emphasized and are poorly developed as a result. A recent editorial in *Science* noted:

“There are major technological challenges in realizing these goals. Genetic improvement of energy crops such as switchgrass, poplar, and jatropha has barely begun. It will be important to increase the yield and environmental range of energy crops while reducing agricultural inputs. Plant development, chemical composition, tolerance of biotic and abiotic stresses, and nutrient requirements are important traits to be manipulated. The combination of modern breeding and transgenic techniques should result in achievements greater than those of the Green Revolution in food crops, and in far less time.” (Koonin 2006)

— Steven E. Koonin, Chief Scientist, BP, London

The Department of Energy (DOE) mission-oriented research program envisioned herein is designed to supplement current investment in plants, with the focus on facilitating rapid progress in formulating biomass feedstock crops, also referred to as “energy crops.” One workshop goal was to identify specific areas in which a focused research investment would speed progress toward an optimized feedstock supply for conversion to biofuels. *In general terms, the goal of feedstock development is to obtain maximum usable*

“Large and cost-effective energy production on a scale that significantly impacts petroleum use calls for new crops with yield and productivity not currently available....” (EERE 2003).

References: p. 80

organic carbon per acre in an environmentally and economically sustainable way. Many previous studies have indicated that minimizing such inputs as annual field preparation and fertilization implies the use of such perennials as switchgrass and poplar, thus decreasing costs (see Fig. 1. Switchgrass Bales from a 5-Year-Old Field in Northeast South Dakota, this page). A joint document of DOE and the U.S. Department of Agriculture (USDA)—*Biomass as a Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply* (Perlack et al. 2005)—called for perennial crops to provide about one-third of biomass-derived fuels for the initial phase of bioethanol development.

In addition, because transition to large-scale cultivation of dedicated energy crops may take years if not decades, research imperatives must be explored for optimizing the use of currently available agricultural crop and forestry residues. Sustainability will be a key issue in implementing the use of crop and forestry residues for biofuel, since the removal of crop residues can reduce organic carbon and nutrient levels in the soil and affect soil microbial community health (see section, Ensuring Sustainability and Environmental Quality, p. 68). More information about the composition and population dynamics of soil microbial communities is needed to facilitate modeling of long-term effects on soil fertility.

Current knowledge indicates that perennial species expected to be used for biofuel production improve soil carbon content and make highly efficient use of mineral nutrients. Development of perennial energy crops also may facilitate use of genetically diverse mixed stands rather than monocultures of single cultivars. Because conventional crops are grown as monocultures, relatively little research has been carried out on issues associated with growing mixed stands (see sidebar, A Billion-Ton Annual Supply of Biomass, p. 10, and sidebar, The Argument for Perennial Biomass Crops, p. 59).

Fig. 1. Switchgrass Bales from a 5-Year-Old Field in Northeast South Dakota in 2005. Each 1200-lb. bale represents 48 gallons of ethanol at a conversion rate of 80 gallons per ton. The cultivar used in this field has a yield potential of 5 to 6 tons per acre (corresponding to 400 to 500 gallons per acre) because it was bred for use as a pasture grass. In experimental plots, 10 tons per acre have been achieved. Processing goals target 100 gallons per ton of biomass, which would increase potential ethanol yield to 1000 gallons per acre. [Source: K. Vogel, University of Nebraska]



The Argument for Perennial Biomass Crops

Many major agricultural crops today are annual plants propagated from seed or cuttings at the beginning of each growing season. By contrast, crops developed and grown specifically for biofuel production are expected to be based on perennial species grown from roots or rhizomes that remain in the soil after harvesting the above-ground biomass. Perennial species are considered advantageous for several reasons (see Fig. 2. Attributes of an “Ideal” Biomass Crop, p. 61). First, input costs are lower than for annuals because costs of tillage are eliminated once a perennial crop is established. Additionally, long-lived roots of perennials may establish beneficial interactions with root symbionts that facilitate acquisition of mineral nutrients, thereby decreasing the amount of fertilizer needed. Some perennials also withdraw a substantial fraction of mineral nutrients from above-ground portions of the plant at the end of the season but before harvest (see Fig. A. Nitrogen Use Efficiency Theory for Perennials, below).

Perennial plants in temperate zones also may have significantly higher total biomass yield per unit of land area than comparable annual species. Perennials establish a photosynthetically active canopy more quickly in the spring and may persist longer in the fall (see Fig. B. Comparing Net Photosynthesis of Corn and Several Perennial Species, below). Thus, their annual solar-energy conversion efficiency is higher than that of annual plants with similar capabilities.

Perennial species have much lower fertilizer runoff than do annuals. For instance, comparing the native perennial switchgrass with corn indicates that switchgrass has about one-eighth the nitrogen runoff and one-hundredth the soil erosion. Perennial grasses harvested for biomass exhibit increased soil-carbon levels and provide habitat for up to five times as many species of birds. Finally, in contrast to annual row crops that typically are monocultures, increasing habitat diversity by growing several intermixed species of perennials may prove more feasible.

Fig. A. Nitrogen Use Efficiency Theory for Perennials.
[Source: S. Long, University of Illinois]

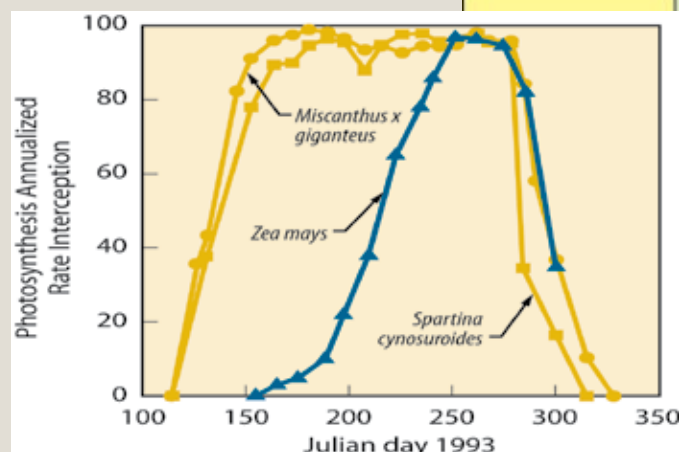
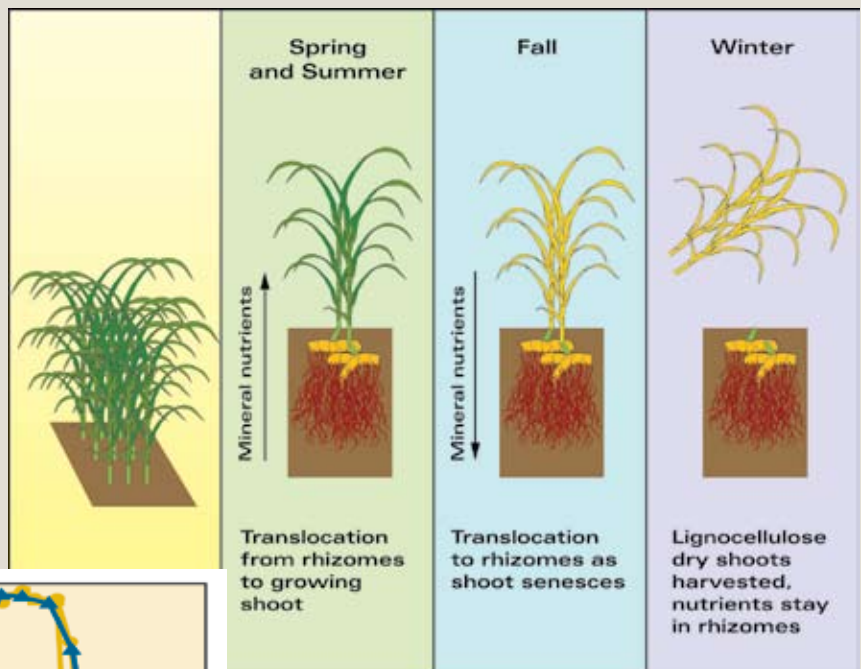


Fig. B. Comparing Net Photosynthesis of Corn and Several Perennial Species. Annualized net photosynthesis is proportional to the area under the curve. Thus, if maximal rates of photosynthesis are similar, the perennial crops (yellow) have much higher annualized net photosynthesis than the annual crop, corn (*Zea mays*, blue). [Source: S. Long, University of Illinois]

Among many factors in plant productivity, several are thought to be of central importance in biofuel production. Because of evidence that most plants do not routinely achieve maximal photosynthetic CO₂ fixation rates, understanding factors limiting the overall process is important. Emphasis should be placed on determining how plants allocate recently fixed carbon to products such as storage polysaccharides (e.g., starch) and structural polysaccharides such as cellulose. Energy crops will be grown on marginal, excess, or surplus agricultural lands, so identifying factors that facilitate tolerance and survival during exposure to drought, freezing, and other abiotic stresses will be vital. Issues for perennials may be quite different from those for annuals, which have been the subject of most research and crop experimentation. These issues may be particularly acute in regard to pressure from pests and pathogens that can be controlled to some extent by crop rotation in annual species. During the past 25 years, DOE's offices of Energy Efficiency and Renewable Energy (EERE) and Basic Energy Sciences have been a primary source of research support on centrally important topics in developing bioenergy feedstocks. Research on key issues identified here constitutes a compelling opportunity. Unique capabilities developed in both the Biomass and Genomics:GTL (GTL) programs can be brought to bear on remaining practical and fundamental problems in producing feedstocks. The following sections outline these issues and opportunities in the context of tangible goals, timelines, and milestones.

Creation of a New Generation of Lignocellulosic Energy Crops

Three distinct goals are associated with development of biofuel feedstocks:

- Maximizing the total amount of biomass produced per acre per year,
- Maintaining sustainability while minimizing inputs, and
- Maximizing the amount of fuel that can be produced per unit of biomass.

Exact values for each of these parameters will vary from one type of energy crop and one growing zone to another. A yield of 20 dry tons per acre per year may be considered a reasonable target in areas of the country with adequate rainfall and good soils, whereas 10 dry tons per acre per year may be acceptable in drier or colder zones.

Thus, the overall objective of developing feedstocks must be focused on broadly useful insights applicable to a variety of plant species grown under various growing conditions and exhibiting beneficial attributes (see Fig. 2. Attributes of an "Ideal" Biomass Crop, p. 61). This is best accomplished by working toward systems-level predictive models that integrate deep knowledge of underlying mechanisms for guiding cultivar and process development. This ambitious goal is only now beginning to be realized by companies that have bred advanced cultivars for major agricultural commodities.

This systems-level approach is feasible because of the last decade's biology revolution in genomic sequencing of higher plants and microbes. Sequencing

provides a means for connecting knowledge about all organisms into a common framework for understanding all forms of life. Current and future DOE investments in plant DNA sequencing afford the opportunity to create the mechanistic knowledge of energy crops needed for cost-effective and practical feedstocks. This enabling information must be elaborated further by strategic investments in understanding aspects of basic biology specifically relevant to energy crops (see sidebar, Enhancing Poplar Traits for Energy Applications, p. 62).

Maximizing Biomass Productivity

Domestication of Energy Crops

The continental United States is composed of a number of growing zones or agroecoregions that vary with such factors as mean temperature, rainfall, and soil quality. No single plant species is optimal for all zones, so using different species as energy crops will be necessary. Previous DOE studies have identified a number of promising plant species, and academic studies have suggested additional ones (see Fig. 3. Geographic Distribution of Biomass Crops, this page). In general, energy crops can be divided into two types: Those, such as maize, which are used for agricultural food and feed production but produce substantial amounts of usable biomass as a by-product (Type I), and those used only for energy (Type II). Type I plants are highly developed from many decades of research and study for another purpose (food or fiber).

The "Ideal" Biomass Crop?	Corn	Short-Rotation Coppice*	Perennial Grass
C4 photosynthesis	★		★
Long canopy duration		★	★
Recycles nutrients to roots			★
Clean burning			★
Low input		★	★
Sterile (noninvasive)	N/A	(★)	<i>M.g.**</i>
Winter standing		★	★
Easily removed	★		★
High water-use efficiency			★
No known pests or diseases			<i>M.g.</i>
Uses existing farm equipment	★		★

* Coppice is a grove of densely growing small trees pruned to encourage growth

** *Miscanthus giganteus*

Fig. 2. Attributes of an "Ideal" Biomass Crop. [Table adapted from S. Long, University of Illinois]



Fig. 3. Geographic Distribution of Biomass Crops. Multiple types designed for various agroecosystems probably will be required to obtain enough biomass for large-scale production of liquid fuels. [Source: Adapted from ORNL Biomass Program]

Enhancing Poplar Traits for Energy Applications

Gaining a better understanding of genes and regulatory mechanisms that control growth, carbon allocation, and other relevant traits in the poplar tree (*Populus trichocarpa*) may lead to its use as a major biomass feedstock for conversion to bioethanol. An international team led by the DOE Joint Genome Institute recently completed sequencing the poplar genome, making it the first tree (and fourth plant, after the mustard weed *Arabidopsis thaliana*, rice, and the alga *Chlamydomonas*) to have its complete genome sequenced (Tuskan et al., in press). These data now offer the molecular access needed to define, quantify, and understand—at a mechanistic level—basic biological processes that impact important traits.

Poplar was chosen for sequencing because of its relatively compact genome (500 million bases), only 2% that of pine. Moreover, many species are available worldwide, and their rapid growth allows meaningful measures of important traits within a few years. Extensive genetic maps already available include initial identification of markers associated with such traits.

Early comparative sequence analyses of poplar and *Arabidopsis* genomes are providing insights into genome structure and gene-family evolution; biosynthetic processes such as cell-wall formation, disease resistance, and adaptation to stress; and secondary metabolic pathways. Comparisons of gene-family sizes show substantial expansion of poplar genes involved in carbon to cellulose and lignin biosynthesis.

Moving from a descriptive to predictive understanding of poplar growth, development, and complex function will require integration of sequence information with functional data. These data will be generated by such new tools and approaches as gene and proteome expression studies, metabolic profiling, high-throughput phenotyping and compositional analysis, and modeling and simulation. Ultimately, this information will lead to the engineering of faster-growing trees that produce more readily convertible biomass (see below, Fig A. Vision for the Future). The International *Populus* Genome Consortium has produced a science plan to guide

postsequencing activities (*The Populus Genome Science Plan 2004–2009: From Draft Sequence to a Catalogue of All Genes Through the Advancement of Genomics Tools*, www.ornl.gov/ipgc).

Other areas to be addressed for poplar and other potential bioenergy crops include sustainability for harvesting biomass, harvesting technologies to remove biomass at low costs, and infrastructure technologies that allow biomass to be transported from harvest locations to conversion facilities.

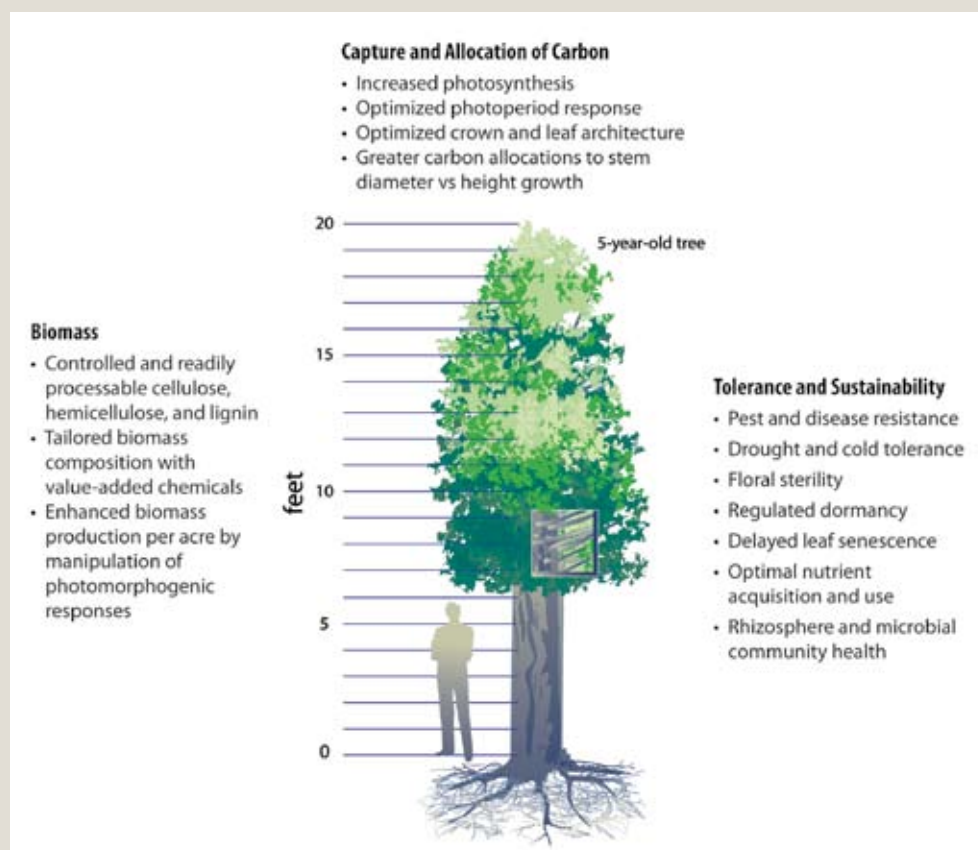


Fig. A. Vision for the Future: Desired Traits of the Domesticated Energy Poplar.

They were bred as monocultures under intensive agriculture. As noted elsewhere in this document, research priorities regarding production of energy from these crops are concerned largely with three issues. These issues are sustainability following removal of previously underutilized straw or stover, processing the residues, and the possibility of making improvements in cell-wall composition for enhanced conversion to fuels.

By contrast, Type II plants are relatively poorly developed as agronomic crops and generally are perennials. Perennial herbaceous and woody plants have several properties that make them better suited for biofuel production than are annual crops (see sidebar, *The Argument for Perennial Biomass Crops*, p. 59). First, because they typically retain a significant tissue mass below ground, they rapidly form a canopy in the spring and accumulate biomass when many annuals are still seedlings. Thus, they may exhibit higher rates of net photosynthetic CO₂ fixation into sugars when measured annually, resulting in higher amounts of total biomass accumulation per acre per year. Second, perennials require little if any tillage, saving energy and labor and significantly reducing soil erosion and nutrient loss. Perennials such as switchgrass and *Miscanthus* can be harvested annually without replanting. Third, perennials typically withdraw mineral nutrients into roots at the end of a growing season, thereby reducing fertilizer costs. Perennial herbaceous and woody plants represent a critical component of our bioenergy future.

Advances in conventional breeding, coupled with molecular tools and high-throughput transformation systems (Busov et al. 2005) will be required to accelerate domestication of species having promise as energy crops. Associating genotypes with phenotypes will require high-throughput technologies for genotyping (i.e., identifying genes and alleles) and phenotyping a wide array of relevant traits (e.g., biomass yield, cell-wall composition) with “molecular” techniques such as gene expression (transcriptomics), proteomics, and metabolomics. Availability of whole-genome sequences and deep expressed sequence tag (EST) libraries of all potential energy crops will be necessary to provide genetic reagents for marker-aided selection, association genetics, and transformation studies (see sidebar, *Marker-Assisted Breeding*, p. 64). Association genetics identifies allelic variants [e.g., single nucleotide polymorphisms (SNPs)] with empirically determined phenotypic effects on traits of interest and allows selection for favorable alleles using molecular markers (Rafalski 2002; Neale and Savolainen 2004; Remington et al. 2001). This form of marker-assisted selection is direct (as opposed to indirect on quantitative trait loci) and allows simultaneous selection of many genes. The result is a rational, marker-aided breeding and selection approach with the expectation of significantly enhanced genetic gain and an accelerated development process (Brown et al. 2003).

For most Type II biomass crops, significantly enhanced breeding, testing, and selection populations will be necessary, along with appropriate infrastructure to breed for desired traits and adaptability across a wide array of environments in multiple physiographic regions. In addition to targeted

Marker-Assisted Breeding

One of the ways in which DNA sequencing of biomass species impacts biofuel feedstock development is by enabling accelerated breeding methods to be applied to plants used for biomass production. The goal in plant breeding is to first identify useful genetic variation for traits of interest (e.g., disease resistance and drought tolerance) by screening natural or mutagenized populations of individuals. Many such traits are controlled by multiple genes. Individuals with useful variations are intercrossed to produce progeny with new combinations of the useful variation. Usually, many traits are of interest and many sources of variation are used, making the overall breeding process very time consuming and expensive. The availability of large amounts of sequence information facilitates identification of DNA polymorphisms—small differences in the DNA sequences of individuals within a species. Having complete genomic sequences also makes possible the identification of genes located near polymorphisms on chromosomes.

This knowledge has practical applications in “marker-assisted breeding,” a method in which a DNA polymorphism, closely linked to a gene encoding a trait of interest, is used to track the trait among progeny of sexual crosses between plant lines. The method allows breeders to monitor plants for a trait that may be expressed only in certain tissues or developmental stages or may be obscured by environmental variation. Similarly, by correlating traits and DNA polymorphisms in individuals from genetically diverse natural populations, associating a particular chromosomal region with a trait of interest frequently is possible. If a large number of polymorphisms are available, the amount of time required to breed an improved plant cultivar is greatly reduced.

In principle, plants with optimal combinations of parental genes may be identifiable within the first several generations following a sexual cross rather than eight generations or more following conventional breeding. In the case of species such as trees, marker-assisted breeding could eliminate many decades of expensive steps to develop more highly productive plants.

breeding, many crops will require fertility control in the field, either to ensure parentage or prevent gene flow to wild populations. A holistic approach for improving feedstocks includes molecular resources, high-throughput screening tools, and well-characterized breeding populations. GTL resources and technologies are well suited to creating comprehensive sets of molecular markers (i.e., SNPs and single sequence repeats) and high-throughput, low-cost phenotyping tools. These efficient deliverables will be used to develop appropriate cultivars and varieties to meet target goals. GTL genomic and other resources will play critically important roles via implementation of deep EST sequencing, marker identification, high-throughput genotyping, and development and application of analytical tools. These tools will be used for high-throughput molecular phenotyping of biomass composition and plant structure and a high-throughput transformation process for major biomass energy species. Such capability suites also will provide centralized bioinformatics support for analysis and archiving of genome data.

By definition cultivar development, field and plantation establishment and growth, and biomass conversion to biofuels necessitate a holistic, systems biology approach. Integrating the smaller subcomponents will be challenging, requiring a coordinated and focused program to facilitate exchange of information and genetic materials across organizations and institutions.

Enhancing the Yield of Biomass Crops

The yield of biomass crops can be defined as the amount of fixed carbon per acre per year. Achieving the maximal yield of a dedicated energy crop (Type II) is a significantly different goal from maximizing the yield of most existing crop species (Type I), where only the number of reproductive or storage organs is considered. The yield of a Type II species is a function of the total number of cells per acre multiplied by the mean amount of accumulated carbon per cell. Thus, biomass yield can be enhanced by increasing the number of cells per acre per year, the amount of carbon per cell, or both. Achieving either type of enhancement is a complex systems problem. At the core of the problem, however, is the need to maximize photosynthetic CO₂ fixation to support carbon accumulation. Additionally, fixed carbon must be directed into either cell-wall polymers or storage carbohydrates or used to support extra cell division. Cell-wall

polymers include cellulose, hemicellulose, and lignin; storage carbohydrates include sugars and starches.

Plants are regulated to fix only the carbon needed for normal growth and development. This generally is referred to as “source-sink” regulation, a poorly understood phenomenon. Plants can fix considerably more carbon, however; the actual photosynthetic CO₂ fixation rate of most or all plants is significantly below (i.e., ~50%) the rate observed following experimental partial defoliation. That is, plants appear to accumulate more carbon per unit of leaf area following defoliation than they normally would without any changes in architecture or photosynthetic electron transport. When mechanisms underlying this regulation are understood, plants can be developed that exhibit significantly higher rates of net photosynthetic CO₂ fixation and higher amounts of total carbon accumulation per acre per year. Therefore, a high-priority research goal is to understand mechanisms that regulate net photosynthetic CO₂ fixation. A closely related priority is to identify factors that limit carbon flux into cell-wall polysaccharides and storage polymers.

A complementary approach is to identify factors that regulate plant growth rate and duration. Different plant species vary widely in growth rates, suggesting that growth rates are under genetic control and, therefore, subject to modification. Recently, several genes have been identified in functional genomics screens that cause significant increases in growth rates of different types of plants. Identifying other genes that control growth and development and understanding gene action may create new opportunities to develop highly productive energy crops (see Fig. 4. Growth Rate Modification, this page).

Enhancing Abiotic Stress Tolerance of Biomass Species

Water availability is a major limitation to plant productivity worldwide, generally in two ways. First, because water escapes from plant leaves through stomata when CO₂ enters, a certain amount of water is required to support a unit of photosynthetic CO₂ fixation. Plants with C₄ photosynthesis (e.g., corn, sugarcane, switchgrass, and *Miscanthus*) typically require less water per unit of CO₂ fixed than do C₃ species (e.g., wheat and soybean) because C₄

plants can achieve high rates of CO₂ fixation with partially closed stomata. Other plants such as cacti close their stomata during the day to reduce water loss but open them at night to take in CO₂ for photosynthesis the next day. This phenomenon



Fig. 4. Growth Rate Modification. The *Arabidopsis* plant on the right has been modified by altering the expression of regulatory genes controlling growth. [Source: Mendel Biotechnology]

also inhibits carbon loss by photorespiration. Except for the possibility of enhancing these adaptations, current theory implies no other options for significantly reducing a plant's water requirement to obtain maximal yields.

The water problem's second component, however, concerns the effects of temporal variation in soil-water content. In rain-fed agriculture, periods of low soil-water content are frequent because of irregularities in rainfall. The ability of plants to survive extended periods of low soil water can be a critical factor in a crop's overall yield (see Fig. 5. Corn Yield on a Missouri Experiment Station, this page). Furthermore, different plants exhibit widely different abilities to survive extended periods of drought, indicating that drought-tolerant energy crops may be possible (see Fig. 6. Modification in Drought-Stress Tolerance, this page).

Currently, the most productive farmland is used for food production, with an amount held in reserve [Conservation Reserve Enhancement Program (CREP) lands]. While initial energy crops would be grown in highly productive CREP land, a substantial proportion of biomass crops will be grown on marginal land that is suboptimal in water availability, soil quality, or both. Recent progress in understanding the mechanistic bases of plant drought, salt, and cold tolerance has raised the possibility of modifying plants to enhance productivity under these and other stress conditions. A priority in dedicated energy crops is to understand mechanisms by which plants survive drought and adapt this knowledge to improving energy crops.

Fig. 5. Corn Yield on a Missouri Experiment Station.

This figure illustrates wide differences in yield observed in various annual growing conditions, including periods of drought. The strong upward trend in yield reflects advances in breeding and agronomical practices for corn. [Figure adapted from Q. Hu and G. Buyanovsky, "Climate Effects on Corn Yield in Missouri," *J. Appl. Meteorol.* 42(11), 1623–35 (2003).]

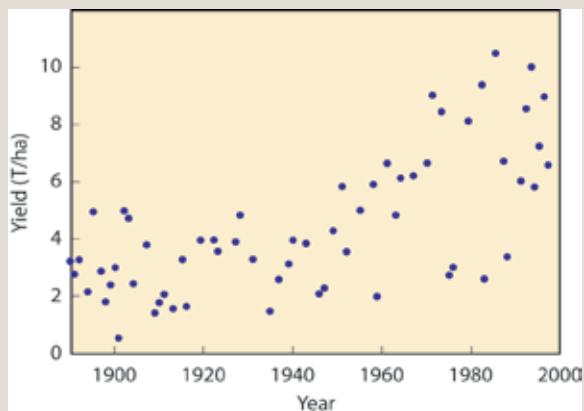
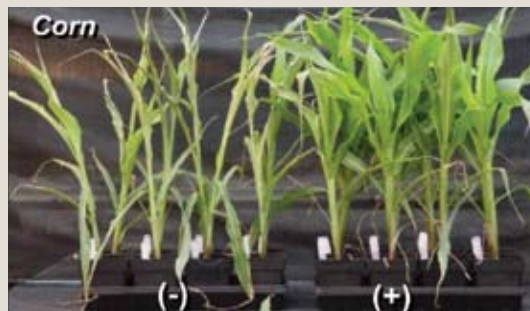


Fig 6. Modification in Drought-Stress Tolerance.

The plants on the left of each photograph are unmodified, and the plants on the right have been altered genetically for drought-stress tolerance. All plants were subjected to a severe drought. [Photos ©2003 Monsanto Company. Used with permission.]



Understanding and Eliminating Undesirable Biomass Crop Characteristics

Research should be conducted to understand and eliminate such undesirable plant traits as inappropriate residues and invasiveness of non-native energy-crop species. Two examples are described here.

Relatively large amounts of silica in plants such as rice lead to accumulations of ash when the plant biomass is burned for energy. When the biomass is converted to liquid fuels, unusable solids are produced. Since many plants do not accumulate silica, the trait appears to be dispensable; developing cultivars with reduced silica accumulation by genetic methods may be possible. The roles of silica in plant growth and development, however, are poorly understood and need further investigation.

Many features considered ideal for a biomass crop (see Fig. 2, p. 61) are characteristic of invasive weeds, particularly perennial C4 grasses. Thus, a key consideration in adapting these grasses for use as dedicated energy crops is to ensure that the species can be contained and will not become a problem. Some highly productive perennial grasses, such as *Miscanthus giganteus*, have been studied intensively in Europe for more than a decade and are thought not to exhibit invasive characteristics. All candidate energy crops, however, should be studied directly for potential invasiveness at diverse locations within the United States. These studies also can provide insights into pests and pathogens that might pose a threat to productivity.

Technical Milestones

Within 5 years

- Facilitate the initiation and advancement of biomass breeding programs in key U.S. geographic regions in cooperation with USDA, private companies, and universities.
- Develop appropriate test populations for conducting association genetics and quantitative trait locus identification (QTL, direct and indirect marker-assisted selection).
- Identify and target for selection and improvement key traits that affect biomass yield and conversion efficiency.
- In support of marker development for gene discovery, carry out targeted sequencing (SNPs, SSRs, ESTs, and cDNAs) for potential biomass species having large and complex genomes; sequence whole genomes for species with modest genome sizes.
- Develop markers that can differentiate superior parents and offspring.
- Initiate and validate methods for high-throughput screening for specific traits such as improved cell-wall digestibility.
- Formulate the infrastructure for functional screening of thousands of potentially useful genes in selected species.

Within 10 years

- Apply high-throughput phenotyping tools to integrated conventional and molecular breeding programs.

- Gain new understanding of genome structure and gene expression in bio-energy crops, including the genetic basis of heterosis (hybrid vigor). Major barriers to facile transformation of select genotypes will be overcome.

Within 15 years

- Identify and integrate major new domestication genes into energy crops.
- Integrate enabling technologies with breeding programs to allow for deployment of genetically superior materials over large land bases dedicated to biomass production. Time to commercial deployment will be shortened dramatically through highly reliable screening methods, genetic control of flowering and sexual reproduction, and effective early-selection models for predicting performance and yield. Genetically enhanced cultivars, hybrids, and varieties capable of meeting targeted goals for ethanol yield per acre per year will be available for major biomass crop species.

Ensuring Sustainability and Environmental Quality

To ensure the viability of bioethanol to meet the large national need for transportation fuels, we must understand the effects of long-term biomass harvesting on soil fertility and other aspects of sustainability. Objectives for this work are to determine how to maintain soil ecosystem function and productivity. A further objective is to develop management practices that can optimize sustainability. Because very long periods of time may be required for directly testing the effects of biomass removal on soil quality, a mechanistic understanding of this issue is essential to obtaining predictive models and monitoring procedures. The critical question in using crop residues and dedicated crops for biomass energy is, How much, if any, of above-ground biomass needs to be left on a field to protect soil from erosion and sustain soil function (soil quality)?

In the past 30 years, crop and soil research has emphasized using crop residues in minimum and no-till farming operations to improve soil organic carbon (SOC) and soil quality. Research has demonstrated that corn stover removal in some production systems can reduce grain yield (Wilhelm et al. 2004). The amount of stover removal and lower yield were associated with the amount of SOC (Maskina et al. 1993). At any time, SOC content is the balance between the rates of input and decomposition (Albrecht 1938). If all other cultural practices are unchanged, removal of crop residue will further decrease carbon inputs and SOC will decline (Follett 2001) (see Fig. 7. Soil Carbon Alterations with Management Changes, p. 69). Loss of SOC typically has detrimental effects on soil productivity and quality, presumably because microbial communities are impacted negatively. For example, considerable carbon is translocated to mycorrhizal fungus communities that play a significant role in soil nutrient cycling (Fitter et al. 2005).

This research must be completed for each major agroecosystem in the United States where crop residues and dedicated crops are a feasible supply source for biomass energy. Sustainability analysis of both dedicated energy crops and of stover or straw removal from agricultural crops should be carried out since the effects may be significantly different.

Soils are complex ecosystems composed of unknown numbers of organisms. The effects of crop-residue quantity and composition on soil ecosystems are largely unknown. Soil microbiologists need access to large-scale genomics and other analytical facilities to conduct soil microbial research. Systems biology analyses of highly productive soils, including their ecosystem-scale genomic characterization, are accessible with current sequencing capabilities. Questions include the following.

- What characteristics and functions of soil microbial communities are needed to maintain soil-ecosystem function and productivity?
- How do soil microbial communities function?
- What are the interactions, positive and negative, among microbes, fungi, and roots in the rhizosphere?
- How do nutrient levels affect microbial communities, and how do microbial communities affect nutrient availability?
- How can management of microbial communities improve productivity?
- How much carbon from crop residues and dedicated energy crops is needed to maintain soil ecosystem function and productivity?
 - Does the composition of residue have an effect?
 - What is happening both physically and biologically in soils as various levels of plant biomass are removed?
- Can management factors, no tillage, modified minimum tillage, or special practices affect the amount of needed residue?
- How does soil composition vary between annual and perennial crops in response to varying levels of biomass removal?
- What are the microbial-community characteristics of marginal and severely depleted lands, and how can they be restored to support energy crops?
- Can energy crops be used to restore lands for food and fiber crops by building up carbon and nutrients in marginal lands? Finally, how might these processes be utilized for reduction in greenhouse gas emissions or carbon sequestration?

GTL facility capabilities can be used in many ways to answer these general questions. In particular, GTL facilities could support ecosystem-analysis studies, including sequencing of microbial communities and genomics approaches for analysis of microbial-community functioning. GTL could support

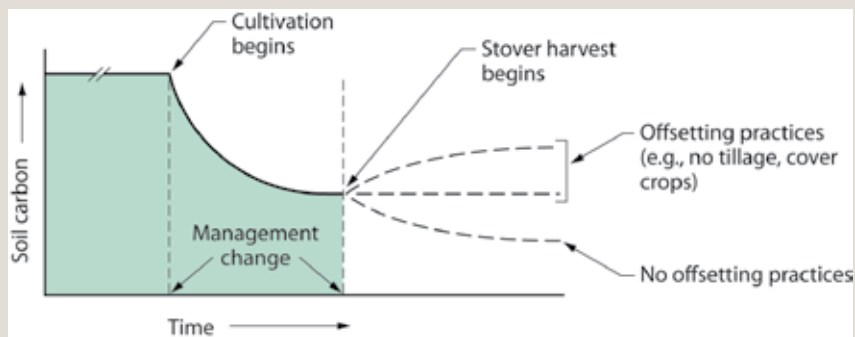


Fig. 7. Soil Carbon Alterations with Management Changes. Cultivation generally leads to reduction of soil organic carbon, which, without offsetting practices, is exacerbated by corn-stover removal. A mechanistic understanding of long-term harvesting effects on soil fertility and other aspects of sustainability is essential for predictive models and monitoring procedures. [Source: W. W. Wilhelm, USDA Agricultural Research Service, Lincoln, Nebraska.]

technologies to evaluate how microbial communities function by, for example, developing gene chips with millions of diagnostic probes.

Determination of microbial-community physiology must be linked with analysis of soil physicochemical states—facilitating analysis of soil emissions or carbon sequestration in soils. Improved methods of soil carbon analyses need to be developed and made available in high-throughput format. Current methods are extremely laborious. Improving analytical methods for analyzing composition of plant cell walls, described below, also may facilitate analysis of soil carbon composition.

Technical Milestones

Within 5 years

- Determine the effects of corn-stover and other crop-residue removal on soil productivity in each major agroecosystem of the United States.
- Use existing study sites where stover has been removed for 5 or more years to initiate comprehensive analysis of microbial communities (and soil carbon and mineral content) in underlying soils, in comparison with land where stover has not been removed.
- Establish long-term study sites for future sustainability studies on prospective energy crops in lands in all major agroecosystems.
- Pursue long-term contracts with private or public providers. The European *Miscanthus* Productivity Network, supported by the European Commission, provides a useful model. Sequence analysis of DNA extracted from soils should be carried out to survey diversity (Handelsman et al. 1998). Sequence analysis should be used to develop diagnostic methods for examining the dynamics of soil microbial composition and abundance.

Within 10 years

- Conduct additional studies with varying harvest rates and management practices, including cropping systems and analyses of microbial communities and of carbon levels in the soils.
- Perform comprehensive ecological system analyses of soils as a living medium.
- Include not only microorganisms present but also how their complex interactions contribute to their net function as a community. Markers of optimally functioning microbial communities should be developed to enable identification of sustainability requirements for soils in each major agroecosystem. In addition to its value to the energy-production function of croplands, this research should enhance the sustainability of food production.

Within 15 years

- Formulate management guidelines on the amount of biomass that can be removed in each major U.S. agroecosystem.

Model Systems for Energy Crops

Application of model systems toward the study of both basic and applied problems in plant biology has become routine and can quickly bring the techniques of 21st Century systems biology to bear on the complex problems associated with domesticating energy crops (see sidebar, Translational Research: The Path from Discovery to Applications, p. 72). Researchers using the model dicot *Arabidopsis thaliana* have made tremendous strides in understanding areas of plant biology ranging from nutrient uptake and metabolism to plant-pathogen interactions. Unfortunately, as an annual dicot, *Arabidopsis* is not an optimal model to study questions unique to potential woody and grassy perennial energy crops (e.g., wood formation in trees or cell-wall composition in grasses). Despite its sequenced genome and genetic resources, rice also is not an ideal model for grassy perennial energy crops because it is a specialized semiaquatic tropical grass. In addition, its large size, long generation time, and demanding growth requirements make experiments expensive. Using *Brachypodium* and *Populus* as model systems would provide researchers working to domesticate energy crops with some of the most powerful tools developed by the highly successful *Arabidopsis* community. These model systems would help identify genes controlling traits relevant to energy-crop productivity and quality, including such global processes as cell-wall biosynthesis, nutrient uptake, carbon flux, and plant architecture. The models can be used for rapid testing of strategies to improve the usefulness of grasses and trees as energy crops. Such tools would allow scientists to use both forward and reverse genetic approaches and modern molecular genetic methods to identify genes controlling traits relevant to the design of superior energy crops. This is important because of the problems (i.e., difficult transformation, large size, long generation time, and self-incompatibility) associated with working directly with energy crops.

Brachypodium distachyon is a small temperate grass with all attributes needed to be a modern model organism, including simple growth requirements, fast generation time, small stature, small genome size, and self-fertility (Draper et al. 2001). *Brachypodium* also is transformed readily by *Agrobacterium* (Vogel et al. 2006) or biolistics (Christiansen et al. 2005), thus facilitating many biotechnological applications. *Brachypodium* is now scheduled to be sequenced by the DOE Joint Genome Institute. The *Populus* genome has been sequenced (Brunner, Busov, and Strauss 2004; Tuskan, DiFazio, and Teichmann 2004), assembled, and annotated through investments from DOE, NSF, and a large international consortium (Tuskan et al., in press). This resource is available to researchers working on a dedicated woody crop for biofuels and biomass energy applications. Despite these advances, many molecular tools and resources currently not available would greatly enhance the discovery and use of genes and gene families related to energy traits and the development of poplar as an energy crop (see sidebar, Enhancing Poplar Traits for Energy Applications, p. 62). As with other model organisms, research on poplar would be facilitated by organism-specific tools such as full-genome DNA

Translational Research: The Path from Discovery to Applications

Model organisms play an important role in both basic and applied research. The most highly developed plant model is *Arabidopsis thaliana*, a small plant in the mustard family (Fig. 4. Growth Rate Modification, p. 65). It was adopted as a model in the early 1980s because of its rapid life cycle, simple genetics, small genome (125 Mb), easy transformability, and traits typical of flowering plants in most respects. Genome sequencing was completed in 2000 by an international consortium, and about 13,000 researchers worldwide currently use the plant for studies of plant biology.

The *Arabidopsis* community has developed a suite of very powerful experimental resources that include several hundred thousand sequenced insertion mutations available through stock centers, a full-genome DNA chip that can be used to measure the expression of most genes, a high-density polymorphism map, and a heavily curated database (arabidopsis.org) that provides access to genomic information. These and other resources have resulted in an explosion of fundamental discoveries about all aspects of plant growth and development. In addition, several companies, founded to exploit knowledge gleaned from *Arabidopsis*, have translated the new information into such crop species as corn, soybean, and canola (see Fig. A. Drought Resistance in Canola Caused by Directed Modification of a Single Gene, below). Plant-improvement goals that were unattainable by applied research on crop plants have been realized by translating basic research on model species to crops.

Much accumulated knowledge about crop species and model organisms will be applicable to improving species suited for use as dedicated biomass crops. However, some traits of interest in that regard have not been a priority in crop research and are poorly understood. For instance, knowledge about cell-wall structure, function, and synthesis is very underdeveloped and will need extensive research. One important species suited for use as a biofuel is poplar, a model woody species that has a relatively small genome. Poplar was sequenced recently at the DOE Joint Genome Institute (see sidebar, Enhancing Poplar Traits for Energy Applications, p. 62).

The DNA sequence of many poplar genes is similar to that of corresponding *Arabidopsis* genes. This comparability will help link mechanistic knowledge about these species to a holistic understanding of common processes. Poplar is relatively easy to transform, which greatly facilitates experimental tests of theories about protein and gene function. Most important, poplar has several important traits, such as a long life cycle and the formation of wood, that cannot be studied in herbaceous plants.

Brachypodium distachyon is another potentially important model for such highly productive grasses as switchgrass and *Miscanthus*. Interest in *Brachypodium* arises because its very small genome has a DNA content about 2.5 times larger than that of *Arabidopsis*. Additionally, its simple growth requirements, small stature, self-fertility, and ready transformability make it well suited to become a model organism. Because *Arabidopsis* is distantly related to and differs from grasses in a number of important respects (e.g., cell-wall composition), *Brachypodium* could

become a powerful new model for cell and molecular biological studies of grasses. A high priority in facilitating the development of *Brachypodium* is sequencing its genome.

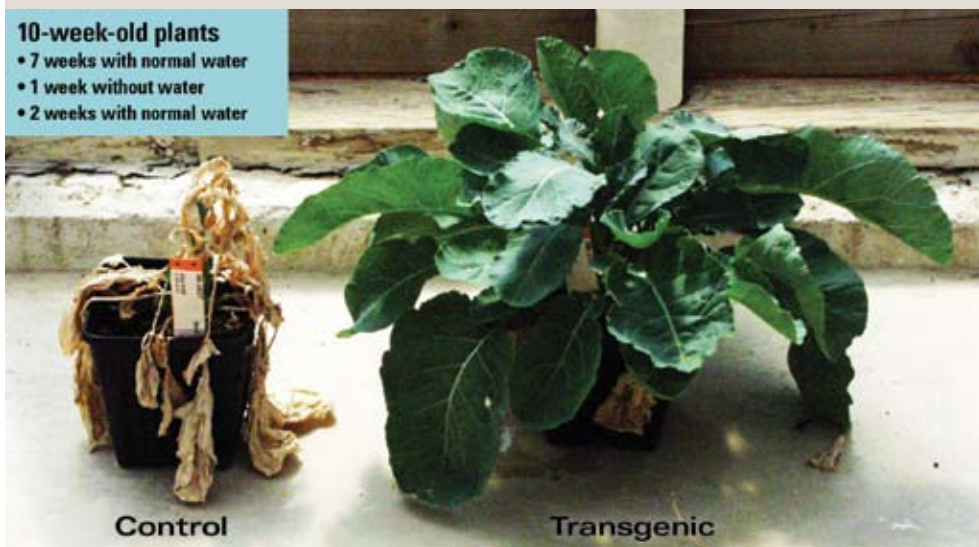


Fig. A. Drought Resistance in Canola Caused by Directed Modification of a Single Gene. [Photo ©2003 Monsanto Company. Used with Permission.]

chips, sequence-indexed mutations, facile transformation methods, and specialized libraries for methods including two-hybrid screens.

Brachypodium and *Populus* must be accelerated into powerful model systems by first sequencing *Brachypodium* and then creating extensive EST databases (i.e., 1 to 2 million per species) from diverse (including full-length) cDNA libraries for both organisms (Sterky et al. 2004). *Agrobacterium*-mediated transformation of both *Brachypodium* and *Populus* has been developed but must be optimized for high-throughput approaches to gene discovery. High-throughput capacity for transforming diverse genotypes would enable forward and reverse genetic approaches by generating collections of sequence-indexed insertional mutants, including but not limited to activation tagging (Groover et al. 2004, Busov et al. 2005), and by allowing targeted silencing or overexpression of large numbers of genes. Transposon-based methods for generating insertional mutants also should be developed. High-throughput resequencing capacities should be used for identifying molecular markers (e.g., SNPs) associated with biomass and bioenergy traits. Similarly, phenotyping (transcript and protein profiling) capacities that identify genotypes associated with relevant cell-wall composition or agronomic traits would be useful.

Using this suite of tools, the function of all *Brachypodium* and *Populus* genes will be evaluated for their contribution to biomass energy-relevant traits. Genes deemed relevant will be used to create advanced genotypes in energy crops. Public organism-specific databases should be created and curated in parallel with advanced experimental capabilities for these organisms.

To set research priorities, the DOE Office of Science's GTL program is interacting with the conversion program managed through EERE's Office of the Biomass Program and other interested parties (e.g., USDA, NSF, and private industry).

Technical Milestones

Within 5 years

- Complete *Brachypodium* genome sequencing, plus EST sequencing and molecular-marker identification for both *Brachypodium* and *Populus*.
- Optimize transformation methods for diverse genotypes of *Brachypodium* and *Populus*, and generate sequence-indexed insertion mutant populations of >100,000 events for both organisms.
- Develop tools for transcriptional and protein profiling and carry out first-stage profiling of key tissues (e.g., cambial development in *Populus*).
- Develop organism-specific databases around annotated genome sequences.

Within 10 years

- Identify genes relevant to biomass production (e.g., those that control cell-wall composition, nutrient uptake, carbon partitioning, flowering, stress tolerance, and disease resistance) using forward and reverse

genetic screens and natural populations based on high-throughput phenotyping [e.g., use Fourier transform infrared spectroscopy (called FTIR) or other techniques that can be automated to identify genotypes with altered cell-wall composition and advanced image analysis to examine plant architecture].

- Develop and test strategies using these genes to improve energy crops in *Brachypodium* and *Populus*.

Within 15 years

- In breeding programs, use genes and strategies shown to affect biomass production and quality in *Brachypodium* and *Populus* to improve energy crops, either by transgenic approaches or by marker-assisted selection of naturally occurring variability in orthologous genes.
- Transfer improved energy crops and genetic information to the conversion program managed through EERE and the Office of the Biomass Program and to other interested parties (e.g., USDA, NSF, and private industry).

The Role of GTL Capabilities for Systems Biology

Capabilities of GTL (within DOE's Office of Biological and Environmental Research) will support or enable the attainment of lignocellulosic characteristics (see Lignocellulosic Biomass Characteristics chapter, p. 39) and feedstock objectives in numerous ways. Potential contributions of each of these, whether distributed or consolidated in a facility, are described below.

Protein Production Capabilities

Most enzymes of interest in cell-wall biosynthesis are thought to be membrane associated and, therefore, difficult to purify and characterize by conventional methods. Based on preliminary experiments with membrane-localized glycosyltransferases, a potentially powerful tool in characterizing the function of all plant glycosyltransferases is to express them in suitable hosts such as insect cells, in which endogenous activities will not interfere with assays. This capability could undertake the high-throughput expression of all glycosyltransferases and other cell wall-active enzymes (e.g., peroxidases and laccases) from higher plants (e.g., *Arabidopsis*, *Populus*, *Brachypodium*, or rice) or specific biomass species in one or more suitable surrogate hosts. Such expression systems will greatly facilitate function identification of enzymes that catalyze synthesis of cell walls. Heterologously expressed proteins would be tested for their substrate specificity and catalytic activity and also will be used as a resource for generating antibodies and tags.

Antibody resources would have multiple applications. A collection of fluorescent- or epitope-tagged proteins in a plant such as *Arabidopsis* or *Brachypodium* also would be very useful for a variety of reasons. First, the tagged proteins could be used to recover complexes for assay by mass spectrometric and imaging methods. Fluorescent tags could be utilized for cellular localization or collocation [e.g., confocal microscopy or fluorescence

resonance energy transfer (called FRET)]. Additionally, if certain proteins are not active in a surrogate, the tagged version in a plant host might be used to identify function by in vitro assays of affinity-purified proteins. Developing such a resource would require a high-throughput transformation capability that also would be useful for generating insertion mutants in *Brachypodium* and *Populus*.

Molecular Machines Capabilities

Cellulose is synthesized by an intricate multienzyme complex in the plasma membrane, and matrix polysaccharides probably are synthesized by multienzyme complexes in the Golgi. Understanding the function of these complexes and their interactions with metabolic pathways that produce sugar nucleotides in the cytosol will be important for understanding the control of polysaccharide biosynthesis. In addition, cellulose synthase interacts with the cytoskeleton to control the orientation of cellulose fibrils and possibly influence fibril length. Many questions about this and related processes will benefit from the application of new technologies for imaging and manipulating protein complexes. Since most or all such complexes are membrane bound, specialized methods must be developed to work with the complexes. Also, because plant cell walls are intricate mechanical assemblies, tools developed for investigating protein machines may be used to gain new insights into assembly and function of the large polysaccharide complexes that comprise cell walls.

Proteomic Capabilities

These capabilities will facilitate identification of proteins that make up complexes in such model plants as *Arabidopsis*, *Populus*, and other biomass crops. Since many complexes are membrane bound, innovative methods will be needed to purify membrane complexes using an array of capabilities. Also, because of poor correlation between mRNA and protein abundance, proteomic capabilities will enable documentation of the proteome of living cell types found in vascular and other tissues, including ray parenchyma and phloem, as a function of time and conditions. Proteomic analysis of plants with fully sequenced genomes will identify post-translationally modified proteins. Preliminary evidence indicates that some key proteins involved in cell-wall synthesis are regulated by modifications such as phosphorylation. Finally, comprehensive proteomic analysis coupled with imaging analysis is required to identify proteins located in such compartments as the cell wall, nuclear membrane, Golgi, and endoplasmic reticulum.

Cellular System Capabilities

These capabilities are envisioned to help investigators understand how complicated microbial communities in soils respond to various cropping regimes by developing baseline analyses of species composition and abundance in suitable experimental plots. Also needed are cost-effective tools (e.g., diagnostic DNA chips) for monitoring these and other plots over

long time periods and for integrating resulting data into a view of how biomass cropping alters soil ecology.

In addition to analyzing soil microbial communities, acquisition and analysis of transcriptomic and proteomic data from model and crop plants will be a powerful tool for assigning probable function to genes implicated in cell-wall synthesis. As a function for each enzyme emerges, the cellular systems capability through imaging and dynamical analyses will be used to develop a systems model to incorporate the wall's biophysical aspects with structural properties and knowledge of the functions of proteins involved in wall synthesis. This model will facilitate the rational development of feedstock species based on "design principles," in which the wall's chemical composition and structure are optimized while, at the same time, not compromising maximal plant productivity. The cellular systems' analytical and modeling resource and proteomic capabilities will carry out a systematic approach for identifying plant biomarkers to guide protein production and the generation of specific molecular tags utilizing protein production capabilities.

The long-term vision is to make available in situ images of living plant cell walls that show all key molecular processes occurring in real time. Studies will cover the full life cycle of cell-wall formation, maturation, transformation, dehydration, and processing into simple feedstocks for conversion into ethanol or production of other fuels and chemicals. The understanding obtained through research that uses such imaging is expected to result in quantitative, predictive modeling as a guide to developing advanced feedstocks and processing them into fuels. These tools will take advantage of various chemically specific imaging tags created with protein production resources.

DOE Joint Genome Institute

DOE JGI will sequence, assemble, and annotate the gene space or entire genomes of model and emerging bioenergy crops and will resequence additional cultivars or ecotypes for marker discovery. Identification of genes that control cell-wall polysaccharide synthesis or modifications in biomass species and the development of tools such as gene chips will depend on the availability of nucleotide sequences. Micro-RNAs are expected to play a role in expression control of many relevant genes, so DNA sequencing must be deep enough to identify them in these species. JGI also will sequence community genomes, including those of the rhizosphere associated with key proposed energy crops and trees.

Other Needed Capabilities

GTL capabilities as currently envisioned do not encompass all resources needed for development of energy crops. To accomplish objectives associated with the energy mission, the current vision should be expanded to include the three capabilities described below.

Transformation Services

Substantial capabilities will be needed to carry out transformation of model and applied species important to feedstock development on behalf of the R&D community and to warehouse genetic resources for these species.

Chemical Phenotyping Services

Feedstock improvement for increased yield and processing will require manipulation and assessment of multiple genes in candidate organisms or model systems. New tools are needed to detect, quantify, and compare changes in cell-wall composition and 3D architecture in primary and secondary walls during assembly and before and during processing. The tools would facilitate feedstock development for improved performance in biomass conversion. Detailed analysis of cell-wall composition by 2D nuclear magnetic resonance (NMR) imaging or other methods is beyond the scope of routine processing in most research laboratories. Capabilities for such analyses would broadly facilitate research.

Synthetic Carbohydrate Chemistry

Analyses of cell walls in their native state and as a function of treatment and processing require advanced synthetic carbohydrate chemistry capabilities to create standards and models. NMR and mass spectroscopy facilities will be necessary to support the synthetic chemistry work.

Other Biofuel Opportunities: Development of High-Productivity Biodiesel Crops

To maximize solar energy use and storage, an ideal biomass crop would carry out photosynthesis at the theoretical maximum every day of the year and would store fixed carbon in a directly useful and easily harvestable form, even without mineral nutrients. In carrying out this process, many plants accumulate large amounts of oils or waxes in such specialized storage tissues as seeds or mesocarp tissues. In some cases, oil accounts for more than half the dry weight of these tissues. Most oilseed species, however, are not as productive as others such as maize that store primarily starch rather than oil. In theory, obtaining higher yields of oil should be possible by genetically modifying already highly productive species (e.g., sugar beet, potato, and maize) so they will accumulate oils or waxes instead of carbohydrates in their storage and vegetative tissues.

Other plants produced by genetic modifications of developmental processes accumulate oil in their roots (Ogas et al. 1997), suggesting the feasibility of this strategy. Acetyl-CoA carboxylase is the first committed step in fatty-acid biosynthesis in potato tubers, which accumulate starch almost exclusively. Overexpression of this enzyme resulted in a fivefold increase in triacylglycerol accumulation and provided an example of oil deposition in a carbohydrate-storage tissue (Klaus et al. 2004). However,

Decoding the DNA of Soybean—A Source of Biodiesel

DOE and the U.S. Department of Agriculture will support genome sequencing of the soybean as the first project in an agreement to share resources and coordinate studies of plant and microbial genomics. The soybean, *Glycine max*, is the world's most valuable legume crop and the principal U.S. source of biodiesel, a renewable alternative fuel. Diesel engines inherently are more thermodynamically efficient than combustion engines. Biodiesel has the highest energy content of any alternative fuel and is significantly more environmentally friendly than comparable petroleum-based fuels, since it degrades rapidly in the environment. It also burns more cleanly than conventional fuels, releasing only half the pollutants and reducing the production of carcinogenic compounds by more than 80%. Sequencing will take place at the DOE Joint Genome Institute, supported by DOE's Office of Science. The soybean genome is about 1.1 billion base pairs in size, less than half the size of maize and human genomes.

understanding how to reprogram plant cells to store oil rather than carbohydrate is very challenging. It requires not only large-scale changes in the complement of expressed metabolic enzymes but also changes in cellular structure typically associated with cell identity.

Because recovery of oil from plants is technically simple and efficient, processing oil-accumulating plant material on the farm may be possible so mineral nutrients and soil adjuvants can be retained at the farm for return to the soil (see sidebar, *Decoding the DNA of Soybean—A Source of Biodiesel*, p. 77). In principle, oil could be recovered during harvest by inexpensive screw-press technology, greatly reducing transport and processing costs and enhancing sustainability.

The lipids that comprise most plant oils are highly reduced forms of carbon and therefore represent the most energy-dense plant-storage compounds. Derivatives such as biodiesel can be produced easily by inexpensive and efficient conversion of plant-derived oil to fatty acid methyl esters. Production of high-yielding, oil-accumulating plants could result in a new biofuel source that would reduce loss of carbon associated with fermenting sugars to ethanol as well as costs associated with converting lignocellulosic feedstocks to liquid fuels. The plant that most nearly meets this description is oil palm (see sidebar, *Oil Palm*, this page).

Oil Palm: An Important Biofuel Plant

Oil palm, *Elaeis guineensis* Jacq., is a tropical tree species that produces bunches of oil-rich fruit resembling avocados. The plants grow in lowlands of the humid tropics (15°N to 15°S), where rainfall of 1800 to 5000 mm is evenly distributed throughout the year. Oil palms mature slowly but, once established, yield as much as 10.6 tonnes of oil per hectare (ha) per year, although the average is less than half that amount. They begin to bear fruit after about 3 years and remain in use for some 25 years, so annual maintenance costs are low.

Palm oil from the mesocarp, which contains 45 to 55% oil, is similar in composition to soy oil. Oil from palm kernels, extracted from the endosperm, contains about 50% oil rich in medium-chain fatty acids and well suited for biodiesel applications. Malaysia is the major source of these oils, producing 13.4 billion pounds of mesocarp oil and 3.5 billion pounds of kernel oil from about 3.8 million ha. Plants are harvested by hand, but typical planting density is only 150 plants/ha, and labor costs are not a major factor in production. Because of the relatively straightforward conversion of palm oils to diesel and food uses, palm acreage in the tropics is expected to expand significantly. [Picture source: C. Somerville, Stanford University]



Some questions to be addressed by basic research:

- How is cell identity controlled? More specifically, what high-level, regulatory genetic controls program a cell to express genes involved in lipid synthesis and accumulation?
- What regulates carbon flux from photosynthetic-source leaves to storage organs?
- What regulates metabolic partitioning of carbon among proteins, starch, and oil production?
- What species are most promising for conversion to oil accumulation? For instance, is oil potato a better candidate than oil beet or very high oil maize, or should plant stems be engineered to deposit extremely thick wax-rich cuticles?
- What are the opportunities for developing new high-yielding oilseed species tailored to energy production by uncoupling oil accumulation from seed carbohydrate or protein accumulation?
- What is the impact of high lipid levels on plant cells? Most cellular mechanisms operate in aqueous environments.

Cited References

- Albrecht, W. A. 1938. "Loss of Soil Organic Matter and Its Restoration," pp. 347–60 in *Soils and Men*, U.S. Department of Agriculture.
- Brown, G. R., et al. 2003. "Identification of Quantitative Trait Loci Influencing Wood Property Traits in Loblolly Pine (*Pinus taeda* L.) III. QTL Verification and Candidate Gene Mapping," *Genetics* **164**, 1537–46.
- Brunner, A. M., V. Busov, and S. H. Strauss. 2004. "The Poplar Genome Sequence: Functional Genomics in a Keystone Plant Species," *Trends Plant Sci.* **9**, 49–56.
- Busov, V. B., et al. 2005. "Insertional Mutagenesis in *Populus*: Relevance and Feasibility," *Tree Genet. Genomes* **1**, 135–42.
- Christiansen, P., et al. 2005. "A Rapid and Efficient Transformation Protocol for the Grass *Brachypodium distachyon*," *Plant Cell Rep.* **23**, 751–58.
- Draper, J., et al. 2001. "*Brachypodium distachyon*: A New Model System for Functional Genomics in Grasses," *Plant Phys.* **127**, 1539–55.
- EERE. 2003 (Review Draft). *Biomass Program Multi-Year Technical Plan*, U.S. DOE, Energy Efficiency and Renewable Energy (http://devafdc.nrel.gov/biogeneral/Program_Review/MYTP.pdf).
- Fitter, A. H., et al. 2005. "Biodiversity and Ecosystem Function in Soil," *Funct. Ecol.* **19** (3), 369–77.
- Follett, R. F. 2001. "Soil Management Concepts and Carbon Sequestration in Cropland Soils," *Soil Tillage Res.* **61**, 77–92.
- Groover, A., et al. 2004. "Gene and Enhancer Trap Tagging of Vascular-Expressed Genes in Poplar Trees," *Plant Physiol.* **134**, 1742–51.
- Handelsman, J., et al. 1998. "Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products," *Chem. Biol.* **5**, R245–49.
- Klaus, D., et al. 2004. "Increased Fatty Acid Production in Potato by Engineering of Acetyl-CoA Carboxylase," *Planta* **219**, 389–96.
- Koonin, S. E. 2006. "Getting Serious About Biofuels," *Science* **311**, 435.
- Maskina, M. S., et al. 1993. "Residual Effects on No-Till Crop Residues on Corn Yield and Nitrogen Uptake," *Soil Sci. Soc. Am. J.* **57**, 1555–60.
- Neale, D. B., and O. Savolainen. 2004. "Association Genetics of Complex Traits in Conifers," *Trends Plant Sci.* **9**, 325–30.
- Ogas, J., et al. 1997. "Gibberellin Mediates the Transition from Embryonic to Vegetative Cell Identity During Germination of the *Arabidopsis* pickle Mutant," *Science* **277**, 91–94.
- Perlack, R. D., et al. 2005. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*, DOE/GO-102005-2135, Oak Ridge National Laboratory, Oak Ridge, Tennessee (http://feedstockreview.ornl.gov/pdf/billion_ton_vision.pdf).

- Rafalski, A. 2002. "Applications of Single Nucleotide Polymorphisms in Crop Genetics," *Curr. Opin. Plant Biol.* **5**, 94–100.
- Remington, D. L., et al. 2001. "Structure of Linkage Disequilibrium and Phenotypic Associations in the Maize Genome," *Proc. Natl. Acad. Sci. USA* **98**, 11479–84.
- Sterky, F., et al. 2004. "A *Populus* Expressed Sequence Tag Resource for Plant Functional Genomics," *Proc. Natl. Acad. Sci. USA* (www.pnas.org/cig/reprint/101/38/13951).
- Tuskan, G. A., et al. In press 2006. "The Genome of Black Cottonwood, *Populus trichocarpa*," *Science*.
- Tuskan, G. A., S. P. DiFazio, and T. Teichmann. 2004. "Poplar Genomics is Getting Popular: The Impact of the Poplar Genome Project on Tree Research," *Plant Biol.* **6**, 2–4.
- Vogel J. P., et al. 2006. "Agrobacterium-Mediated Transformation and Inbred Line Development in the Model Grass *Brachypodium distachyon*," *Plant Cell Tiss. Org. Cult.* (www.springerlink.com, doi: 10.1007/s11240-005-9023-9).
- Wilhelm, W. W., et al. 2004. "Crop and Soil Productivity Response to Corn Residue Removal: A Literature Review," *Agron. J.* **96**, 1–17.

For Further Reading

- American Society of Plant Biologists. 2005. "President Bush Cites Importance of Plant-Based Biofuels," *ASPB News* **32** (4), 16–17.
- Bar-Joseph, Z., et al. 2003. "Computational Discovery of Gene Modules and Regulatory Networks," *Nat. Biotechnol.* **21**, 1337–42.
- Brown, G. R., et al. 2004. "Nucleotide Diversity and Linkage Disequilibrium in Candidate Genes Influencing Wood Formation in Loblolly Pine," *Proc. Natl. Acad. Sci.* **101**(42), 15255–260.
- Brummer, E. C. 1999. "Capturing Heterosis in Forage Crop Cultivar Development," *Crop Sci.* **39**, 943–54.
- Busov, V. B., et al. 2005. "Genetic Transformation: A Powerful Tool for Dissection of Adaptive Traits in Trees," *New Phytol.* **167**, 219–28.
- Cardon, L. R., and J. L. Bell. 2001. "Association Study Designs for Complex Diseases," *Nat. Rev. Genet.* **2**, 91–99.
- Del Grosso, S. J., et al. 2005. "DAYCENT Model Analysis of Past and Contemporary Soil N₂O and Net Greenhouse Gas Flux for Major Crops in the USA," *Soil Tillage Res.* **83**, 9–24.
- DiFazio, S. P. 2005. "A Pioneer Perspective on Adaptation," *New Phytol.* **165**, 661–64.
- Flint-Garcia, S. A., J. M. Thornsberry, and E. S. Buckler, IV. 2003. "Structure of Linkage Disequilibrium in Plants," *Annu. Rev. Plant Biol.* **54**, 357–74.

Follett, R. F., et al. 2005. "Research and Implementation Needs to Mitigate Greenhouse Gas Emissions from Agriculture in the USA," *Soil Tillage Res.* **83**, 159–66.

Follett, R. F., J. Z. Castellanos, and E. D. Buenger. 2005. "Carbon Dynamics and Sequestration in an Irrigated Vertisol in Central Mexico," *Soil Tillage Res.* **83**, 148–58.

Franzluebbers, A. J. 2005. "Soil Organic Carbon Sequestration and Agricultural Greenhouse Gas Emissions in the Southeastern USA," *Soil Tillage Res.* **83**, 120–47.

Franzluebbers, A. J., and R. F. Follett. 2005. "Greenhouse Gas Contributions and Mitigation Potential in Agricultural Regions of North America: Introduction," *Soil Tillage Res.* **83**, 1–8.

Gregorich, E. G., et al. 2005. "Greenhouse Gas Contributions of Agricultural Soils and Potential Mitigation Practices in Eastern Canada," *Soil Tillage Res.* **83**, 53–72.

Harbison, C. T., et al. 2004. "Transcriptional Regulatory Code of a Eukaryotic Genome," *Nature* **431**, 99–104.

Hooker, B. A., et al. 2005. "Long-Term Effects of Tillage and Corn Stalk Return on Soil Carbon Dynamics," *Soil Sci. Soc. Am. J.* **69**, 188–96.

Hu, W.-J., et al. 1999. "Repression of Lignin Biosynthesis Promotes Cellulose Accumulation and Growth in Transgenic Trees," *Nat. Biotechnol.* **17**, 808–12.

Huggins, D. R., et al. 1998. "Soil Organic C in the Tallgrass Prairie-Derived Region of the Cornbelt: Effects of Long-Term Crop Management," *Soil Tillage Res.* **47**, 219–34.

Jawson, M. D., et al. 2005. "GRACEnet: Greenhouse Gas Reduction Through Agricultural Carbon Enhancement Network," *Soil Tillage Res.* **83**, 167–72.

Johnson, J. M. F., et al. 2005. "Greenhouse Gas Contributions and Mitigation Potential of Agriculture in the Central USA," *Soil Tillage Res.* **83**, 73–94.

Johnson, J. M. F., R. R. Allmaras, and D. C. Reicosky. 2006. "Estimating Source Carbon from Crop Residues, Roots and Rhizodeposits Using the National Grain-Yield Database," *Agron. J.* (in press).

Lal, R. 2004. "Is Crop Residue a Waste?" *J. Soil Water Conserv.* **59**, 136–39.

Larson, W. E., et al. 1972. "Effects of Increasing Amounts of Organic Residues on Continuous Corn. II. Organic Carbon, Nitrogen, Phosphorus, and Sulfur," *Agron. J.* **64**, 204–8.

Liebig, M. A., et al. 2005. "Greenhouse Gas Contributions and Mitigation Potential of Agricultural Practices in Northwestern USA and Western Canada," *Soil Tillage Res.* **83**, 25–52.

Linden, D. R., C. E. Clapp, and R. H. Dowdy. 2000. "Long-Term Corn Grain and Stover Yields as a Function of Tillage and Residue Removal in East Central Minnesota," *Soil Tillage Res.* **56**, 167–74.

- Mann, L. K., V. R. Tolbert, and J. Cushman. 2002. "Potential Environmental Effects of Corn (*Zea mays* L.) Stover Removal with Emphasis on Soil Organic Matter and Erosion: A Review," *Agric. Ecosyst. Environ.* **89**, 149–66.
- Martens, D. A., et al. 2005. "Atmospheric Carbon Mitigation Potential of Agricultural Management in the Southwestern USA," *Soil Tillage Res.* **83**, 95–119.
- Mukherjee, S., et al. 2004. "Rapid Analysis of the DNA-Binding Specificities of Transcription Factors with DNA Microarrays," *Nat. Genet.* **36**, 1331–39.
- Nelson, R. G. 2002. "Resource Assessment and Removal Analysis for Corn Stover and Wheat Straw in the Eastern and Midwest United States—Rainfall and Wind-Induced Soil Erosion Methodology," *Biomass Bioenergy* **22**, 349–63.
- Nordborg, M., et al. 2002. "The Extent of Linkage Disequilibrium in *Arabidopsis thaliana*," *Nat. Genet.* **30**, 190–93.
- Strauss, S. H., et al. 2004. "Ten Lessons from 15 Years of Transgenic *Populus* Research," *Forestry* **77**, 455–65.
- Thornsberry, J. M., et al. 2001. "*Dwarf8* Polymorphisms Associate with Variation in Flowering Time," *Nat. Genet.* **28**, 286–89.
- Tong, A. H. Y., et al. 2004. "Global Mapping of the Yeast Genetic Interaction Network," *Science* **303**, 808–13.
- U.S. DOE Office of Science. 2005. "Making Bioethanol Cost Competitive," DOE Genomics:GTL Program, Systems Biology for Energy and Environment (www.doe.genome.gov/pubs/Bioethanol10_27_05_8.5.pdf).
- Venter, J. C., et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea," *Science* **304**, 66–74.
- Vogel, K. P. 2004. "Switchgrass," pp. 561–88 in *Warm-Season (C4) Grasses*, ed. L. E. Moser, L. Sollenberger, and B. Burson, ASA-CSSA-SSSA Monograph No. 45, Madison, Wis.
- Vogel, K. P. 2000. "Improving Warm-Season Grasses Using Selection, Breeding, and Biotechnology," pp. 83–106 in *Native Warm-Season Grasses: Research Trends and Issues*, ed. K. J. Moore and B. Anderson, Crop Science Special Publication No. 30, Crop Science Society of America and American Society of Agronomy, Madison, Wis.
- Vogel, K. P., and B. Burson. 2004. "Breeding and Genetics," pp. 51–96 in *Warm-Season (C4) Grasses*, ed. L. E. Moser, L. Sollenberger, and B. Burson, ASA-CSSA-SSSA Monograph No. 45, Madison, Wis.
- Vogel, K. P., and H. G. Jung. 2001. "Genetic Modification of Herbaceous Plants for Feed and Fuel," *Crit. Rev. Plant Sci.* **20**, 15–49.
- Walsh, M. E., et al. 2003. "Bioenergy Crop Production in the United States," *Environ. Resour. Econ.* **24**(4), 313–33.