

BIOMASS MODIFICATION, CHARACTERIZATION AND PROCESS MONITORING ANALYTICS TO SUPPORT BIOFUEL AND BIOMATERIAL PRODUCTION

EDITED BY : Robert Henry, Blake Simmons and Jason Lupoi

PUBLISHED IN: Frontiers in Bioengineering and Biotechnology and
Frontiers in Energy Research





frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-867-2

DOI 10.3389/978-2-88919-867-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

BIOMASS MODIFICATION, CHARACTERIZATION AND PROCESS MONITORING ANALYTICS TO SUPPORT BIOFUEL AND BIOMATERIAL PRODUCTION

Topic Editors:

Robert Henry, The University of Queensland, Australia

Blake Simmons, Joint BioEnergy Institute, USA

Jason Lupoi, Joint BioEnergy Institute, USA & University of Queensland, Australia



Sugarcane a lignocellulosic biomass feedstock

Image by Robert Henry

The conversion of lignocellulosic biomass into renewable fuels and other commodities has provided an appealing alternative towards supplanting global dependence on fossil fuels. The suitability of multitudes of plants for deconstruction to useful precursor molecules and products is currently being evaluated. These studies have probed a variety of phenotypic traits, including cellulose, non-cellulosic polysaccharide, lignin, and lignin monomer composition, glucose and xylose production following enzymatic hydrolysis, and an assessment of lignin-carbohydrate and lignin-lignin linkages, to name a few. These quintessential traits can provide an assessment of biomass recalcitrance, enabling researchers to devise appropriate deconstruction strategies. Plants with high polysaccharide and lower lignin contents have been shown to breakdown to monomeric sugars more readily. Not all plants contain ideal proportions of the various cell wall constituents, however. The capabilities of

biotechnology can alleviate this conundrum by tailoring the chemical composition of plants to be more favorable for conversion to sugars, fuels, etc. Increases in the total biomass yield, cellulose content, or conversion efficiency through, for example, a reduction in lignin content, are pathways being evaluated to genetically improve plants for use in manufacturing biofuels and bio-based chemicals. Although plants have been previously domesticated for food and

fiber production, the collection of phenotypic traits prerequisite for biofuel production may necessitate new genetic breeding schemes. Given the plethora of potential plants available for exploration, rapid analytical methods are needed to more efficiently screen through the bulk of samples to hone in on which feedstocks contain the desired chemistry for subsequent conversion to valuable, renewable commodities.

The standard methods for analyzing biomass and related intermediates and finished products are laborious, potentially toxic, and/or destructive. They may also necessitate a complex data analysis, significantly increasing the experimental time and add unwanted delays in process monitoring, where delays can incur in significant costs. Advances in thermochemical and spectroscopic techniques have enabled the screening of thousands of plants for different phenotypes, such as cell-wall cellulose, non-cellulosic polysaccharide, and lignin composition, lignin monomer composition, or monomeric sugar release. Some instrumental methods have been coupled with multivariate analysis, providing elegant chemometric predictive models enabling the accelerated identification of potential feedstocks. In addition to the use of high-throughput analytical methods for the characterization of feedstocks based on phenotypic metrics, rapid instrumental techniques have been developed for the real-time monitoring of diverse processes, such as the efficacy of a specific pretreatment strategy, or the formation of end products, such as biofuels and biomaterials. Real-time process monitoring techniques are needed for all stages of the feedstocks-to-biofuels conversion process in order to maximize efficiency and lower costs by monitoring and optimizing performance. These approaches allow researchers to adjust experimental conditions during, rather than at the conclusion, of a process, thereby decreasing overhead expenses.

This Frontiers Research Topic explores options for the modification of biomass composition and the conversion of these feedstocks into to biofuels or biomaterials and the related innovations in methods for the analysis of the composition of plant biomass, and advances in assessing up- and downstream processes in real-time. Finally, a review of the computational models available for techno-economic modeling and lifecycle analysis will be presented.

Citation: Henry, R., Simmons, B., Lupoi, J., eds. (2016). Biomass Modification, Characterization and Process Monitoring Analytics to Support Biofuel and Biomaterial Production. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-867-2

Table of Contents

- 06 Editorial: Biomass Modification, Characterization, and Process Monitoring Analytics to Support Biofuel and Biomaterial Production**
Jason Lupoi, Blake Simmons and Robert Henry
- 08 Efficient Eucalypt Cell Wall Deconstruction and Conversion for Sustainable Lignocellulosic Biofuels**
Adam L. Healey, David J. Lee, Agnelo Furtado, Blake A. Simmons and Robert J. Henry
- 22 Optimization of alkaline and dilute acid pretreatment of agave bagasse by response surface methodology**
Abimael I. Ávila-Lara, Jesus N. Camberos-Flores, Jorge A. Mendoza-Pérez, Sarah R. Messina-Fernández, Claudia E. Saldaña-Duran, Edgar I. Jimenez-Ruiz, Leticia M. Sánchez-Herrera and Jose A. Pérez-Pimienta
- 32 Evaluating lignocellulosic biomass, its derivatives, and downstream products with Raman spectroscopy**
Jason S. Lupoi, Erica Gjersing and Mark F. Davis
- 50 Standard flow liquid chromatography for shotgun proteomics in bioenergy research**
Susana M. González Fernández-Niño, A. Michelle Smith-Moritz, Leanne Jade G. Chan, Paul D. Adams, Joshua L. Heazlewood and Christopher J. Petzold
- 57 Development of a high throughput platform for screening glycoside hydrolases based on Oxime-NIMS**
Kai Deng, Joel M. Guenther, Jian Gao, Benjamin P. Bowen, Huu Tran, Vimalier Reyes-Ortiz, Xiaoliang Cheng, Noppadon Sathitsuksanoh, Richard Heins, Taichi E. Takasuka, Lai F. Bergeman, Henrik Geertz-Hansen, Samuel Deutsch, Dominique Loqué, Kenneth L. Sale, Blake A. Simmons, Paul D. Adams, Anup K. Singh, Brian G. Fox and Trent R. Northen
- 67 Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases**
Kai Deng, Taichi E. Takasuka, Christopher M. Bianchetti, Lai F. Bergeman, Paul D. Adams, Trent R. Northen and Brian G. Fox
- 77 Immunological approaches to biomass characterization and utilization**
Sivakumar Pattathil, Utku Avci, Tiantian Zhang, Claudia L. Cardenas and Michael G. Hahn
- 91 Relationships between biomass composition and liquid products formed via pyrolysis**
Fan Lin, Christopher L. Waters, Richard G. Mallinson, Lance L. Lobban and Laura E. Bartley

- 106 Potential for Genetic Improvement of Sugarcane as a Source of Biomass for Biofuels**
Nam V. Hoang, Agnelo Furtado, Frederik C. Botha, Blake A. Simmons and Robert J. Henry
- 121 Identification and molecular characterization of the switchgrass AP2/ERF transcription factor superfamily, and overexpression of PvERF001 for improvement of biomass characteristics for biofuel**
Wegi A. Wuddineh, Mitra Mazarei, Geoffrey B. Turner, Robert W. Sykes, Stephen R. Decker, Mark F. Davis and C. Neal Stewart Jr.
- 142 Phenotypic Changes in Transgenic Tobacco Plants Overexpressing Vacuole-Targeted *Thermotoga maritima* BglB Related to Elevated Levels of Liberated Hormones**
Quynh Anh Nguyen, Dae-Seok Lee, Jakyun Jung and Hyeun-Jong Bae
- 154 Book review: Socio-economic impacts of bioenergy production**
Sheikh Adil Edrisi and P. C. Abhilash



Editorial: Biomass Modification, Characterization, and Process Monitoring Analytics to Support Biofuel and Biomaterial Production

Jason Lupoi^{1,2,3}, Blake Simmons^{1,2,4} and Robert Henry^{2,3*}

¹ Joint BioEnergy Institute, Emeryville, CA, USA, ² University of Queensland, Brisbane, QLD, Australia, ³ Queensland Alliance for Agriculture and Food Innovation, Brisbane, QLD, Australia, ⁴ Sandia National Laboratory, Livermore, CA, USA

Keywords: high-throughput screening assays, genetic manipulation, glycoside hydrolases, Raman spectroscopy, NIMS, cell-wall degrading enzymes, biomass pretreatment and fractionation

The Editorial on the Research Topic

Biomass Modification, Characterization, and Process Monitoring Analytics to Support Biofuel and Biomaterial production

OPEN ACCESS

Edited by:

Uwe Schröder,
Technische Universität Braunschweig,
Germany

Reviewed by:

Falk Harnisch,
UFZ – Helmholtz-Centre for
Environmental Research, Germany

*Correspondence:

Robert Henry
robert.henry@uq.edu.au

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 08 January 2016

Accepted: 22 February 2016

Published: 10 March 2016

Citation:

Lupoi J, Simmons B and Henry R
(2016) Editorial: Biomass
Modification, Characterization, and
Process Monitoring Analytics
to Support Biofuel and
Biomaterial Production.
Front. Bioeng. Biotechnol. 4:25.
doi: 10.3389/fbioe.2016.00025

This Frontiers Research Topic journeys through various challenges facing researchers seeking to develop fuels and products derived from lignocellulosic biomass. These challenges include: the rapid quantification of plant cell wall chemistry, enabling yields of potential monomeric sugars to be assessed, identification of plants possessing ideal trait that can be brought to the forefront of research efforts; once the native plant chemistry is known, how can yields be improved by chemically or genetically altering plant cell walls to reduce recalcitrance; does genetic modification of plants to increase accessibility to saccharification enzymes hinder the plant's growth and/or function; are the innovative methods identified by researchers cost-effective and scalable to a commercial level? These topics are a sampling of the obstacles researchers combat when nominating a specific plant for downstream applications or implementing new deconstruction, genetic, or measurement strategies. How well can different plants can be broken down into useful, downstream precursor molecules? Efforts to develop a lucid picture of the chemical composition of abundant, diverse plants being explored as potential starting feedstocks has resulted in the evolution of high-throughput techniques that permit many more samples to be screened in much shorter periods of time. Advances in thermochemical and spectroscopic techniques have enabled the screening of thousands of plants for different phenotypes, such as cell-wall composition and monomeric sugar release. Some instrumental methods have been coupled with multivariate analysis, providing elegant chemometric predictive models enabling the accelerated identification of potential feedstocks. Rapid instrumental techniques have been developed for real-time monitoring of diverse processes, such as the efficacy of specific pretreatment strategies, or downstream products, such as biofuels and biomaterials. Real-time process monitoring techniques are needed for all stages of the feedstocks-to-biofuels conversion process to maximize efficiency and lower costs by monitoring and optimizing performance. These approaches allow researchers to adjust experimental conditions during, rather than at the conclusion, of processes, thereby decreasing overhead expenses.

The article in this book entitled “Evaluating Lignocellulosic Biomass, Its Derivatives, and Downstream Products with Raman Spectroscopy” illustrates how advances in Raman instrumentation, and the use of multivariate analysis has aided researchers in developing efficient, rapid methods to quickly evaluate diverse varieties of plants. Another hot research topic is how can we make better, more efficient enzymes beyond what may be naturally available? Just as high-throughput and

automated methods are desirable to partition a diverse portfolio of potential feedstocks, gaging enzymatic performance and efficiency presents similar challenges that can be alleviated by employing techniques such as that illustrated in the two articles “Use of Nanostructure-Initiator Mass Spectrometry to Deduce Selectivity of Reaction in Glycoside Hydrolases” and “Development of a High Throughput Platform for Screening Glycoside Hydrolases Based on Oxime-NIMS”. Studies like these have probed a variety of phenotypic traits, including the assessment of dominant plant cell wall chemical constituents (glucan, xylan, lignin, etc.), and the quantification of products from enzymatically deconstructing the plants into simpler, molecular building blocks. “Immunological Approaches to Biomass Characterization and Utilization” reviews recent advances in the use of glycan-directed probes to evaluate plant cell walls, including glycan-directed monoclonal antibodies, in efforts to understand cell-wall composition and structure, and how manipulations to these affect the biosynthetic processes. Assessing the study and function of proteins, called proteomics, is typically throughput limited. “Standard Flow Liquid Chromatography for Shotgun Proteomics in Bioenergy Research” validates the standard flow technique, and shows how 800–1000 different proteins could be identified in complex samples. Pyrolysis of plants to form bio-oil presents an interesting option for the synthesis of fossil fuel replacements. Before these oils can be utilized, they must be fully characterized. How will the intrinsic biomass chemistry affect the resultant bio-oil? The article entitled “Relationships between Biomass Composition and Liquid Products formed via Pyrolysis” explores this topic.

These important traits can aid in gaging biomass recalcitrance, enabling researchers to devise appropriate deconstruction strategies. The article “Optimization of Alkaline and Dilute Acid Pretreatment of Agave Bagasse by Response Surface Methodology” describes a technique for optimizing the deconstruction of agave cell walls to maximize the total reducing sugars produced during saccharification. Various techniques used to combat biomass recalcitrance in eucalypts have been outlined in the article “Efficient Eucalypt Cell Wall Deconstruction and Conversion for Sustainable Lignocellulosic Biofuels”. Plants with high polysaccharide and lower lignin contents are known to breakdown to monomeric sugars more readily, but what happens if a plant does not naturally possess the ideal traits for biofuel production, like high cellulose and low lignin? Researchers can purposely tailor

the plant’s chemical composition to be more favorable for conversion to sugars, fuels, or other value-added products. In “Potential for Genetic Improvement of Sugarcane as a Source of Biomass for Biofuels”, different strategies for improving the degradability of sugarcane are reviewed. Increases in the total biomass yield, cellulose content, or saccharification efficiency through, for example, reductions in lignin content, are pathways being evaluated to genetically improve plants as feedstocks for biofuels and bio-based chemicals. The biological pathways with which the plant cell wall is fabricated can shed light on potential genetic manipulations that can favorably alter the cell wall chemistry. The articles “Phenotypic Changes in Transgenic Tobacco Plants Overexpressing Vacuole-Targeted *Thermotoga maritima* BglB Related to Elevated Levels of Liberated Hormones” and “Identification and Molecular Characterization of the Switchgrass AP2/ERF Transcription Factor Superfamily, and Overexpression of PvERF001 for Improvement of Biomass Characteristics for Biofuel” illustrate this technique. Although plants have been previously domesticated for food and fiber production, the collection of phenotypic traits prerequisite for biofuel production may necessitate new genetic breeding schemes.

Lastly, the economic and social ramifications of using lignocellulosic biomass, such as the food versus fuel conflict, have been well documented. Can marginal lands provide us with the land required to achieve our energy needs in a cost-effective way? Cost-effective enough that we can seriously consider supplanting some fossil fuel usage on a much larger scale? This Research Topic includes a book review “Book Review: Socio-Economic Impacts of Bioenergy Production” of the recent Springer publication *Socio-economic impacts of bioenergy production*.

This collection of papers demonstrates how advances at multiple levels are likely to contribute to the successful industrial scale production of biofuels and biomaterials. Innovations in biomass composition have a major role to play in facilitating conversion and appropriate analytical methodologies are fundamental to managing the industrial process and guiding research and development.

AUTHOR CONTRIBUTIONS

JL wrote the editorial for this research topic. BS and RH critically read and edited the document.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Lupoi, Simmons and Henry. This is an open-access article distributed under the terms of the Creative Commons Attribution License

(CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Efficient Eucalypt Cell Wall Deconstruction and Conversion for Sustainable Lignocellulosic Biofuels

Adam L. Healey^{1*}, David J. Lee^{2,3}, Agnelo Furtado¹, Blake A. Simmons^{1,4,5} and Robert J. Henry¹

¹ Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St. Lucia, QLD, Australia, ² Forest Industries Research Centre, University of the Sunshine Coast, Maroochydore, QLD, Australia, ³ Department of Agriculture and Fisheries, Forestry and Biosciences, Agri-Science Queensland, Gympie, QLD, Australia, ⁴ Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA, ⁵ Biological and Engineering Sciences Center, Sandia National Laboratories, Livermore, CA, USA

OPEN ACCESS

Edited by:

Subba Rao Chaganti,
University of Windsor, Canada

Reviewed by:

Chiranjeevi Thulluri,
Jawaharlal Nehru Technological
University Hyderabad, India
Maria Carolina Quecine,
University of São Paulo, Brazil
Brahmaiah Pendyala,
University of Toledo, USA

*Correspondence:

Adam L. Healey
a.healey1@uq.edu.au

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 15 August 2015

Accepted: 04 November 2015

Published: 20 November 2015

Citation:

Healey AL, Lee DJ, Furtado A,
Simmons BA and Henry RJ (2015)
Efficient Eucalypt Cell Wall
Deconstruction and Conversion for
Sustainable Lignocellulosic Biofuels.
Front. Bioeng. Biotechnol. 3:190.
doi: 10.3389/fbioe.2015.00190

In order to meet the world's growing energy demand and reduce the impact of greenhouse gas emissions resulting from fossil fuel combustion, renewable plant-based feedstocks for biofuel production must be considered. The first-generation biofuels, derived from starches of edible feedstocks, such as corn, create competition between food and fuel resources, both for the crop itself and the land on which it is grown. As such, biofuel synthesized from non-edible plant biomass (lignocellulose) generated on marginal agricultural land will help to alleviate this competition. Eucalypts, the broadly defined taxa encompassing over 900 species of *Eucalyptus*, *Corymbia*, and *Angophora* are the most widely planted hardwood tree in the world, harvested mainly for timber, pulp and paper, and biomaterial products. More recently, due to their exceptional growth rate and amenability to grow under a wide range of environmental conditions, eucalypts are a leading option for the development of a sustainable lignocellulosic biofuels. However, efficient conversion of woody biomass into fermentable monomeric sugars is largely dependent on pretreatment of the cell wall, whose formation and complexity lend itself toward natural recalcitrance against its efficient deconstruction. A greater understanding of this complexity within the context of various pretreatments will allow the design of new and effective deconstruction processes for bioenergy production. In this review, we present the various pretreatment options for eucalypts, including research into understanding structure and formation of the eucalypt cell wall.

Keywords: eucalypts, biotechnology, pretreatment, lignocellulosic biofuel, bioenergy

INTRODUCTION

Currently, approximately 40% of the world's transportation fuels (fossil fuels) are derived from non-renewable sources, the combustion of which directly contributes to global climate change (Simmons et al., 2008; González-García et al., 2012). As such, renewable plant-based feedstocks for fuel synthesis, aptly referred to as "biofuels," are under consideration to alleviate these concerns. The first generation of feedstocks used for biofuel synthesis was mainly derived from sugarcane and corn, as their energy storage polysaccharides are readily available and easily hydrolyzed into

monosaccharides for microbial fermentation. However, as these feedstocks are important links within the human food chain, generation of biofuel from these crops creates a direct competition for resources. Furthermore, in the USA alone, the maximum biofuel yield from the first-generation biofuel feedstocks is roughly 30% of the renewable fuel target (Perlack et al., 2005), creating a large gap that must be filled with alternatives. Plant cell wall structural polysaccharides, although more complex than starch molecules, represent the most abundant biopolymers in the world, containing large stores of carbon for conversion into liquid fuels, such as ethanol and butanol (Wyman, 1999). As structural polysaccharides represent the non-edible portions of plants, fuel synthesized from cellulose and hemicellulose can help alleviate the competition between energy and agriculture. Crops intended for this purpose are known as the second-generation biofuel feedstocks.

There are numerous advantages to using the second-generation feedstocks as a source of renewable energy. Combustion of fossil fuels adds carbon dioxide to the atmosphere, the main contributor to the greenhouse effect and subsequent climate change. Biofuel crops help to mitigate the effect of CO₂ by sequestering more carbon within their biomass than is released during biofuel combustion, thus creating a net reduction in CO₂ levels (Rubin, 2008; Shepherd et al., 2011; Soccol et al., 2011). High production grassy species, such as those belonging to the *Miscanthus* and *Saccharum* genera, are high-value bioenergy crops due to their exceptional growth rate and desirable biomass composition that is relatively easy to deconstruct for polysaccharides using mild pretreatments (Rubin, 2008). However, high production crops, such as these require nutrient rich soils, normally reserved for intensive agriculture. This indirect competition for land and soil between food or fuel crops can be avoided through the cultivation of feedstocks that grow well on marginal land, of which there is approximately 1.4 billion hectares available globally (Carroll and Somerville, 2009; Somerville et al., 2010). Fuel production from woody (or lignocellulosic) biomass also offers several advantages over grassy biomass. Growing trees for energy production allows biomass to be “stored on the stump” to be harvested when needed (Shepherd et al., 2011), a luxury not afforded by grasses, which must be harvested at particular times during the year and must be processed immediately before fungal degradation begins. Also, woody biomass can be transported to processing facilities more economically, as it more energy dense than grassy biomass which requires greater amounts of fuel to move the biomass than can be generated from its fibers (Kaylen et al., 2000; Somerville et al., 2010).

Eucalypts, a native Australian taxon that includes genera *Eucalyptus*, *Corymbia*, and *Angophora*, are an attractive prospective biofuel crop, being the most widely planted hardwood trees in the world (Myburg et al., 2007; Grattapaglia and Kirst, 2008). Having adapted to the terrestrial environment of Australia, eucalypts are well suited for plantations in a wide variety of climates, soil types, and rainfall conditions (Ladiges et al., 2003; Myburg et al., 2007; Grattapaglia and Kirst, 2008). They are grown commercially in over 100 countries with well-established silviculture practices already in place, such as clonal propagation, allowing plantations to achieve high rates of productivity, up to 25 dry

tonnes/hectare/year (Stricker et al., 2000; Rockwood et al., 2008), more than double the required productivity rate estimated by the US Department of Energy for a long-term renewable energy crop (Hinchee et al., 2009). Furthermore, many eucalypt species also regenerate shoots after harvesting, which ensures ease of management by potentially eliminating the need for re-planting (Shepherd et al., 2011).

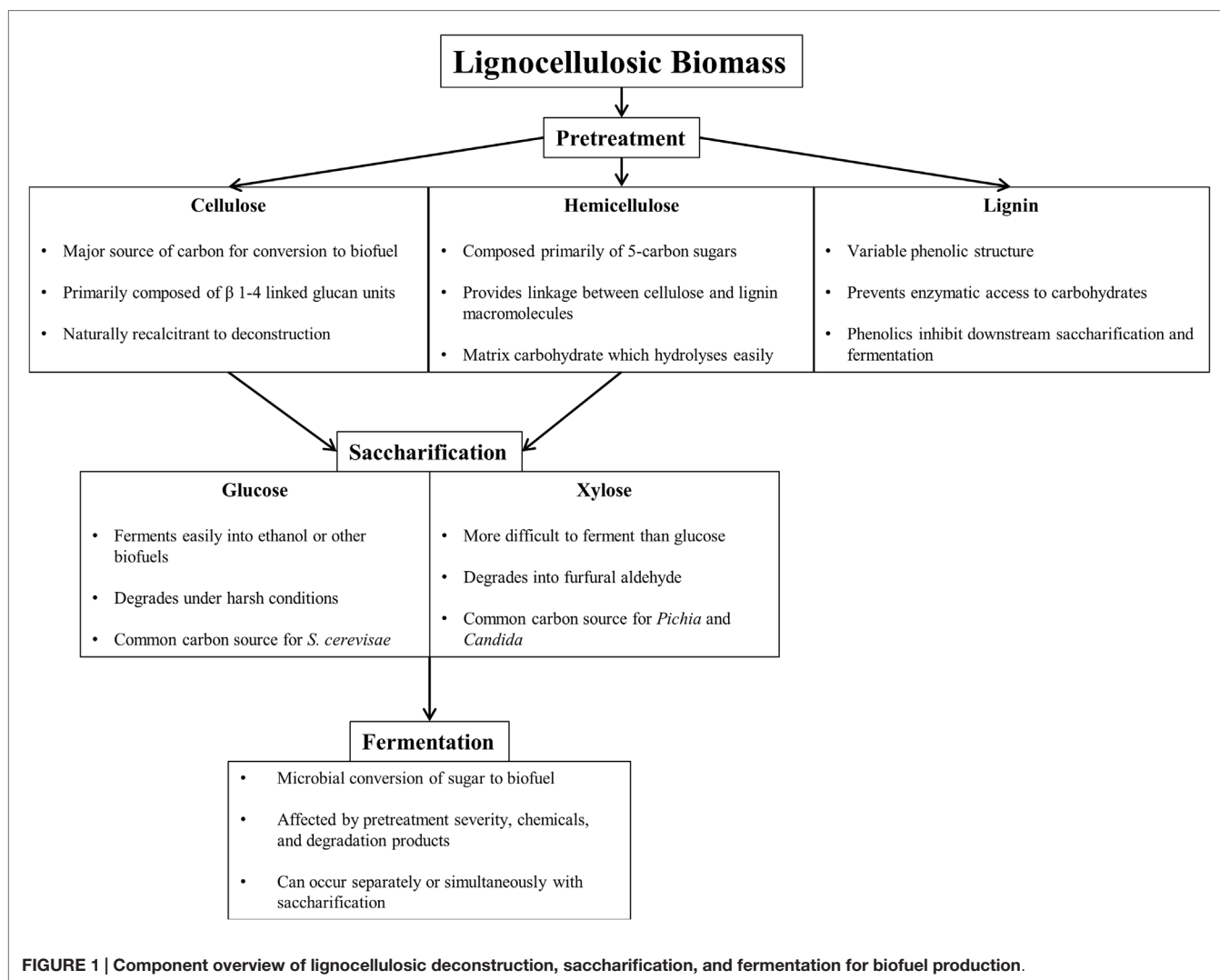
Eucalypts, due to differences in flowering times (protandry) and self-incompatibility, are predominately out-crossing species which maintains high levels of heterozygosity in their genomes and encourages genetic diversity and phenotypic variation (Horsley and Johnson, 2007; Grattapaglia and Kirst, 2008). This variation is exploited by breeders through selection and combination of desirable traits for industrial application, such as controlled-cross hybrids that combine the high cellulose and fiber content of *Eucalyptus globulus* with the growth rate and form of *E. grandis* (Poke et al., 2005; Grattapaglia, 2008). Phenotypic traits that are desirable for efficient biofuel production are closely aligned with those sought by the pulp and paper industry. High-quality wood pulp is primarily composed of cellulosic fibers, which upon enzymatic hydrolysis releases monomeric glucose subunits, which serve as the main substrate for microbial fermentation and conversion to liquid fuel (Hisano et al., 2009; Wegrzyn et al., 2010).

Despite the advantages of lignocellulosic biofuel crops, woody biomass conversion into a source of renewable energy is hindered through its natural complexity and recalcitrance to deconstruction (Ramos and Saddler, 1994; Blanch et al., 2011). Many of the options for deconstruction require harsh and expensive chemicals (such as acids and alkalis) or energy intensive methods, such as grinding and ball milling. An increased understanding of eucalypt biomass will allow the engineering of more cost-effective pretreatments that can increase fuel production efficiency, while lessening the formation and impact of inhibitory compounds produced during conversion. In this review, we present an overview of the major contributing components of eucalypt cell wall recalcitrance, and the current research surrounding eucalypt biomass pretreatment for fuel production.

CHALLENGES TO LIGNOCELLULOSIC BIOFUEL CONVERSION

Efficient conversion of lignocellulosic biomass to biofuel requires pretreatment, saccharification, and fermentation, each presenting unique challenges (**Figure 1**). Pretreatment breaks down and separates each of the major components of biomass (cellulose, hemicellulose, and lignin), either through mechanical or through chemical means to reduce cellulose crystallinity, increase surface area, and remove lignin, the largest barrier to efficient enzymatic saccharification (Furtado et al., 2014). **Table 1** summarizes common lignocellulose pretreatments and highlights pros and cons of each process.

Despite the low cost of producing lignocellulosic biomass, the economic cost of producing biofuel remains high (Lange, 2007). Pretreatment, a required process for increasing saccharification is costly, requiring large amounts of energy or expensive chemicals



(e.g., sulfuric acid) to promote enzymatic access to polysaccharides. Fermentation also represents a significant cost to biofuel production as the production of enzymes is expensive, and the efficiency at which microorganisms can convert sugars into fuel is dependent on pretreatment (Hamelinck et al., 2005). Therefore, harsh pretreatments that are used for biomass deconstruction in other industrial processes (e.g., pulping) may not be appropriate for biofuel production. There are also significant operational costs associated with biofuel conversion, including capital costs, labor, and waste water processing. As such, the development of simple, cost-effective, and environmentally safe pretreatments is critical for large-scale sustainable production. Given that pretreatment and fermentation represent the highest costs of producing biofuel, feedstock selection is also critical for fuel production as well.

The simplest pretreatment option is grinding and milling of biomass to increase reactive surface area for hydrolysis. However, the energy required to generate small enough particles is often too high to be a cost-effective option (Zheng et al., 2009; Talebnia et al., 2010). A more common pretreatment is acid hydrolysis, where strong acids (e.g., H_2SO_4) solubilize the hemicellulose

polysaccharide matrix, leaving behind cellulose and lignin (Galbe and Zacchi, 2007). Although effective, acid pretreatment generates compounds that inhibit downstream biomass conversion processes through reduction of microbial growth and enzymatic release (Jönsson et al., 2013). For instance, while the majority of lignin present in the cell wall is acid insoluble (Klason lignin), upon pretreatment, a small portion hydrolyzes releasing phenolics, such as vanillin, trans-cinnamic acid, and 4-hydrobenzoic acid (Palmqvist and Hahn-Hägerdal, 2000; Ximenes et al., 2010). Additionally, monomeric subunits of cellulose and hemicellulose degrade in low pH conditions, generating aldehydes [furfural and 5-hydroxymethyl-2-furaldehyde (5-HMF)] and organic acids. Formation of these degradation products inhibits fermentation by reducing available sugars and limiting microbial growth (Zheng et al., 2009; Socol et al., 2011; Puri et al., 2012). Similarly, organosolv pretreatment combines an organic solvent (e.g., ethanol) with an inorganic acid catalyst (e.g., sulfuric acid) to destroy internal lignin and hemicellulose bonds, resulting in effective recovery of high-quality cellulose and lignin portions of biomass. Although an effective pretreatment for both hardwood and

TABLE 1 | Summary and assessment of common pretreatment options for lignocellulose.

Pretreatment	Summary	Pros	Cons
Grinding and milling	Mechanical disruption of biomass to increase surface area	No chemicals required No degradation products generated	Energy inefficient Lignin structure remains
Concentrated acid	Relatively complete hydrolysis of biomass with hydrochloric or sulfuric acid	Complete biomass hydrolysis Low inhibitory product formation under low temperature conditions	High cost and loss of acid High environmental impact Phenolic release Inhibition of fermentation
Dilute acid	Combination of acid and high temperature to solubilize hemicellulose	Low acid concentrations required (<1%) Short reaction times	Sugar degradation and loss Release of phenolics
Alkaline	Cleaves linkages within lignin and between hemicellulose and lignin	Swells biomass Established pulping practice Works with various feedstocks Low temperature, low pressure reaction	High environmental impact Low recovery Requires neutralization
Organosolv	Aqueous/organic solvent at high temperatures break hemicellulose–lignin bonds	Allows intact lignin recovery Works well across various feedstocks	Organic solvents are expensive and inhibit fermentation High temperatures (250°C) required
Steam explosion	Biomass explosion of biomass by high temperature/pressure coupled with rapid decompression	Solubilization of hemicellulose and reduced cellulose crystallinity Short reaction time	High temperatures generate inhibitory products
Autohydrolysis	Pressurized, high temperature water solubilizes hemicellulose with <i>in situ</i> acids	No chemicals needed Low environmental impact	Requires a low lignin feedstock to be efficient High temperature and pressure required
Ionic liquids	Room-temperature organic liquid salts dissolve biomass	Selective precipitation of cellulose Lignin recovery Stable, low volatility chemicals Works well regardless of varying wood properties	High cost of chemicals Inhibition of microbial fermentation

Hendriks and Zeeman (2009), Alvira et al. (2010), and Blanch et al. (2011).

softwood biomass, downstream ethanol production still suffers from the formation of inhibitory products (Sun and Cheng, 2002; Zhu and Pan, 2010).

Alkaline pretreatment, which employs chemicals, such as sodium hydroxide, lime and hydrazine, to disrupt the linkage between hemicellulose and lignin, reduces the formation of inhibitory products but nonetheless remains an expensive option that is dependent on lignin content which determines its efficacy (Blanch et al., 2011). An alternative method, which seeks to work universally well regardless of biomass composition, is ionic liquid (IL) pretreatment. ILs are non-volatile, stable compounds that solubilize lignocellulosic biomass, allowing selective precipitation of components for easy recovery. Once dissolved, cellulose precipitates from solution upon addition of an antisolvent (e.g., water or ethanol) while lignin and other solutes remain intact (Zhu et al., 2006; Singh et al., 2009).

CELLULOSE CRYSTALLINITY

Cellulose, the most abundant biopolymer on earth, is composed of thousands of glucose monomers linked together by β 1–4 glycosidic bonds. Its function within the cell wall is to provide strength and rigidity, while remaining flexible during cell expansion and growth (Mutwil et al., 2008; Mansfield, 2009). Sucrose, generated through photosynthesis, supplies the glucose molecule required for cellulose synthesis, which is phosphorylated by hexokinase, and is incorporated into growing cellulose

microfibrils by cellulose synthase (CESA) enzymes (Somerville, 2006; Joshi and Mansfield, 2007; Mohnen et al., 2008). During synthesis, each cellulose microfibril associates with other glucan chain through extensive hydrogen bonding and Van der Waals forces, creating a highly compact polysaccharide. Within the cell wall, cellulose exists in primarily two forms, a highly ordered crystalline structure that lacks surface area and a less ordered, amorphous type (Harris and DeBolt, 2010). The highly compact crystalline structure lends itself toward the natural recalcitrance of woody biomass to deconstruction, as it prevents cellulase enzymes to accessing microfibrils, thus inhibiting efficient saccharification (Mosier et al., 2005; Hall et al., 2010).

Crystalline cellulose formation in eucalypts has been traditionally researched through the formation of tension wood. Tension wood, characterized by the formation a gelatinous layer of crystalline cellulose (G-layer), serves to re-direct a growing stem upwards in response to gravitational stress (Jourez et al., 2001). As tension wood can be artificially induced, Paux et al. (2005) investigated tension wood formation in *E. globulus* by tying the growing stems of 2-year-old trees to the adjacent tree, bending their trunks to a 45° angle. By extracting RNA from the xylem of the bent trees on either side of the bend (tension wood and opposite wood) at various timepoints (0, 6, 24, and 168 h), the authors were able to identify differentially expressed genes during cellulose formation using a xylem complementary DNA (cDNA) array. As evidenced by a much larger bent-stem experiment performed in *Eucalyptus nitens* with 4,900 xylem cDNAs, Qiu

et al. (2008) found tension wood, although lacking the characteristic “G-layer,” contained high concentrations of cellulose and low amounts of Klason lignin. Additionally, X-ray diffraction of upper and lower bent stems revealed that the cellulose microfibril angle (MFA) on the upper branch was much less than that of the lower branch. MFA, the angle at which cellulose polymers are synthesized within the cell wall affects their tendency to form hydrogen bonds. MFA, which affects wood stiffness (Schimleck et al., 2001), is an indirect biofuel trait as cellulose content negatively correlates with MFA and lignin content (Plomion et al., 2001). Qiu et al. (2008) also found that in tension wood, the highest expression profiles belonged to β -tubulin genes and fasciclin-like arabinogalactan (FLA) proteins. β -Tubulin proteins are responsible for transporting cellulose synthesis machinery to the plasma membrane, which may in-turn affect MFA. FLA genes, known to associate pectic side-chains and other structural polysaccharides also affect MFA, as demonstrated through transformation of *E. nitens* with FLA3, identified from the *E. grandis* genome (Macmillan et al., 2015).

To investigate the effect of tension and opposite wood on saccharification and fermentation, Muñoz et al. (2011) treated *E. globulus* biomass to organosolv (ethanol/water) pretreatment, followed by simultaneous saccharification and fermentation (SSF) (discussed later). The authors found that tension wood (as compared to opposite wood) contained similar glucan content (46–47%), higher xylan amounts (16.0 and 12.0%, respectively), and lower lignin content (22.1 and 26.1%, respectively). Upon pretreatment, remaining residual lignin was lower in tension wood and required less time and cooking (as expressed by H factor, a single variable calculated from the combination of cooking temperature and time) for delignification. Pulp from tension and opposite wood were assayed for glucose conversion by enzymatic hydrolysis, finding that despite similar or higher lignin content, glucan to glucose conversion was more efficient in opposite wood. However, investigation into pulp viscosity showed that tension wood glucans were of higher molecular mass, which may have influenced their rate of conversion. Upon submission of pulps from tension and opposite wood for SSF, the authors found that harsh pretreatment conditions (H factor – 12,500) outperformed milder conditions (H factor – 3,900) to produce 35 and 30 g/L of ethanol, respectively. Considering the maximum theoretical conversion of ethanol from glucose is 51%, these concentrations represent 95 and 85% conversion efficiency, which scales to a yield of 290 L of ethanol/tonne of biomass. Considering the formation of tension wood is undesirable from a timber standpoint and good management practices within plantations dictate that trees of low economic value are removed to increase the growth of high-value trees (McIntosh et al., 2012), ethanol production from eucalypt plantation thinnings is a potential option for bio-energy production, dependent on distance required for biomass transport, growth rate, and stocking rate.

NON-CELLULOSIC POLYSACCHARIDES

Before the formation of the secondary cell wall, the plant primary cell wall is a thin yet flexible structure that resists gravity and internal pressure while allowing growth and expansion (Cosgrove,

2005). Cellulose, being the core of the internal structure, provides the scaffold that non-cellulosic polysaccharides, such as hemicellulose and pectin, surround within a polysaccharide matrix (Carpita and Gibeau, 1993; Mellerowicz and Sundberg, 2008). Although hemicellulose and pectin are polysaccharides, and thus can hydrolyze into monomeric subunits, these monomers consist mainly of pentose sugars which are more difficult to ferment than glucose. As such, based on their difficulty to ferment and how they reduce access to cellulose, hemicellulose and pectin also contribute to biomass recalcitrance (Himmel et al., 2007; Sticklen, 2008).

Xyloglucan is the most abundant hemicellulose polysaccharide of woody dicot species, with a repeating structure of β 1–4 glucan residues with various side-chains, predominantly unbranched glycosyl residues or α 1–6 xylose. Other side-chain molecules include galactose, fructose, and arabinose (Harris and DeBolt, 2010; Scheller and Ulvskov, 2010). Xyloglucan interacts with cellulose by crosslinking with non-crystalline regions or through hydrogen bonding with the microfibrils themselves (Cosgrove, 2005). For further reinforcement and strength, woody plant cell walls synthesize a secondary cell wall of cellulose, hemicellulose, and lignin. However, unlike the primary cell wall with a repetitive hemicellulose structure, the secondary cell wall polysaccharide matrix is composed of highly variable xylan molecules. This varied structure is highly substituted, with the most common modification in woody dicots being glucuronosyl residues which generates glucuronoxylan (Li et al., 2006; Scheller and Ulvskov, 2010). Given that *E. globulus* is a major source of fiber for the pulp and paper industry, the structure of its non-cellulosic polysaccharides has been extensively researched. Originally, eucalypts were believed to possess glucuronoxylan as found in woody dicot species, but investigations by Shatalov et al. (1999) and Evtuguin et al. (2003) found that *E. globulus* xylan structure was highly substituted by galactosyl and acetyl residues. These residues, although not targets for saccharification, can affect downstream conversion efficiency. Galactose is one of the most difficult sugars to ferment (Lee et al., 2011), while acetyl groups can contribute acetic acid during fermentation conditions which inhibits ethanol production in *Pichia* (Ferrari et al., 1992) and *Saccharomyces* (Taherzadeh and Karimi, 2007).

Acid pretreatment, designed to hydrolyze the hemicellulose matrix surrounding cellulose, requires various acid concentrations, pretreatment times, and temperatures to be effective. To examine these parameters on various eucalypt species, McIntosh et al. (2012) conducted a 3³ factorial design (acid concentration, temperature, and pretreatment time) to understand sugar solubilization and degradation, enzymatic saccharification in response to pretreatment, and the fermentation of various hydrolyzates. Thinned trees of *Eucalyptus dunnii* and *Corymbia citriodora* subsp. *variegata* at ages 6 and 10 were tested within the factorial design, finding their biomass composition contained approximately 47–48% glucan, 16–17% xylan, 5% minor sugars, and 30% lignin. The authors found that under the mild pretreatment conditions [expressed as a combined severity factor (CSF)], monomeric xylose was the first to solubilize. However, as pretreatment became more severe, recovered xylose yields decreased, likely lost to degradation. Glucose release correlated

with CSF increase, with temperature being the main contributing factor, followed by acid concentration and reaction time. In the presence of crude *E. dun nii* hydrolyzate, *Saccharomyces cerevisiae* could be cultured for fermentation, although the time (30 h) at which the organism was able to convert 38 g of glucose into 18 g/L of ethanol (92% efficiency) was double when compared to starch-fed fermentations (Sánchez and Cardona, 2008). This study highlights the cost/benefit analysis of biomass conversion, where more severe treatments will result in greater glucose yields but will generate more degradation products from matrix polysaccharides that inhibit fermentation. The authors also encountered significant differences in saccharification yield between biomass of different ages. After two pretreatment severity conditions (CSF 1.60 and 2.48), 6-year-old eucalypt biomass yielded greater amounts of glucose than their 10-year-old counterparts, despite similar chemical composition. These differences were attributed to changes in cellulose crystallinity, which may be species specific based on similar studies in *Populus* (DeMartini and Wyman, 2011).

Although xylose, the main monosaccharide present within hemicellulose, is more difficult to ferment by fungi due to an overproduction of nicotinamide adenine dinucleotide (NADH) under anaerobic conditions (Bruinenberg et al., 1983), hemicellulose exists as a matrix polysaccharide and is thus far less resistant to pretreatment than cellulose. To demonstrate the ease at which xylose, generated from residual *E. grandis* wood chips during pulp production, could be fermented into fuel, Silva et al. (2011) optimized ethanol production from hemicellulose hydrolyzate, generated from mild acid pretreatment. Dilute sulfuric acid was mixed with the wood chips and was then autoclaved (121°C, 45 min) to allow separation from the hemicellulose hydrolyzate portion from the solids' (cellulose and lignin) portion. The hydrolyzate was then fermented to ethanol by a *Pichia stipitis* strain, known for its ability to ferment xylose, to achieve an ethanol concentration of 15.3 g/L (100 L/tonne of biomass). As a comparison, the solids' portion, which was delignified using an alkaline NaOH pretreatment step (4%, w/v, 121°C, 20 min), was fermented by *S. cerevisiae* by an SSF process yielded a final ethanol concentration of 28.7 g/L.

Although this study demonstrates eucalypt biomass conversion from debarked biomass, bark accounts for approximately 10–12% of tree biomass residue processed from a plantation (Perlack et al., 2005; Zhu and Pan, 2010), which contains considerable levels of glucose (40%) and xylose (10%) (Lima et al., 2013). Given that bark is often not considered or optimized during lignocellulose pretreatment, Lima et al. (2013) tested various options for bark deconstruction from commercial *E. grandis* (EG) and *E. grandis* × *urophylla* (EGU) trees. The authors tested both one- and two-step acid and alkaline combinations in order to maximize sugar recovery. A combination of acid (1%) and NaOH (4%) pretreatment resulted in a solids fraction containing high concentrations of glucose from EG and EGU (78 and 81% dry weight, respectively); however, only 54.2 and 66.6% of total glucose was actually recovered after treatment. Upon saccharification, 65.4 and 84.5% of glucose was released from the acid + alkaline-treated bark samples. Alternatively, a single NaOH (4%) pretreatment step, while retaining lesser amounts of glucose

within the solids fraction (56 and 62%), resulted in higher total recovered glucose (63.4 and 73.1%) and more efficient enzymatic saccharification (78.5 and 98.6%).

Although alkaline pretreatments are widely used, particularly in the pulp and paper industry, the chemicals required are considered pollutants and require multiple purification steps for removal from hydrolyzate. More recently, ILs, organic salts that are liquid at room temperature act as a solvent to solubilize cellulose, hemicellulose, and lignin without degradation, have been used as an effective pretreatment (Zhu et al., 2006). ILs, although not yet developed for large-scale use, are prized for their stability, recyclability, and low volatility during biomass solubilization (Zhu et al., 2006; Shi et al., 2015). As an emerging, pretreatment option, their exact interaction with biomass during solubilization is not well understood. To examine changes in cell wall structure and composition in woody biomass in response to IL pretreatment, Çetinkol et al. (2010) compared the cell wall of *E. globulus* before and after exposure to IL 1-ethyl-3-methylimidazolium acetate [C2min][OAc]. Using a variety of imaging and spectroscopy techniques [2-dimensional nuclear magnetic resonance spectroscopy (2D-NMR), Fourier transform infrared spectroscopy, scanning electron microscopy, small angle neutron scattering, and X-ray diffraction], they found IL pretreatment resulted in the deacetylation of xylan, acetylation of lignin, and the selective removal of G lignin monomers thereby increasing the S/G ratio. Subsequent saccharification of the treated biomass showed a significant increase in glucose (5×) yield after 1 h saccharification, which authors attributed to a decrease in cellulose crystallinity. Xylose yield was also increased after IL treatment, which was undetectable after saccharification of untreated biomass.

Depending on their chemistry, ILs interact with biomass differently. Protic ILs (PILs) can be prepared via a one-step process with low-cost acids and bases and preferentially solubilize lignin, while aprotic IL (AIL) preparation is a multistep process and preferentially dissolve carbohydrate macromolecules (Greaves et al., 2006; Zhang et al., 2015). Zhang et al. (2015) developed a concerted IL pretreatment (CIL) for *Eucalyptus* bark, combining pyrrolidinium acetate ([Pyr][AC]; PIL) with 1-butyl-3-methylimidazolium acetate ([BMIM][AC]; AIL). Compared to untreated bark, each IL pretreatment alone ([Pyr] or [BMIM]) or separate combinations of each ([Pyr] and [BMIM]), the CIL pretreatment ([Pyr]/[BMIM]) resulted in 91% enzymatic hydrolysis of cellulose, as compared to 5, 67, 50, and 77%. The same trend (13, 48, 65, and 79%) was observed during enzymatic hemicellulose hydrolysis as well (untreated biomass, [Pyr], [BMIM], [Pyr] and [BMIM], and [Pyr]/[BMIM]). Reduced lignin content correlated with cellulose conversion, which was further enhanced through the removal of hemicellulose. These strategies of converting underutilized (bark, thinned trees, and hemicellulose hydrolyzate) or undesirable (tension wood) lignocellulose will be a key for the sustainable generation of biofuels through coupling bioenergy production with traditional industrial forestry practices (van Heiningen, 2006).

While acid pretreatment remains a common method of pretreatment due to its effectiveness, strong industrial acids are expensive to generate and difficult to recycle and neutralize

(Menon et al., 2010). An alternative pretreatment method utilizes residues on the xylan backbone to disrupt the structure of lignocellulose. Hot water pretreatment, or autohydrolysis, is a cost-effective pretreatment option that mixes pressurizes hot water with biomass in a reaction vessel, causing acetyl residues on the xylan backbone to generate *in situ* acetic acid. The internal generation of acetic acid reduces the pH of the biomass liquor and accelerates delignification and the solubilization of hemicellulose (Galbe and Zacchi, 2007). To demonstrate the effectiveness of liquid hot water pretreatment for eucalypt biomass, Yu et al. (2010) developed a two-step pretreatment assay (step 1: 180–200°C, 0–60 min and step 2: 180–240°C; 0, 20, 40, and 60 min) to achieve maximize xylose recovery and minimize cellulose degradation. Their results demonstrated that during the first pretreatment step, degradation of xylose to furfural increases linearly with reaction severity, a trend which continues during the second pretreatment step where furfural concentration increases between 180 and 200°C then seemingly decreases through the formation of other aldehyde products. During the second pretreatment step, furfural and 5-HMF production increased steadily over time at constant temperature (200°C), demonstrating that extended pretreatments are detrimental for recovery of monomeric sugars. Temperature had the greatest effect on the formation of inhibitory products, with authors finding that shorter reaction times and lower temperatures (180°C, 20 min; 200°C, 20 min) maximized sugar recovery (96.6%) and enzymatic digestion (81.5%).

Although autohydrolysis pretreatment can effectively solubilize hemicellulose, cellulose will remain in its recalcitrant, crystalline form after pretreatment. To reduce cellulose crystallinity in conjunction with autohydrolysis pretreatment, Inoue et al. (2008) used ball milling to improve saccharification yield from *Eucalyptus* biomass. The authors demonstrated that milling alone for short periods of time (20 min) could dramatically reduce cellulose crystallinity from 59.7 to 7.6%, although only 44.2% of sugars were captured after saccharification. To achieve higher rates of enzymatic saccharification from ball-milled biomass (86.2%), restrictively long milling times were required (120 min). To combat this, the authors combined a hot water pretreatment (160°C, 30 min) and ball milling (20 min) step to yield approximately 70% of total sugars with a low enzyme loading [4 filter paper units (FPU)/g substrate]. By comparison, the same yields were achieved by hot water pretreatment (160°C, 30 min) or ball milling (40 min) separately, each requiring 10× enzyme loading (40 FPU/g). This study demonstrates how combining methods can effectively reduce the severity of the pretreatment required to deconstruct biomass, which will lessen the formation of inhibitory products and the costs associated with enzymatic saccharification.

Traditionally, lignocellulosic biofuel production required separated process vessels where polysaccharide hydrolysis was carried out independently from microbial fermentation. Separate hydrolysis and fermentation (SHF) required additional processing and distilling steps to remove contaminants that prevent biofuel production (Olofsson et al., 2008). To improve biomass conversion efficiency and reduce fuel production costs, SSF processes generate liquid fuel from sugars as they are hydrolyzed from a polysaccharide. The advantages of SSF over SHF include

the use of a single reactor for production to reduce capital costs, lower accumulation of sugars which bolsters saccharification rate and yield, and the presence of ethanol in the reaction vessel helps reduce microbial contamination (Krishna and Chowdary, 2000; Olofsson et al., 2008). To examine the efficiency of organosolv (in this case, ethanol and water) pretreatment with SSF processes, Yáñez-S et al. (2013) pretreated *E. globulus* biomass using an SSF process with various substrate loadings (10 and 15%, w/v), thermostable yeast concentrations (6 and 12 g/L), and enzyme loadings (as expressed as cellulase FPU/β-glucosidase IU [10/20, 20/40, and 30/60]). The authors found that the highest ethanol concentration (42 g/L) was obtained from 15% (w/v) substrate loading, 20 FPU/40 IU enzyme loading, at either yeast concentration. Although higher substrate loading decreased the overall ethanol yield, ethanol concentration within the reaction vessel was increased. Furthermore, mass balance calculation from 15% substrate loading within SSF and SHF processes suggested that greater ethanol amounts could be achieved by SSF (164 and 107 L/tonne, respectively).

The strategy of increasing the solids loading during an SSF is another strategy to further reduce operation costs associated with fuel production. By increasing the weight of solids to 15–20% of the SSF reaction, the energy required to heat and distil the reaction is dramatically reduced (Wang et al., 2011). Of course, this requires optimization of process parameters, such as liquid-to-solid ratio (LSR) and enzyme-to-substrate ratio (Romaní et al., 2011). Optimization of these parameters with *E. globulus* biomass, as well as autohydrolysis pretreatment severity, allowed Romaní et al. (2012) to reach an ethanol concentration of 67.4 g/L, representing 91% conversion of ethanol from cellulose, which scales to 291 L of ethanol per tonne of biomass.

Steam explosion (SE), another cost-effective pretreatment that is similar to autohydrolysis, solubilizes hemicellulose and disrupts the structure of biomass through the breakage of linkages caused by a sudden drop in pressure. SE pretreatment is often combined with alkaline or dilute acid catalysts to increase saccharification through either delignification or increased recovery of xylose (respectively). However, addition of catalysts increase biofuel production costs either through the cost of the chemical itself or through the additional washing and neutralization steps. Thus, optimization of SE pretreatment can provide an environmentally friendly process for biofuel production. Romaní et al. (2013) optimized the temperature (173–216) and pretreatment time ranges (6–34 min) with fixed enzyme loadings (15 FPU/10 IU) to improve ethanol production from *E. globulus* biomass. Using a scanning electron micrograph (SEM) to visualize the biomass after explosion, the authors observed that exposure to a temperature of 210°C for 30 min completely opened up the fibular structure of the biomass. Although, maximum ethanol production of the SE treated material was achieved under less severe conditions (210°C, 10 min) which produced 50.9 g/L from an SSF reactor. This represents again a 91% theoretical conversion of ethanol from cellulose, scaling to 248 L/tonne of biomass.

Microbial fermentation efficiency is another limiting step during lignocellulosic biofuel production. High-fuel production strains of yeast can readily convert glucose to ethanol while withstanding ethanol toxicity but are largely unable to utilize

hemicellulose derived pentose sugars (Lange, 2007). Alternate strains, belonging to *Pichia* and *Candida* genera, are capable of xylose fermentation but lack productivity. Metabolic engineering achieved through transformation to generate an organism capable of efficiently utilizing multiple carbon sources will greatly increase lignocellulosic fuel production, particularly one unfettered by high concentrations of ethanol or aldehydes, such as furfural and 5-HMF (Sun and Cheng, 2002; Wen et al., 2009). Despite eucalypt's desirable biofuel characteristics, their preferred climate ranges from cool temperate to tropical rainforest (Grattapaglia and Kirst, 2008; Shepherd et al., 2011). As such, their productivity as an energy crop outside of these climates is limited. To combat this limitation, Castro et al. (2014) investigated *E. benthamii*, a naturally cold resistant species that is commercially grown in Southeast USA, as a potential biofuel feedstock. To maximize biomass conversion, authors used a process known as liquefaction plus simultaneous saccharification and cofermentation (L + SSf), which combines dilute acid SE pretreatment with SSF processes with an inhibitor-resistant *E. coli* strain (SL100) capable of dual glucose/xylose fermentation. In addition, the authors used phosphoric acid instead of sulfuric acid, as it forms fewer inhibitory products during deconstruction and it allows the use of lower grades of stainless steel in reaction vessels, which saves on capital costs. Through optimization of temperature, acid concentration and pretreatment time (combined as a function of CSF), Castro et al. (2014) found that sugar yields were affected primarily by pretreatment time and temperature, with acid concentration having the smallest impact. Within the reaction vessel during fermentation, glucose was completely consumed within 48 h of fermentation, at which point the SL100 strain began fermenting xylose for the remainder of the 96 h fermentation. The cofermentation strategy to utilize all available carbon for conversion was successful, producing 240 g of ethanol/kg of raw biomass (304 L/tonne). For comparison, average ethanol production from sugarcane bagasse using the same process achieved 270–280 g/kg (342 SSf 355 L/tonne) (Geddes et al., 2013). Given the low costs of producing woody biomass (Hamelinck et al., 2005), this combination of strategies to employ alternative chemicals, SSF reaction vessels and cofermentation microbial strains that are engineered to withstand the detrimental effects of inhibitors demonstrates the feasibility of using eucalypts as a cost-effective crop for bioenergy production.

While ethanol is the most widely produced biofuel due to its ease of production, butanol is another fermentation product that can be used as a liquid fuel. Butanol is less volatile, hygroscopic, corrosive, and explosive than ethanol, can be transported with current infrastructure, and has similar energy content to gasoline (Antoni et al., 2007; Dürre, 2007; Ezeji et al., 2007; Fortman et al., 2008). Despite its advantages, microbial fermentation to butanol lacks efficiency given butanol's toxicity and often requiring nutrient supplementation which increasing operating costs (Zheng et al., 2015). Zheng et al. (2015) demonstrated the feasibility of acetone–butanol–ethanol (ABE) production from *Clostridium saccharoperbutylacetonicum* from steam exploded *Eucalyptus* biomass without nutrient supplementation. Various glucose concentrations (30–75 g/L) were achieved though varying solid

loadings (6.7–25%) finding that a hydrolyzate loading of 10% (39.5 g/L) generated the highest concentration of ABE (acetone 4.07 g/L, butanol 7.72 g/L, and ethanol 0.467 g/L). However, further optimization of glucose concentration (dilution of 75–45 g/L) produced the highest ABE concentrations (4.27 g/L acetone, 8.16 g/L butanol, and 0.643 g/L ethanol). Solids loading beyond 10% had a detrimental effect on ABE production, likely due to formation of fermentation inhibitors such as 5-HMF and phenolics.

LIGNIN

Lignin, being the second most abundant biopolymer in plant tissue, accounts for roughly 25% of biomass. Its primary role is to provide strength and rigidity to the plant, as well as assisting in vascular water transport and protection from pathogens (Boerjan et al., 2003; Ralph et al., 2004). While providing critical functions for the plant, lignin effectively surrounds structural polysaccharides within the secondary cell wall, resulting in inefficient release of fermentable sugars from chemical or enzymatic hydrolysis (Hinchee et al., 2010; Jönsson et al., 2013).

Lignin synthesis begins with the conversion of phenylalanine to trans-cinnamic acid, catalyzed by the enzyme phenylalanine ammonia lyase (PAL). The remaining enzymatic steps have been well-reviewed (Ona et al., 1997; Li et al., 2006; Déjardin et al., 2010; Vanholme et al., 2010), but ultimately this biosynthetic pathway ends with the generation of the main precursors of the lignin molecule: coniferyl, *p*-coumaryl, and sinapyl alcohol (Bonawitz and Chapple, 2010). Upon transportation to the secondary cell wall, each alcohol precursor undergoes an oxidation reaction, mediated by laccase and peroxidase enzymes, which destabilize the monolignol causing it to form a covalent bond with another monolignol. Once bonded, these subunits form *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin (Ralph et al., 2004; Bonawitz and Chapple, 2010; Vanholme et al., 2010). The most common covalent bond to occur, particularly in eucalypt lignin, is the β - θ -4 linkage, which is predominately formed from S lignin monomers. Other linkages are present, such as β - β and β -5 dimers, but β - θ -4 linkages are preferential for pulp and biofuel production as they are less stable than other bonds, branch less frequently, and are more easily broken during alkaline pretreatment (Huntley et al., 2003; Hinchee et al., 2010).

Lignin represents the largest barrier to efficient deconstruction of woody biomass. Studies performed in transgenic lines of alfalfa, poplar and *Arabidopsis* have demonstrated how slight alterations in the quantity and composition of lignin can result in large downstream effects for the saccharification of biomass (Chen and Dixon, 2007; Leplé et al., 2007; Eudes et al., 2012). Given its importance to the survivability of the plant, genetic control of cell lignification is tightly regulated. Using the promoter region of cinnamoyl CoA reductase (CCR) from *E. gunnii*, paired with a reporter gene (*GUS*), Lacombe et al. (2000) demonstrated using transgenic tobacco plates that *EgCCR* was highly activated during development and lignification of xylem tissues. Control of the lignin biosynthetic pathway is achieved through AC-rich elements within gene promoters. These AC elements serve as a binding platform for transcription factors (such as LIM and

MYB) that modulate gene expression (Rogers and Campbell, 2004; Zhong and Ye, 2007). The LIM transcription factor, first identified in tobacco, upregulates lignin genes. When silenced in tobacco using antisense *NtLIM1* constructs, transcripts for phenylpropanoid genes *PAL*, 4 coumarate CoA ligase (*4CL*), and cinnamyl alcohol dehydrogenase (*CAD*) were also downregulated, resulting in plants with 27% less lignin than wild type (Kawaoka et al., 2000). Similarly, suppression of the *LIM1* ortholog in *E. camaldulensis* also downregulated the *PAL*, *4CL*, and *CAD* gene pathways, resulting in plants with not only 29% less lignin but also 5% higher structural polysaccharides. The polysaccharide increase could be a result of shifting carbon resources as a result of downregulating the phenylpropanoid pathways (Kawaoka et al., 2006).

The MYB transcription factor, first discovered as a regulator of the lignin pathway in snapdragons, also affects the transcription of the lignin gene pathways. Identified from cDNA libraries of differentiating xylem tissue, the *E. grandis* MYB2 gene when overexpressed in tobacco resulted in abnormal secondary cell wall thickening and altered lignin composition. Interestingly, while the expression of phenylpropanoid genes was unaltered, downstream genes responsible for monolignol synthesis [*4CL*, *p*-coumarate 3-hydroxylase (*C3H*), hydroxycinnamoyl:shikimate hydroxycinnamoyl transferase (*HCT*), caffeoyl CoA *O*-methyltransferase (*CCoAOMT*), ferulate 5-hydroxylase (*F5H*), caffeic acid *O*-methyltransferase (*COMT*), *CCR*, and *CAD*] were upregulated, increasing the S/G ratio composition of the lignin (Goicoechea et al., 2005). Another MYB transcription factor, identified from *E. grandis*, *EgMYB1*, when overexpressed in poplar and *Arabidopsis* resulted in plants with dwarfed leaves and stems and downregulated lignin and cellulose and hemicellulose transcripts. Given that the upregulation of *EgMYB1* resulted in the alteration of the major components of secondary cell wall structures suggests that MYB1 is a weak activator of lignocellulose genes, and its upregulation outcompetes stronger activators, thereby reducing overall transcription (Rogers and Campbell, 2004; Legay et al., 2010).

To investigate the effects of various wood properties on the enzymatic saccharification of woody biomass, such as lignin content, S/G ratio, cellulose crystallinity, fiber pore size, and enzyme adsorption, Santos et al. (2012) characterized the biomass of nine woody plants, including *E. nitens*, *E. globulus*, and *E. urograndis*. Using a Kraft alkaline pretreatment and fixed enzyme loading, the authors found that of all the parameters investigated, lignin content is the most significant contributing factor for saccharification. *E. globulus* biomass conversion resulted in the highest sugar recovery, efficient enzymatic conversion, and least residual lignin (75.2, 97.9, and 6.9%, respectively). However, lignin content alone did not fully explain saccharification yields, as biomass with similar lignin levels released much less glucose than *E. globulus*. Lignin S/G ratios were also found to impact enzymatic hydrolysis, as increased S lignin monomers undergo less frequent branching, producing a more linear polymer which increases enzymatic access to polysaccharides. Although, the effect of S/G ratio on saccharification appears to be dependent on biomass pretreatment, as acid hydrolysis has been shown to have a greater effect on low S/G lignin (Davison et al., 2006) while Papa et al. (2012)

demonstrated using three mutant lines of *E. globulus* with varying S/G ratios (0.94, 1.13, and 2.15) that lignin composition did not affect saccharification after IL pretreatment.

Given that lignin remains the largest barrier to effective deconstruction of woody biomass for fermentation, treatments to increase the efficiency at which it can be removed from biomass will aid biofuel production. To improve enzymatic saccharification of eucalypt biomass, Sykes et al. (2015) generated transgenic *E. grandis* × *urophylla* hybrids with RNA interference (RNAi)-downregulated lignin biosynthetic genes *C3H* and cinnamate 4-hydroxylase (*C4H*). Total lignin content in transgenic lines was reduced by 8–9%, and after hot water pretreatment (designed as a mild, cost-effective method for biomass disruption) and enzymatic saccharification, both *C3H* (94%) and *C4H* (97%) transgenic lines released higher total sugars than control biomass (80% saccharification). However, transgenic lines were dwarfed (*C3H* – 2.0 m and *C4H* – 3.4 m) as compared to controls (6.0 m), a common issue for lignin transgenic plants that could be alleviated through silviculture practices.

Until low lignin transgenic plants are further developed, large-scale biofuel production will depend on harsher pretreatments that inhibit microbial growth and enzymatic action through solubilization of phenolics (Ximenes et al., 2010; Jönsson et al., 2013). An alternative option to aid in delignification of biomass is the addition of laccase to destabilize the lignin network through phenol oxidation. Gutiérrez et al. (2012) and Rico et al. (2014) tested the potential of a laccase enzyme to increase saccharification from *E. globulus* biomass. Tested in the presence of an enzyme mediator, either 1-hydroxybenzotriazole (HBT) or methyl syringate (respectively), both studies reported lignin reduction (~48%) in *E. globulus* substrate and increased glucose and xylose yields after saccharification. Using pyrolysis-gas chromatography/mass spectroscopy to understand the effect of the laccase treatment, authors found an increased S/G composition (4.9 vs. 4.0) within the lignin because of preferential hydrolysis of G lignin subunits, resulting in a less condensed phenolic polymer. Continued investigation of laccase pretreatment with mediators was conducted by Rico et al. (2015) using 2D-NMR to characterize each step of delignification by fungal enzymes with *E. globulus* biomass and cellulolytic lignin. The low redox potential *M. thermophila* laccase enzyme and methyl syringate mediator pretreatment was tested against a high redox potential laccase, isolated from *Pycnoporus cinnabarinus*, with HBT mediator across several stages of pretreatment and alkaline extraction. Though various structural changes occurred throughout each stage of the fungal pretreatments, the most striking effects involved the preferential removal of guaiacyl units, reduced β -0-4 alkyl-aryl ether linkages, and S/G ratio increase. Syringyl lignin subunits underwent C_{α} oxidation during laccase pretreatment, which were incompletely removed through alkaline extraction. Both fungal enzyme treatments achieved similar delignification results (~50%), although multistage analysis suggests that the rate of oxidation by *P. cinnabarinus* laccase + HBT was greater. The 50% delignification result correlated with a 30% increase in glucose yield after enzymatic saccharification. These results suggest that the largest gains in sugar release from biomass result from total delignification of biomass rather than the alteration of lignin composition.

TABLE 2 | Advances in lignocellulosic biofuel production from eucalypt biomass.

Reference	Strategy	Pretreatment and fermentation conditions	Conclusions	Result
Inoue et al. (2008)	Pretreatments without acids/bases/solvents are cheaper with fewer environmental impacts	Autohydrolysis + milling	Dual pretreatment required 10× less enzyme for saccharification	70% sugar recovery
Yu et al. (2010)	Two-step liquid hot water hydrolysis of biomass	Autohydrolysis	Temperature affects degradation products formation Short reaction times and low temperatures maximize recovery	96.6% sugar recovery; 81.5% saccharification
Çetinkol et al. (2010)	IL pretreatment of biomass	1-Ethyl-3-methyl imidazolium acetate	Deacetylation of xylan Acetylation of lignin Increased S/G ratio	5× glucose yield
Silva et al. (2011)	Hemicellulose deconstruction and fermentation from residual wood chips	Dilute sulfuric acid <i>P. stipitis</i> (<i>S. cerevisiae</i> fermentation of solids)	Hemicellulose was separated from cellulose and lignin	15.3 g/L ethanol (100 L/tonne biomass) 28.7 g/L ethanol (obtained from solids)
Muñoz et al. (2011)	Fermentation of tension and opposite wood	Organosolv SSF fermentation with <i>S. cerevisiae</i>	Tension wood required milder conditions to delignify	35 g/L ethanol (290 L/tonne biomass)
McIntosh et al. (2012)	Optimization of acid concentration, temperature, and pretreatment time	Sulfuric acid <i>S. cerevisiae</i>	Hemicellulose solubilizes and degrades first Temperature contributes most to glucose release	18 g/L ethanol
Santos et al. (2012)	Screened various woody feedstocks with varying for wood properties	Alkaline pretreatment	Lignin content, enzyme adsorption, and S/G ratio contribute most saccharification	<i>E. globulus</i> biomass (low lignin content 7%, 98% saccharification, and 75% sugar recovery)
Papa et al. (2012)	Investigate effects of S/G ratio on IL pretreatment efficiency	1-Ethyl-3-methyl imidazolium acetate	S/G ratio did not affect IL pretreatment efficiency	Glucose yield of 759–897 g/kg cellulose after 24 h saccharification
Yáñez-S et al. (2013)	SSF optimization of substrate loading, yeast concentration, and enzyme loading	Organosolv	Higher substrate loading and midrange enzyme loading maximize yield	42 g/L ethanol (164 L/tonne of biomass)
Romaní et al. (2012)	SSF optimization of substrate and enzyme loading	Autohydrolysis and SSF reaction	91% conversion of cellulose to ethanol	67.4 g/L ethanol (291 L/tonne of biomass)
Romaní et al. (2013)	Optimization of temperature and pretreatment time	Steam explosion and SSF reaction	Maximum ethanol yield is achieved at 210°C for 10 min	50.9 g/L ethanol (248 L/tonne of biomass)
Lima et al. (2013)	Optimization of pretreatment for <i>Eucalyptus</i> bark	One/two-step acid/alkaline pretreatment	Single alkaline step recovered most glucose	73.1% glucose recovery and 98.6% saccharification
Castro et al. (2014)	SSF fermentation with inhibitor-resistant cofermentation <i>E. coli</i> strain	Steam explosion + phosphoric acid SSF fermentation + cofermentation <i>E. coli</i>	Sugar yield is primarily determined by pretreatment time and temperature	240 g ethanol/kg biomass (304 L/tonne biomass)
Rico et al. (2014, 2015)	Fungal laccases with mediator pretreatment	Laccase pretreatment + alkaline extraction	Preferential G unit removal S unit oxidation Increased S/G ratio	~50% lignin reduction and 30% increase in saccharification
Martín-Sampedro et al. (2015)	Screening, isolation, and pretreatment with endophytic fungal laccases	Fungal pretreatment + autohydrolysis	Endophytic fungi outperformed white rot reference <i>Trametes</i> strain	3.3 and 2.9× increase in total sugar release after pretreatment
Sykes et al. (2015)	RNAi downregulation of lignin genes <i>C3H</i> and <i>C4H</i>	Hot water pretreatment	Transgenic lines had less lignin and underwent more efficient saccharification Transgenic plants were dwarfed	<i>C3H</i> (94% saccharification) <i>C4H</i> (97% saccharification) Control (80% saccharification)
Zhang et al. (2015)	IL pretreatment of eucalyptus bark	Pyrrolidinium acetate and 1-butyl-3-methylimidazolium acetate	IL combinations had a synergistic effect on pretreatment	91% enzymatic hydrolysis of cellulose
Zheng et al. (2015)	ABE production without nutrients	Steam explosion and <i>Clostridium</i> fermentation	Solids loading and glucose concentration are critical for microbial inhibition	4.27 g/L acetone, 8.16 g/L butanol, and 0.643 g/L ethanol

IL, ionic liquid; S, *Syringyl*; G, *Guaiacyl*; SSF, simultaneous saccharification fermentation; RNAi, RNA interference; *C3H*, *p-coumarate 3-hydroxylase*; *C4H*, *cinnamate 4-hydroxylase*; ABE, acetone/butanol/ethanol.

While *M. thermophila* is a commercially available strain, its laccase enzymes may lack specificity when applied to various lignocellulose feedstocks. To investigate novel laccase enzymes from endophytic fungi, occurring in symbiosis with *Eucalyptus* trees, Martín-Sampedro et al. (2015) screened more than 100 strains, selecting five for their ligninolytic enzymes. These strains, tested against a white rot *Trametes* sp. reference, were combined with 10 g of *Eucalyptus* wood chips, before or after mild autohydrolysis pretreatment (selected to minimize the production of fungal inhibitory products). Enzymatic saccharification of each pretreatment released greater sugar yields from combination of treatments (fungal + autohydrolysis) than either pretreatment alone. Endophytic fungi strains *Ulocladium* sp. and *Hormonema* sp. outperformed the *Trametes* sp. reference strain, resulting in 3.3- and 2.9-fold increase of total sugars (compared to a 2.3-fold increase) as compared to autohydrolyzed control biomass (~3 g/L). The authors postulated that the specific activity of the ligninolytic enzymes could be a result of evolutionary processes, and endophytic fungi represent a large reservoir of biodiversity to aid biofuel production.

CONCLUSION

Given the global demand and potential for lignocellulosic biofuels, selection and research into alternative feedstocks is essential. Eucalypts, given their wide range of phenotypic diversity, genetic potential, environmental adaptability, and desirable cell wall chemistry, are excellent candidates for bioenergy production (Table 2). While eucalypt biomass is highly prized for other industrial processes, such as pulp and paper and timber production, the most economical way to introduce lignocellulose into the energy supply chain will be in conjunction with other plantation practices where thinned and undesirable trees are removed to promote growth of high-value trees. In addition, the production of fuel from waste wood chips and bark within pulping factories will help convert mills into complete biorefineries. Indeed, as global paper consumption diminishes, alternative uses for eucalypt biomass will require research and development. While pulping plants are efficient at deconstruction, harsh pretreatments are not suitable for downstream microbial conversion of polysaccharides to monosaccharides to fuel. High temperatures and pressures, while effective for deconstruction, generate inhibitory compounds from lignin and carbohydrates that result in sugar losses and inefficient downstream processes. Lignocellulosic fuel will require mild, low-cost pretreatments, coupled with SSF or “one-pot” processes to promote efficient biofuel production.

Genetic and chemical exploitation of eucalypt cell walls has allowed the design of mild and environmentally friendly pretreatments, such as autohydrolysis and SE, relying on *in situ* acid generation to aid deconstruction without expensive and caustic

chemicals. Although these pretreatments help to reduce the formation of inhibitory products, aldehydes and phenolics formed from cellulose, hemicellulose, and lignin will likely remain in low concentrations within reaction vessels, necessitating the need for robust fermentive microbial strains. Metabolic engineering to exploit genetic variation has great potential to overcome the largest barriers to fuel conversion. These techniques have already generated dual fermentation strains to utilize all present carbon sources and resist the effects of degradation products within reaction vessels to main productivity. Application of the same principles to feedstocks have downregulated lignin gene pathways, designing plant cell walls that deconstruct with ease under mild conditions. Coupled with screening and isolation of endophytic fungi with specific ligninolytic enzymes, lignin deconstruction and removal from process vessels will maximize enzyme adsorption, sugar recovery, and fermentation.

Ionic liquids are the most promising for biomass pretreatment, given their stability and low volatility, and action at low temperatures. Despite the commercial use of cold resistant *E. benthamii*, eucalypts are not the ideal biofuel feedstock in all climates. ILs work universally well regardless of feedstock composition, solubilizing whole biomass without degradation, and selectively precipitating cellulose upon the addition of an antisolvent. Efficient saccharification of the cellulose precipitate maximizes sugar recovery and maintains intact lignin for alternate chemical processing.

In addition to the biological components of biofuel production, process optimization, such as single reaction SSF and high solids' loading, increases achievable ethanol concentrations and lower capital costs for production. Additional savings will be gained through the combination of pretreatments to reduce energy costs and enzyme loading for efficient saccharification. Increased understanding of eucalypt cell wall formation, particularly lignin formation, will allow the engineering of new and effective pretreatment options to make biofuel production suitable for a wide range of lignocellulosic feedstocks to provide renewable fuels for the future.

ACKNOWLEDGMENTS

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

REFERENCES

- Alvira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresour. Technol.* 101, 4851–4861. doi:10.1016/j.biortech.2009.11.093
- Antoni, D. W., Zverlov, V. V., and Schwarz, W. H. (2007). Biofuels from microbes. *Appl. Microbiol. Biotechnol.* 77, 23–35.
- Blanch, H. W., Simmons, B. A., and Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnol. J.* 6, 1086–1102. doi:10.1002/biot.201000180
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi:10.1146/annurev.arplant.54.031902.134938

- Bonawitz, N. D., and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* 44, 337–363. doi:10.1146/annurev-genet-102209-163508
- Bruinenberg, P. M., Bot, P. H. M., Dijken, J. P., and Scheffers, W. A. (1983). The role of redox balances in the anaerobic fermentation of xylose by yeasts. *Eur. J. Appl. Microbiol. Biotechnol.* 18, 287–292. doi:10.1007/BF00500493
- Carpita, N. C., and Gibeau, D. M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 3, 1–30. doi:10.1111/j.1365-313X.1993.tb00007.x
- Carroll, A., and Somerville, C. (2009). Cellulosic biofuels. *Annu. Rev. Plant Biol.* 60, 165–182. doi:10.1146/annurev-arplant.043008.092125
- Castro, E., Nieves, I. U., Mullinnix, M. T., Sagues, W. J., Hoffman, R. W., Fernández-Sandoval, M. T., et al. (2014). Optimization of dilute-phosphoric-acid steam pretreatment of *Eucalyptus benthamii* for biofuel production. *Appl. Energy* 125, 76–83. doi:10.1016/j.apenergy.2014.03.047
- Çetinkol, ÖP., Dibble, D. C., Cheng, G., Kent, M. S., Knierim, B., Auer, M., et al. (2010). Understanding the impact of ionic liquid pretreatment on *Eucalyptus*. *Biofuels* 1, 33–46. doi:10.4155/bfs.09.5
- Chen, F., and Dixon, R. A. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–761. doi:10.1038/nbt1316
- Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6, 850–861. doi:10.1038/nrm1746
- Davison, B. H., Drescher, S. R., Tuskan, G. A., Davis, M. F., and Nghiem, N. P. (2006). Variation of S/G ratio and lignin content in a *Populus* family influences the release of xylose by dilute acid hydrolysis. *Appl. Biochem. Biotechnol.* 130, 427. doi:10.1385/ABAB
- Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G., and Leplé, J.-C. (2010). Wood formation in angiosperms. *C. R. Biol.* 333, 325–334. doi:10.1016/j.crvi.2010.01.010
- DeMartini, J. D., and Wyman, C. E. (2011). Changes in composition and sugar release across the annual rings of *Populus* wood and implications on recalcitrance. *Bioresour. Technol.* 102, 1352–1358. doi:10.1016/j.biortech.2010.08.123
- Dürre, P. (2007). Biobutanol: an attractive biofuel. *Biotechnol. J.* 2, 1525–1534. doi:10.1002/biot.200700168
- Eudes, A., George, A., Mukerjee, P., Kim, J. S., Pollet, B., Benke, P. I., et al. (2012). Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.* 10, 609–620. doi:10.1111/j.1467-7652.2012.00692.x
- Evtuguin, D. V., Tomás, J. L., Silva, A. M. S., and Neto, C. P. (2003). Characterization of an acetylated heteroxylan from *Eucalyptus globulus* Labill. *Carbohydr. Res.* 338, 597–604. doi:10.1016/S0008-6215(02)00529-3
- Ezeji, T. C., Qureshi, N., and Blaschek, H. P. (2007). Bioproduction of butanol from biomass: from genes to bioreactors. *Curr. Opin. Biotechnol.* 18, 220–227. doi:10.1016/j.copbio.2007.04.002
- Ferrari, M. D., Neirrotti, E., Alborno, C., and Saucedo, E. (1992). Ethanol production from *Eucalyptus* wood hemicellulose hydrolysate by *Pichia stipitis*. *Biotechnol. Bioeng.* 40, 753–759. doi:10.1002/bit.260400702
- Fortman, J. L., Chhabra, S., Mukhopadhyay, A., Chou, H., Lee, T. S., Steen, E., et al. (2008). Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol.* 26, 375–381. doi:10.1016/j.tibtech.2008.03.008
- Furtado, A., Lupoi, J. S., Hoang, N. V., Healey, A., Singh, S., Simmons, B. A., et al. (2014). Modifying plants for biofuel and biomaterial production. *Plant Biotechnol. J.* 12, 1246–1258. doi:10.1111/pbi.12300
- Galbe, M., and Zacchi, G. (2007). Pretreatment of lignocellulosic materials for efficient bioethanol production. *Adv. Biochem. Eng. Biotechnol.* 108, 41–65. doi:10.1007/10_2007_070
- Geddes, C. C., Mullinnix, M. T., Nieves, I. U., Hoffman, R. W., Sagues, W. J., York, S. W., et al. (2013). Seed train development for the fermentation of bagasse from sweet sorghum and sugarcane using a simplified fermentation process. *Bioresour. Technol.* 128, 716–724. doi:10.1016/j.biortech.2012.09.121
- Goicoechea, M., Lacombe, E., Legay, S., Mihaljevic, S., Rech, P., Jauneau, A., et al. (2005). EgMYB2, a new transcriptional activator from *Eucalyptus* xylem, regulates secondary cell wall formation and lignin biosynthesis. *Plant J.* 43, 553–567. doi:10.1111/j.1365-313X.2005.02480.x
- González-García, S., Moreira, M. T., and Feijoo, G. (2012). Environmental aspects of *Eucalyptus* based ethanol production and use. *Sci. Total Environ.* 438, 1–8. doi:10.1016/j.scitotenv.2012.07.044
- Grattapaglia, D. (2008). *Genomics of Eucalyptus, a Global Tree for Energy, Paper, and Wood* (New York: Springer), 259–297.
- Grattapaglia, D., and Kirst, M. (2008). *Eucalyptus* applied genomics: from gene sequences to breeding tools. *New Phytol.* 179, 911–929. doi:10.1111/j.1469-8137.2008.02503.x
- Greaves, T. L., Weerawardena, A., Fong, C., Krodziewska, I., and Drummond, C. J. (2006). Protic ionic liquids: solvents with tunable phase behavior and physico-chemical properties. *J. Phys. Chem.* 110, 22479–22487. doi:10.1021/jp0634048
- Gutiérrez, A., Rencoret, J., Cadena, E. M., Rico, A., Barth, D., Del Río, J. C., et al. (2012). Demonstration of laccase-based removal of lignin from wood and non-wood plant feedstocks. *Bioresour. Technol.* 119, 114–122. doi:10.1016/j.biortech.2012.05.112
- Hall, M., Bansal, P., Lee, J. H., Realff, M. J., and Bommarius, A. S. (2010). Cellulose crystallinity – a key predictor of the enzymatic hydrolysis rate. *FEBS J.* 277, 1571–1582. doi:10.1111/j.1742-4658.2010.07585.x
- Hamelinck, C. N., Hooijdonk, G. V., and Faaij, A. P. (2005). Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass Bioenergy* 28, 384–410. doi:10.1016/j.biombioe.2004.09.002
- Harris, D., and DeBolt, S. (2010). Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnol. J.* 8, 244–262. doi:10.1111/j.1467-7652.2009.00481.x
- Hendriks, A. T. W. M., and Zeeman, G. (2009). Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresour. Technol.* 100, 10–18. doi:10.1016/j.biortech.2008.05.027
- Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Mark, R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi:10.1126/science.1137016
- Hinchee, M., Rottmann, W., Mullinax, L., Zhang, C., Chang, S., Cunningham, M., et al. (2009). Short-rotation woody crops for bioenergy and biofuels applications. *In Vitro Cell. Dev. Biol. Plant* 45, 619–629. doi:10.1007/s11627-009-9235-5
- Hinchee, M. A. W., Mullinax, L. N., and Rottmann, W. H. (2010). Chapter 7 woody biomass and purpose-grown trees as feedstocks for renewable energy. In *Plant Biotechnology for Sustainable Production of Energy and Co-products*, eds Mascia P. N., Scheffran J., and Widholm J. M. (Berlin: Springer), 155–208.
- Hisano, H., Nandakumar, R., and Wang, Z.-Y. (2009). Genetic modification of lignin biosynthesis for improved biofuel production. *In Vitro Cell. Dev. Biol. Plant* 45, 306–313. doi:10.1007/s11627-009-9219-5
- Horsley, T. N., and Johnson, S. D. (2007). Is *Eucalyptus* cryptically self-incompatible? *Ann. Bot.* 100, 1373–1378. doi:10.1093/aob/mcm223
- Huntley, S. K., Ellis, D., Gilbert, M., Chapple, C., and Mansfield, S. D. (2003). Significant increases in pulping efficiency in C4H-F5H-transformed poplars: improved chemical savings and reduced environmental toxins. *J. Agric. Food Chem.* 51, 6178–6183. doi:10.1021/jf034320o
- Inoue, H., Yano, S., Endo, T., Sakaki, T., and Sawayama, S. (2008). Combining hot-compressed water and ball milling pretreatments to improve the efficiency of the enzymatic hydrolysis of *Eucalyptus*. *Biotechnol. Biofuels* 1, 2–2. doi:10.1186/1754-6834-1-2
- Jönsson, L. J., Alriksson, B., and Nilvebrant, N.-O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* 6, 16–16. doi:10.1186/1754-6834-6-16
- Joshi, C. P., and Mansfield, S. D. (2007). The cellulose paradox – simple molecule, complex biosynthesis. *Curr. Opin. Plant Biol.* 10, 220–226. doi:10.1016/j.pbi.2007.04.013
- Jourez, B., Riboux, A., and Leclercq, A. (2001). Comparison of basic density and longitudinal shrinkage in tension wood and opposite wood in young stems of *Populus euramericana* cv. Ghoy when subjected to a gravitational stimulus. *Can. J. For. Res.* 31, 1676–1683. doi:10.1139/cjfr-31-10-1676
- Kawaoka, A., Kaohien, P., Yoshida, K., Endo, S., Yamada, K., and Ebinuma, H. (2000). Functional analysis of tobacco LIM protein Ntlm1 involved in lignin biosynthesis. *Plant J.* 22, 289–301. doi:10.1046/j.1365-313x.2000.00737.x
- Kawaoka, A., Nanto, K., Ishii, K., and Ebinuma, H. (2006). Reduction of lignin content by suppression of expression of the LIM domain transcription factor in *Eucalyptus camaldulensis*. *Silvae Genet.* 6, 269–277.
- Kaylen, M., Van Dyne, D. L., Choi, Y.-S., and Blase, M. (2000). Economic feasibility of producing ethanol from lignocellulosic feedstocks. *Bioresour. Technol.* 72, 19–32. doi:10.1016/S0960-8524(99)00091-7
- Krishna, S. H., and Chowdhary, G. V. (2000). Optimization of simultaneous saccharification and fermentation for the production of ethanol from lignocellulosic biomass. *J. Agric. Food Chem.* 48, 1971–1976. doi:10.1021/jf991296z

- Lacombe, E., Van Doorselaere, J., Boerjan, W., Boudet, A. M., and Grima-Pettenati, J. (2000). Characterization of cis-elements required for vascular expression of the cinnamoyl CoA reductase gene and for protein-DNA complex formation. *Plant J.* 23, 663–676. doi:10.1046/j.1365-3113.2000.00838.x
- Ladiges, P. Y., Udovicic, F., and Nelson, G. (2003). Australian biogeographical connections and the phylogeny of large genera in the plant family *Myrtaceae*. *J. Biogeogr.* 30, 989–998. doi:10.1046/j.1365-2699.2003.00881.x
- Lange, J. P. (2007). Lignocellulose conversion: an introduction to chemistry, process and economics. *Biofuels Bioprod. Biorefin.* 1, 39–48. doi:10.1002/bbb
- Lee, K. S., Hong, M. E., Jung, S. C., Ha, S. J., Yu, B. J., Koo, H. M., et al. (2011). Improved galactose fermentation of *Saccharomyces cerevisiae* through inverse metabolic engineering. *Biotechnol. Bioeng.* 108, 621–631. doi:10.1002/bit.22988
- Legay, S., Sivadon, P., Blervacq, A.-S., Pavy, N., Baghdady, A., Tremblay, L., et al. (2010). EgMYB1, an R2R3 MYB transcription factor from *Eucalyptus* negatively regulates secondary cell wall formation in *Arabidopsis* and poplar. *New Phytol.* 188, 774–786. doi:10.1111/j.1469-8137.2010.03432.x
- Léplé, J.-C., Dauwe, R., Morreel, K., Storme, V., Lapierre, C., Naumann, A., et al. (2007). Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* 19, 3669–3691. doi:10.1105/tpc.107.054148
- Li, L., Lu, S., and Chiang, V. (2006). A genomic and molecular view of wood formation. *CRC Crit. Rev. Plant Sci.* 25, 215–233. doi:10.1080/07352680600611519
- Lima, M. A., Lavorente, G. B., Da Silva, H. K., Bragatto, J., Rezende, C. A., Bernardinelli, O. D., et al. (2013). Effects of pretreatment on morphology, chemical composition and enzymatic digestibility of *Eucalyptus* bark: a potentially valuable source of fermentable sugars for biofuel production – part 1. *Biotechnol. Biofuels* 6, 75–75. doi:10.1186/1754-6834-6-75
- Macmillan, C. P., Taylor, L., Bi, Y., Southerton, S. G., and Evans, R. (2015). The fasciclin-like arabinogalactan protein family of *Eucalyptus grandis* contains members that impact wood biology and biomechanics. *New Phytol.* 206, 1314–1327. doi:10.1111/nph.13320
- Mansfield, S. D. (2009). Solutions for dissolution – engineering cell walls for deconstruction. *Curr. Opin. Biotechnol.* 20, 286–294. doi:10.1016/j.copbio.2009.05.001
- Martín-Sampedro, R., Fillat, Ú, Ibarra, D., and Eugenio, M. E. (2015). Use of new endophytic fungi as pretreatment to enhance enzymatic saccharification of *Eucalyptus globulus*. *Bioresour. Technol.* 196, 383–390. doi:10.1016/j.biortech.2015.07.088
- McIntosh, S., Vancov, T., Palmer, J., and Spain, M. (2012). Ethanol production from *Eucalyptus* plantation thinnings. *Bioresour. Technol.* 110, 264–272. doi:10.1016/j.biortech.2012.01.114
- Mellerowicz, E. J., and Sundberg, B. (2008). Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Curr. Opin. Plant Biol.* 11, 293–300. doi:10.1016/j.pbi.2008.03.003
- Menon, V., Prakash, G., and Rao, M. (2010). Value added products from hemicelluloses: biotechnological perspective. *Glob. J. Biochem.* 1, 36–67.
- Mohnen, D., Somerville, C., and Bar-Peled, M. (2008). “Biosynthesis of plant cell walls,” in *Biomass Recalcitrance*, ed. Himmel M. (Oxford: Blackwell Publishing), 94–187.
- Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., et al. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* 96, 673–686. doi:10.1016/j.biortech.2004.06.025
- Muñoz, C., Baeza, J., Freer, J., and Mendonça, R. T. (2011). Bioethanol production from tension and opposite wood of *Eucalyptus globulus* using organosolv pretreatment and simultaneous saccharification and fermentation. *J. Ind. Microbiol. Biotechnol.* 38, 1861–1866. doi:10.1007/s10295-011-0975-y
- Mutwil, M., Debolt, S., and Persson, S. (2008). Cellulose synthesis: a complex complex. *Curr. Opin. Plant Biol.* 11, 252–257. doi:10.1016/j.pbi.2008.03.007
- Myburg, A. A., Potts, B. M., Marques, C. M., Kirst, M., Gion, J.-M., and Grima-Pettenatti, J. (2007). “Eucalypts,” in *Genome Mapping and Molecular Breeding in Plants, Volume 7, Forest Trees*, ed. Kole C. (Berlin: Springer-Verlag), 115–160.
- Olofsson, K., Bertilsson, M., and Lidén, G. (2008). A short review on SSF – an interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol. Biofuels* 1, 7–7. doi:10.1186/1754-6834-1-7
- Ona, T., Sonoda, T., Itoh, K., and Shibata, M. (1997). Relationship of lignin content, lignin monomeric composition and hemicellulose composition in the same trunk sought by their within-tree variations in *Eucalyptus camaldulensis* and *E. globulus*. *Holzforschung* 51, 396–404. doi:10.1515/hfsg.1997.51.5.396
- Palmqvist, E., and Hahn-Hägerdal, B. (2000). Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresour. Technol.* 74, 25–33. doi:10.1016/S0960-8524(99)00161-3
- Papa, G., Varanasi, P., Sun, L., Cheng, G., Stavila, V., Holmes, B., et al. (2012). Exploring the effect of different plant lignin content and composition on ionic liquid pretreatment efficiency and enzymatic saccharification of *Eucalyptus globulus* L. mutants. *Bioresour. Technol.* 117, 352–359. doi:10.1016/j.biortech.2012.04.065
- Paux, E., Carocha, V., Marques, C., Mendes De Sousa, A., Borralho, N., Sivadon, P., et al. (2005). Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New Phytol.* 167, 89–100. doi:10.1111/j.1469-8137.2005.01396.x
- Perlack, R. D., Wright, L. L., Turhollow, A. F., Graham, R. L., Stokes, B. J., and Erblich, D. C. (2005). *Biomass as a Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*. Oak Ridge National Lab TN.
- Plomion, C., Leprovost, G., and Stokes, A. (2001). Wood formation in trees. *Plant Physiol.* 127, 1513–1523. doi:10.1104/pp.010816.1
- Poke, F. S., Vaillancourt, R. E., Potts, B. M., and Reid, J. B. (2005). Genomic research in *Eucalyptus*. *Genetica* 125, 79–101. doi:10.1007/s10709-005-5082-4
- Puri, M., Abraham, R. E., and Barrow, C. J. (2012). Biofuel production: prospects, challenges and feedstock in Australia. *Renew. Sustain. Energ. Rev.* 16, 6022–6031. doi:10.1016/j.rser.2012.06.025
- Qiu, D., Wilson, I. W., Gan, S., Washusen, R., Moran, G. F., and Southerton, S. G. (2008). Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. *New Phytol.* 179, 94–103. doi:10.1111/J.1469-8137.2008.02439.X
- Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., et al. (2004). Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochem. Rev.* 3, 29–60. doi:10.1023/B:PHYT.0000047809.65444.a4
- Ramos, L. P., and Saddler, J. N. (1994). “Bioconversion of wood residues mechanisms involved in pretreating and hydrolyzing lignocellulosic materials,” in *Enzymatic Conversion of Biomass for Fuels Production*, ed. Himmel M. (Washington, DC: ACS Symposium Series; American Chemical Society), 325–341.
- Rico, A., Rencoret, J., Del Río, J. C., Martínez, A. T., and Gutiérrez, A. (2014). Pretreatment with laccase and a phenolic mediator degrades lignin and enhances saccharification of *Eucalyptus* feedstock. *Biotechnol. Biofuels* 7, 6–6. doi:10.1186/1754-6834-7-6
- Rico, A., Rencoret, J., del Río, J. C., Martínez, A. T., and Gutiérrez, A. (2015). In-depth 2D NMR study of lignin modification during pretreatment of *Eucalyptus* wood with laccase and mediators. *Bioenerg. Res.* 8, 211–230. doi:10.1007/s12155-014-9505-x
- Rockwood, D. L., Rudie, A. W., Ralph, S. A., Zhu, J. Y., and Winandy, J. E. (2008). Energy product options for *Eucalyptus* species grown as short rotation woody crops. *Int. J. Mol. Sci.* 9, 1361–1378. doi:10.3390/ijms9081361
- Rogers, L. A., and Campbell, M. M. (2004). The genetic control of lignin deposition during plant growth and development. *New Phytol.* 164, 17–30. doi:10.1111/j
- Romaní, A., Garrote, G., Ballesteros, I., and Ballesteros, M. (2013). Second generation bioethanol from steam exploded *Eucalyptus globulus* wood. *Fuel* 111, 66–74. doi:10.1016/j.fuel.2013.04.076
- Romaní, A., Garrote, G., López, F., and Parajó, J. C. (2011). *Eucalyptus globulus* wood fractionation by autohydrolysis and organosolv delignification. *Bioresour. Technol.* 102, 5896–5904. doi:10.1016/j.biortech.2011.02.070
- Romaní, A., Garrote, G., and Parajó, J. C. (2012). Bioethanol production from autohydrolyzed *Eucalyptus globulus* by simultaneous saccharification and fermentation operating at high solids loading. *Fuel* 94, 305–312. doi:10.1016/j.fuel.2011.12.013
- Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature* 454, 841–845. doi:10.1038/nature07190
- Sánchez, Ó.J., and Cardona, C. A. (2008). Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresour. Technol.* 99, 5270–5295. doi:10.1016/j.biortech.2007.11.013
- Santos, R. B., Lee, J. M., Jameel, H., Chang, H.-M., and Lucia, L. A. (2012). Effects of hardwood structural and chemical characteristics on enzymatic hydrolysis for biofuel production. *Bioresour. Technol.* 110, 232–238. doi:10.1016/j.biortech.2012.01.085
- Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289. doi:10.1146/annurev-arplant-042809-112315

- Schimleck, L. R., Evans, R., and Ilic, J. (2001). Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy. *Can. J. For. Res.* 31, 1671–1675. doi:10.1139/cjfr-31-10-1671
- Shatalov, A. A., Evtuguin, D. V., and Pascoal Neto, C. (1999). (2-O-alpha-D-galactopyranosyl-4-O-methyl-alpha-D-glucurono)-D-xylan from *Eucalyptus globulus* Labill. *Carbohydr. Res.* 320, 93–99. doi:10.1016/S0008-6215(99)00136-6
- Shepherd, M., Bartle, J., Lee, D. J., Brawner, J., Bush, D., Turnbull, P., et al. (2011). eucalypts as a biofuel feedstock. *Biofuels* 2, 639–657. doi:10.4155/bfs.11.136
- Shi, J., George, K. W., Sun, N., He, W., Li, C., Stavila, V., et al. (2015). Impact of pretreatment technologies on saccharification and isopentenol fermentation of mixed lignocellulosic feedstocks. *Bioenerg. Res.* 8, 1004–1013. doi:10.1007/s12155-015-9588-z
- Silva, N. L. C., Betancur, G. J. V., Vasquez, M. P., Gomes, E. D. B., and Pereira, N. (2011). Ethanol production from residual wood chips of cellulose industry: acid pretreatment investigation, hemicellulosic hydrolysate fermentation, and remaining solid fraction fermentation by SSF process. *Appl. Biochem. Biotechnol.* 163, 928–936. doi:10.1007/s12010-010-9096-8
- Simmons, B. A., Loque, D., and Blanch, H. W. (2008). Next-generation biomass feedstocks for biofuel production. *Genome Biol.* 9, 242–242. doi:10.1186/gb-2008-9-12-242
- Singh, S., Simmons, B. A., and Vogel, K. P. (2009). Visualization of biomass solubilization and cellulose regeneration during ionic liquid pretreatment of switchgrass. *Biotechnol. Bioeng.* 104, 68–75. doi:10.1002/bit.22386
- Soccol, C. R., Faraco, V., Karp, S., Vandenberghe, L. P. S., Thomaz-Soccol, V., Woiciechowski, A., et al. (2011). *Lignocellulosic Bioethanol: Current Status and Future Perspectives*. San Diego, CA: Elsevier Inc.
- Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78. doi:10.1146/annurev.cellbio.22.022206.160206
- Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi:10.1126/science.1189268
- Sticklen, M. B. (2008). Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat. Rev. Genet.* 9, 433–443. doi:10.1038/nrg2336
- Stricker, J., Rockwood, D. L., Segrest, S. A., Alker, G. R., Prine, G. M., and Carter, D. R. (2000). Short rotation woody crops for Florida. in *Proc. 3rd Biennial Short Rotation Woody Crops Operations Working Group Conference*. p. 15–23.
- Sun, Y., and Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* 83, 1–11. doi:10.1016/S0960-8524(01)00212-7
- Sykes, R. W., Gjersing, E. L., Foutz, K., Rottmann, W. H., Kuhn, S. A., Foster, C. E., et al. (2015). Down-regulation of p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) and cinnamate 4-hydroxylase (C4H) genes in the lignin biosynthetic pathway of *Eucalyptus urophylla* × *E. grandis* leads to improved sugar release. *Biotechnol. Biofuels* 8, 128–128. doi:10.1186/s13068-015-0316-x
- Taherzadeh, M. J., and Karimi, K. (2007). Acid-based hydrolysis processes for ethanol from lignocellulosic materials: a review. *Bioresources* 2, 472–499.
- Talebna, F., Karakashev, D., and Angelidaki, I. (2010). Production of bioethanol from wheat straw: an overview on pretreatment, hydrolysis and fermentation. *Bioresour. Technol.* 101, 4744–4753. doi:10.1016/j.biortech.2009.11.080
- van Heiningen, A. (2006). Converting a kraft pulp mill into an integrated forest biorefinery. *Pulp Pap. Canada* 107, 141–146.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi:10.1104/pp.110.155119
- Wang, W., Kang, L., Wei, H., Arora, R., and Lee, Y. Y. (2011). Study on the decreased sugar yield in enzymatic hydrolysis of cellulosic substrate at high solid loading. *Appl. Biochem. Biotechnol.* 164, 1139–1149. doi:10.1007/s12010-011-9200-8
- Węgrzyn, J. L., Eckert, A. J., Choi, M., Lee, J. M., Stanton, B. J., Sykes, R., et al. (2010). Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, *Salicaceae*) secondary xylem. *New Phytol.* 188, 515–532. doi:10.1111/j.1469-8137.2010.03415.x
- Wen, F., Nair, N. U., and Zhao, H. (2009). Protein engineering in designing tailored enzymes and microorganisms for biofuels production. *Curr. Opin. Biotechnol.* 20, 412–419. doi:10.1016/j.copbio.2009.07.001
- Wyman, C. E. (1999). Biomass ethanol: technical progress, opportunities, and commercial challenges. *Annu. Rev. Energy Environ.* 24, 189–226. doi:10.1146/annurev.energy.24.1.189
- Ximenes, E., Kim, Y., Mosier, N., Dien, B., and Ladisch, M. (2010). Inhibition of cellulases by phenols. *Enzyme Microb. Technol.* 46, 170–176. doi:10.1016/j.enzmictec.2009.11.001
- Yáñez-S, M., Rojas, J., Castro, J., Ragauskas, A., Baeza, J., and Freer, J. (2013). Fuel ethanol production from *Eucalyptus globulus* wood by autocatalyzed organosolv pretreatment ethanol-water and SSF. *J. Chem. Technol. Biotechnol.* 88, 39–48. doi:10.1002/jctb.3895
- Yu, Q., Zhuang, X., Yuan, Z., Wang, Q., Qi, W., Wang, W., et al. (2010). Two-step liquid hot water pretreatment of *Eucalyptus grandis* to enhance sugar recovery and enzymatic digestibility of cellulose. *Bioresour. Technol.* 101, 4895–4899. doi:10.1016/j.biortech.2009.11.051
- Zhang, C., Xu, W., Yan, P., Liu, X., and Zhang, Z. C. (2015). Overcome the recalcitrance of *Eucalyptus* bark to enzymatic hydrolysis by concerted ionic liquid pretreatment. *Process Biochem.* doi:10.1016/j.procbio.2015.09.009
- Zheng, J., Tashiro, Y., Wang, Q., Sakai, K., and Sonomoto, K. (2015). Feasibility of acetone–butanol–ethanol fermentation from *Eucalyptus* hydrolysate without nutrients supplementation. *Appl. Energy* 140, 113–119. doi:10.1016/j.apenergy.2014.11.037
- Zheng, Y., Pan, Z., and Zhang, R. (2009). Overview of biomass pretreatment for cellulosic ethanol production. *Int. J. Agric. Biol. Eng.* 2, 51–68. doi:10.3965/j.issn.1934-6344.2009.03.051-068
- Zhong, R., and Ye, Z.-H. (2007). Regulation of cell wall biosynthesis. *Curr. Opin. Plant Biol.* 10, 564–572. doi:10.1016/j.pbi.2007.09.001
- Zhu, J. Y., and Pan, X. J. (2010). Woody biomass pretreatment for cellulosic ethanol production: technology and energy consumption evaluation. *Bioresour. Technol.* 101, 4992–5002. doi:10.1016/j.biortech.2009.11.007
- Zhu, S., Wu, Y., Chen, Q., Yu, Z., Wang, C., Jin, S., et al. (2006). Dissolution of cellulose with ionic liquids and its application: a mini-review. *Green Chem.* 8, 325–325. doi:10.1039/b601395c

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Healey, Lee, Furtado, Simmons and Henry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Optimization of alkaline and dilute acid pretreatment of agave bagasse by response surface methodology

Abimael I. Ávila-Lara¹, Jesus N. Camberos-Flores¹, Jorge A. Mendoza-Pérez², Sarah R. Messina-Fernández³, Claudia E. Saldaña-Duran³, Edgar I. Jimenez-Ruiz⁴, Leticia M. Sánchez-Herrera⁴ and Jose A. Pérez-Pimienta^{1*}

¹ Department of Chemical Engineering, Universidad Autónoma de Nayarit, Tepic, Mexico, ² Department of Engineering in Environmental Systems, Instituto Politécnico Nacional, Mexico City, Mexico, ³ Cuerpo Académico de Sustentabilidad Energética, Universidad Autónoma de Nayarit, Tepic, Mexico, ⁴ Food Technology Unit, Universidad Autónoma de Nayarit, Tepic, Mexico

OPEN ACCESS

Edited by:

Robert Henry,
The University of Queensland,
Australia

Reviewed by:

Jian Xu,
Chinese Academy of Sciences, China
Maria Gonzalez Alriols,
University of the Basque Country,
Spain

*Correspondence:

Jose A. Pérez-Pimienta,
Department of Chemical Engineering,
Universidad Autónoma de Nayarit,
Ciudad de la Cultura "Amado Nervo"
S/N, Tepic, Nayarit 63155, Mexico
japerez@uan.edu.mx

Specialty section:

This article was submitted to
Bioenergy and Biofuels, a section of
the journal *Frontiers in Bioengineering
and Biotechnology*

Received: 25 July 2015

Accepted: 09 September 2015

Published: 23 September 2015

Citation:

Ávila-Lara AI, Camberos-Flores JN,
Mendoza-Pérez JA,
Messina-Fernández SR,
Saldaña-Duran CE, Jimenez-Ruiz EI,
Sánchez-Herrera LM and
Pérez-Pimienta JA (2015)

Optimization of alkaline and dilute acid
pretreatment of agave bagasse by
response surface methodology.
Front. Bioeng. Biotechnol. 3:146.
doi: 10.3389/fbioe.2015.00146

Utilization of lignocellulosic materials for the production of value-added chemicals or biofuels generally requires a pretreatment process to overcome the recalcitrance of the plant biomass for further enzymatic hydrolysis and fermentation stages. Two of the most employed pretreatment processes are the ones that used dilute acid (DA) and alkaline (AL) catalyst providing specific effects on the physicochemical structure of the biomass, such as high xylan and lignin removal for DA and AL, respectively. Another important effect that need to be studied is the use of a high solids pretreatment ($\geq 15\%$) since offers many advantaged over lower solids loadings, including increased sugar and ethanol concentrations (in combination with a high solids saccharification), which will be reflected in lower capital costs; however, this data is currently limited. In this study, several variables, such as catalyst loading, retention time, and solids loading, were studied using response surface methodology (RSM) based on a factorial central composite design of DA and AL pretreatment on agave bagasse using a range of solids from 3 to 30% (w/w) to obtain optimal process conditions for each pretreatment. Subsequently enzymatic hydrolysis was performed using Novozymes Cellic CTec2 and HTec2 presented as total reducing sugar (TRS) yield. Pretreated biomass was characterized by wet-chemistry techniques and selected samples were analyzed by calorimetric techniques, and scanning electron/confocal fluorescent microscopy. RSM was also used to optimize the pretreatment conditions for maximum TRS yield. The optimum conditions were determined for AL pretreatment: 1.87% NaOH concentration, 50.3 min and 13.1% solids loading, whereas DA pretreatment: 2.1% acid concentration, 33.8 min and 8.5% solids loading.

Keywords: agave bagasse, high solids, biomass pretreatment, optimization, characterization

Introduction

Lignocellulosic biomass is the most abundant renewable carbohydrate source in the world and it is proposed to dominate the biofuel production in the future (Avci et al., 2013). Mainly composed by cellulose, hemicellulose, and lignin, their organization and interaction between these polymeric structures, the plant cell wall is naturally recalcitrant to biological degradation

(da Costa Sousa et al., 2009). A pretreatment step is fundamental to alter the structure of cellulosic biomass to make cellulose more accessible to the enzymes that convert the carbohydrate polymers into fermentable sugars (Mosier et al., 2005).

Many options exist for pretreatment of biomass, increase saccharification efficiency and improve the yields of monomeric sugars; the leading examples use liquid catalysts, such as sulfuric acid, ammonia, ionic liquid, or water, which penetrate the cell wall and alter its chemistry and ultrastructure (Dadi et al., 2006; Chundawat et al., 2011).

Recently, agave bagasse (AGB) byproduct of the Tequila industry that represent 40% of the harvested plant, with an annual generation in Mexico of about $1.12 \text{ kg} \times 10^8 \text{ kg}$ has been studied for biomass conversion using different pretreatment approaches, such as ionic liquid (Perez-Pimienta et al., 2013) and organosolv (Caspeta et al., 2014). Moreover, AGB was also been used with acid and enzymatic hydrolysis followed by a fermentation step using a native microorganism (*Pichia caribbica* UM-5) obtaining ~57% of theoretical ethanol (w/w) (Saucedo-Luna et al., 2011) or for the production of n-butanol and ethanol from different Agave species (Mielenz et al., 2015).

Dilute acid (DA) and alkaline (AL; NaOH) are among the most extensively studied biomass pretreatments in different feedstocks, such as grasses, agricultural residues, and woods (Kumar et al., 2009; Xu et al., 2010; Sathitsuksanoh et al., 2013; Zhang et al., 2014). The mode of action of the DA pretreatment typically use sulfuric acid that removes hemicellulose in a great extent improving the enzyme accessibility to cellulose which its effectiveness depends on the acid concentration and temperature applied during the process, however, if severe conditions are applied several degradation products are formed, mainly furfural, 5-hydroxymethylfurfural, phenolic acids and aldehydes, levulinic acid, and other aliphatic acids, which can inhibit both, enzymatic hydrolysis and fermentation (Mosier et al., 2005; da Costa Sousa et al., 2009). On the other hand, ALs pretreatment uses AL catalyst, such as sodium hydroxide, which are effective depending on the lignin content on the biomass, increasing cellulose digestibility through lignin solubilization/removal, exhibiting minor cellulose and hemicellulose solubilization than acid or hydrothermal processes (Avira et al., 2010).

In recent years, the need to investigate the use of high solids loading ($\geq 15\%$) in biomass pretreatment has increase hence offers many advantaged over lower solids loadings, including increased sugar and ethanol concentrations, which will be reflected in lower capital costs (Modenbach and Nokes, 2012; Li et al., 2013); however, this data is currently limited for DA and AL pretreatments in AGB (Hernández-Salas et al., 2009; Saucedo-Luna et al., 2011).

In the present manuscript, optimization of DA and AL pretreatment strategies for conversion of AGB to sugars using a central composite design (CCD) for response surface methodology (RSM) was studied. The objective of this study was to identify the optimum process conditions for the selected operating variables namely catalyst concentration, retention time, and solid loading for the maximum production of fermentable sugars. Furthermore, the untreated and selected samples from both pretreatments were characterized by calorimetric techniques (TGA), fluorescence and

energy dispersive X-ray spectroscopy (EDS), and scanning electron microscopy (SEM).

Materials and Methods

The biomass used in this study was obtained from Destilería Rubio, a Tequila plant from western Mexico. The AGB was harvested in August 2014. The biomass was milled with a Thomas-Wiley Mini Mill fitted with a 40-mesh screen (Model 3383-L10 Arthur H. Thomas Co., Philadelphia, PA, USA) and stored at 4°C in a sealed plastic bag. Cellic® CTec2 (Cellulase complex for degradation of cellulose) and HTec2 (Endoxylanase with high specificity toward soluble hemicellulose) were a gift from Novozymes (Davis, CA, USA).

Experimental Design

Optimization of processing conditions for fermentable sugars recovery was studied using a factorial CCD of RSM. The independent variables were catalyst concentration, residence time, and solids loading. The experimental data were fit using Eq. 1, a low-order polynomial equation to evaluate the effect of each independent variable to the response, which was later analyzed to obtain the optimum process conditions (Tan et al., 2011). In this study, a polynomial quadratic equation was employed as follows:

$$y = \beta_0 + \sum_{i=1}^3 \beta_i X_i + \sum_{i=1}^3 \beta_{ii} X_i^2 + \sum_{i=1}^3 \sum_{j=1}^3 \beta_{ij} X_i X_j \quad (1)$$

where y is the response, X_i and X_j are independent variables, β_0 is the constant coefficient, β_i is the i th linear coefficient, β_{ii} is the quadratic coefficient, and β_{ij} is the ij th interaction coefficient. CCD consists of 2^k factorial points, $2k$ axial points ($\pm \alpha$), and six central points, where k is the number of independent variables. Each of the variables were investigated at five coded levels ($-\alpha$, -1 , 0 , 1 , α), as listed in **Table 1**, and the complete experimental design matrix for this study is shown in **Table 2**. For each pretreatment (DA and AL), a total of 20 experiments per pretreatment were carried out, including eight per factorial design, six for axial points and six repetitions at the central point.

Alkaline Pretreatment

A NaOH solution at a specific concentration were placed in a serum bottle and mixed with AGB using a glass rod, forming

TABLE 1 | Levels of the pretreatment condition variables tested in the CCD.

Variable	Unit	Coding	Coded level				
			$-\alpha^a$	-1	0	1	$+\alpha^a$
Catalyst concentration	% (w/w)	A	0.15	0.73	1.58	2.42	3.00
Residence time	Min	B	15.00	30.20	52.50	74.80	90.00
Solids loading	% (w/w)	C	3.00	8.47	16.50	24.53	30.00

^a α (axial distance) = $\sqrt[4]{N}$, where N is the number of experiments of the factorial design. In this case, 1.6818.

TABLE 2 | Experimental design matrix of CCD and corresponding results (sugars and solids recovery).

Run	Experimental variables			Solids recovery (%)		TRS yield (mg/g biomass)	
	Catalyst concentration, A (% w/w)	Retention time, B (min)	Solids loading, C (% w/w)	AL	DA	AL	DA
1	1.58	15.00	16.50	70.3	71.9	506.1	419.3
2	2.42	30.20	8.47	74.3	58.5	476.9	391.3
3	0.73	30.20	8.47	75.5	83.0	447.5	410.9
4	2.42	30.20	24.53	72.3	66.8	468.2	371.6
5	0.73	30.20	24.53	84.3	82.7	415.6	385.7
6	0.15	52.50	16.50	86.2	80.8	360.3	339.7
7	1.58	52.50	3.00	65.5	57.6	513.6	457.2
8	3.00	52.50	16.50	60.7	58.3	457.9	364.5
9	1.58	52.50	30.00	80.6	68.7	421.0	399.7
10	2.42	74.80	8.47	63.4	54.4	460.5	360.5
11	0.73	74.80	8.47	72.4	77.3	460.1	428.0
12	2.42	74.80	24.53	74.6	65.3	453.7	353.7
13	0.73	74.80	24.53	87.6	86.1	437.2	409.0
14	1.58	90.00	16.50	72.8	70.4	521.5	397.9
15	1.58	52.50	16.50	76.2	68.6	532.8	394.4
16	1.58	52.50	16.50	76.7	71.6	517.1	415.2
17	1.58	52.50	16.50	78.2	68.8	497.2	443.8
18	1.58	52.50	16.50	79.3	70.6	521.3	431.5
19	1.58	52.50	16.50	79.9	69.5	502.7	439.4
20	1.58	52.50	16.50	77.8	70.4	511.6	435.9
Untreated	–	–	–	100		135.1	

AL, alkaline pretreatment; DA, dilute acid pretreatment.

a slurry at with a precise biomass concentration and the pretreatment was performed in autoclave conditions (121°C and ~15 psi) during the appropriate time according to **Table 1** (Xu et al., 2010). Pretreated biomass was recovered by filtration and washed with 400 mL of distilled water to remove excess alkali and dissolved byproducts. All experiments were conducted in triplicate.

Dilute Acid Pretreatment

The DA pretreatment with H₂SO₄ was conducted using the appropriate acid concentration and solids loading referred to **Table 1** at 130°C and 20 psi in an autoclave for a specific time (Sathitsuksanoh et al., 2013). After DA, the hydrolyzate was separated by filtration and the pretreated AGB was washed with 400 mL of distilled water prior to enzymatic hydrolysis. All experiments were conducted in triplicate.

Scanning Electron Microscopy

The morphology of untreated and selected pretreated AGB solids was analyzed using a high resolution SEM by a JEOL JSM-7800F equipment. The representative images were acquired with a 1 kV accelerating voltage and analysis using 20 kV.

Confocal Fluorescent Microscopy

The confocal fluorescent microscope images of untreated and selected pretreated AGB samples were taken using a Carl Zeiss LSM 710 NLO with two laser sources (405 and 633 nm). To demonstrate the microstructure based on the distribution of lignin (autofluorescence) and cellulose, all samples were labeled with Calcofluor white stain (0.1%) for 5 min, subsequently were

washed four times using distilled water and allowed to dry in the dark until analysis under the confocal microscope.

DSC and TGA Analysis

A differential scanning calorimeter (Pyris 1) from Perkin Elmer was employed with an argon atmosphere in the range of 50–450°C, at 10°C/min ramp. DSC curves were obtained with 3.3 mg. The TGA curves were obtained using around 3.8 mg of AGB as initial sample mass. The samples was tested in a SETARAM thermal analysis instrument, with temperature range of 50–800°C and heating rate of 10°C/min in argon atmosphere. Untreated and selected pretreated samples were measured by DSC and TGA.

Biomass Porosimetry

Nitrogen porosimetry (ASAP 2406) from Mca-Micromeritics was employed to measure the surface area, pore volume and pore size distribution of the untreated and selected pretreated AGB with the following methods from ASTM: ASTM D-3663(R2008), ASTM D-4222-03(R2008), and ASTM D-4641-12(R2008). Samples were degasified at 120°C.

Enzymatic Saccharification

The saccharification was carried out using commercially available Cellic® CTec2 and HTec2 enzyme mixtures of untreated and pretreated AGB samples, which was conducted at 55°C and 150 rpm in 50 mM citrate buffer (pH of 4.8). A 3% biomass loading was used, likewise, untreated AGB were run concurrently with the pretreated samples to eliminate potential differences in temperature history or enzyme loading. The enzyme concentrations of CTec2 and HTec2 were set at 35 FPU/g biomass and 60 CBU/g biomass, respectively. All assays were performed in triplicate.

DNS Assay

The total reducing sugar (TRS) yield of the final hydrolyzate calculated as mg sugar/g biomass was determined by DNS assay (Miller, 1959) on a DTX 880 Multimode Detector (Beckman Coulter, CA, USA) at 550 nm with solutions (0–10 g/L) of D-glucose in water as calibration standards. All assays were performed in triplicate.

Statistical Analysis

Analysis of experimental CCD results was carried out with the software Design-Expert 7.1.5 (Stat-Ease, Minneapolis, MN, USA). Each coefficient in Eq. 1 was calculated and the possible interaction effects of the process variables on the response were obtained. Their significance was checked by variance analysis (ANOVA) of experimental results.

Results and Discussion

Biochemical Composition Analysis of Untreated Agave Bagasse

By following, the National Renewable Energy Laboratory (NREL, Denver, CO, USA) protocols, the composition of untreated AGB in dry basis was 41.5% glucan, 20.3% xylan, 17.0% insoluble lignin, 3.8% soluble lignin, and 5.4% ash, which is consistent with other reported values (Davis et al., 2011; Perez-Pimienta et al., 2013). Glucan and xylan correspond to 61.8% of the total carbohydrates in the AGB.

Model Development

The experimental data were first analyzed, in order to obtain second-order polynomial equations including terms of interaction between the experimental variables using Design-Expert software and the following models for AL and DA pretreatment describes the TRS yield (mg sugar/g biomass) in terms of coded parameters and actual parameters are based on the statistical analysis of the experimental data shown in Table 2.

The final equations for AL pretreatment were as follows:

$$\begin{aligned} \text{TRS yield} = & 513.35 + 21.08 \times A + 3.57 \times B - 16.87 \times C \\ & - 9.95 \times AB + 4.87 \times AC + 1.67 \times BC - 38.44 \times A^2 \\ & - 2.52 \times B^2 - 18.14 \times C^2 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{TRS yield} = & 277.0937 + 203.1844 * \text{NaOH} + 1.0059 * \text{Time} \\ & + 5.6903 * \text{Solids} - 0.4313 * \text{NaOH} * \text{Time} \\ & + 0.7172 * \text{NaOH} * \text{Solids} + 0.0076 * \text{Time} * \text{Solids} \\ & - 53.8367 * \text{NaOH}^2 - 0.0034 * \text{Time}^2 \\ & - 0.2813 * \text{Solids}^2 \end{aligned} \quad (3)$$

In the same way, the final equations for DA pretreatment were as follows:

$$\begin{aligned} \text{TRS yield} = & 427.27 - 5.87 \times A + 0.26 \times B - 12.80 \times C \\ & - 13.65 \times AB + 2.19 \times AC + 2.92 \times BC \\ & - 27.48 \times A^2 - 11.45 \times B^2 - 0.62 \times C^2 \end{aligned} \quad (4)$$

$$\begin{aligned} \text{TRS yield} = & 305.8687 + 137.0487 \times \text{Acid} + 2.1816 \times \text{Time} \\ & - 2.4166 \times \text{Solids} - 0.5917 \times \text{Acid} \times \text{Time} \\ & + 0.3231 \times \text{Acid} \times \text{Solids} \\ & + 0.0133 \times \text{Time} \times \text{Solids} - 384832 \\ & \times \text{Acid}^2 - 0.0154 \times \text{Time}^2 - 0.0097 \times \text{Solids}^2 \end{aligned} \quad (5)$$

where A, B, and C are catalyst concentration (NaOH for AL and H₂SO₄ for DA), retention time and solids loading, respectively. An analysis of variance (ANOVA) was performed to test the significance of the developed model and the results are presented for AL and DA pretreatment in Tables 3 and 4, respectively. If a *p*-value (also known as the Prob > *D*-value) is lower than 0.05 a model is considered significant, indicating only a 5% chance that their respective model could occur due to noise. For both pretreatments, their models effectively describes the response nevertheless the AL pretreatment model have a lower *p*-value (0.0003) than the DA pretreatment model (0.0247). In addition, the Prob > *F* values for each model term in AL pretreatment suggest that A, C, and A², meanwhile for DA pretreatment suggest that only A² are the model terms that have significant effects on the TRS yield. To determine the suitability of the model, the lack of fit test was used, which indicated an insignificant lack of fit with an *F*-value of 0.1393 and 0.3009 for AL and DA pretreatment, respectively. The coefficient of determination (*R*²) of the pretreatment models was 0.9151 for AL and 0.7270 and for DA, implying a good and average correlation between the observed and predicted values of AL and DA respectively, as shown in Figures 1A,B. Finally, the quadratic models developed for AL and DA pretreatment are appropriate for predicting TRS yield under different pretreatment conditions within the range used in the present study.

Effect of Pretreatment Conditions on Solids Recovery

The highest solids recovery for AL and DA was obtained in the same run (13) with 87.6 and 86.1%, respectively, with experimental conditions of 0.73% catalyst concentration, 74.8 min and 24.53% solids loading. On the other hand, the lowest solids recovery for AL pretreatment of 60.7% was obtained during run 8

TABLE 3 | ANOVA table for the quadratic model of alkaline pretreatment.

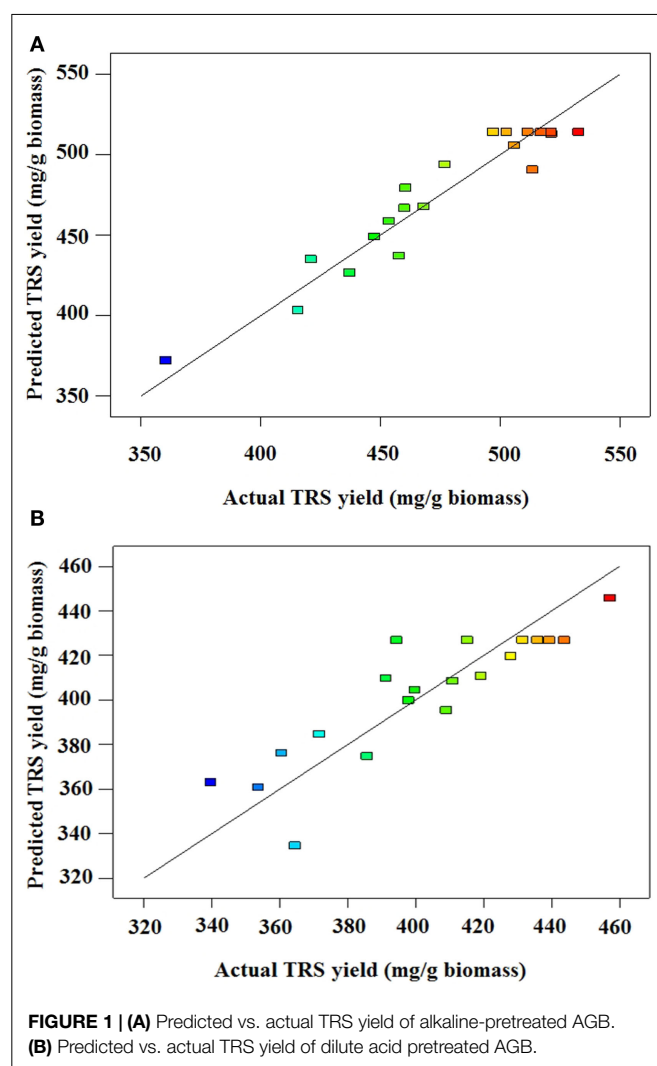
Source	Sum of squares	DF	Mean square	F-value	Prob > F	
Model	34291.73	9	3810.19	11.97	0.0003	Significant
A	5600.42	1	5600.42	17.59	0.0018	
B	97.31	1	97.31	0.31	0.5925	
C	3578.16	1	3578.16	11.24	0.0073	
AB	528.42	1	528.42	1.66	0.2266	
AC	189.46	1	189.46	0.60	0.4583	
BC	14.85	1	14.85	0.047	0.8333	
A ²	21479.83	1	21479.83	67.48	<0.0001	
B ²	40.82	1	40.82	0.13	0.7277	
C ²	4734.21	1	4734.21	14.87	0.0032	
Residual	3183.28	10	318.33			
Lack of fit	2351.53	5	470.31	2.83	0.1393	Not significant
Pure error	831.75	5	166.35			

DF, degree of freedom.

TABLE 4 | ANOVA table for the quadratic model of dilute acid pretreatment.

Source	Sum of squares	DF	Mean square	F-value	Prob > F	
Model	15632.82	9	1736.98	3.79	0.0247	Significant
A	434.69	1	434.69	0.95	0.3529	
B	0.51	1	0.51	0.001	0.9739	
C	2059.98	1	2059.98	4.50	0.0599	
AB	994.66	1	994.66	2.17	0.1713	
AC	38.45	1	38.45	0.084	0.7779	
BC	45.60	1	45.60	0.100	0.7588	
A ²	10975.31	1	10975.31	23.96	0.0006	
B ²	841.41	1	841.41	1.84	0.2051	
C ²	5.61	1	5.61	0.012	0.9140	
Residual	4580.07	10	458.01			
Lack of fit	2843.15	5	568.63	1.64	0.3009	Not significant
Pure error	1736.92	5	347.38			

DF, degree of freedom.



(3.00% catalyst concentration, 52.5 min and 16.5% solids loading), while for DA pretreatment was 54.4% with run 10 using 2.42% catalyst concentration, 74.8 min and 8.47% solids loading. The difference between low and high solids recovery, which represents

process severity are 26.9 and 31.7% for AL and DA pretreatment, respectively.

Effect of Pretreatment Catalyst Concentration and Retention Time

The effect of catalyst concentration and retention time in AL and DA pretreatment on TRS yield during enzymatic saccharification using 3% biomass loading of are shown in **Figure 2**. By means of pretreatment shorter retention times and catalyst concentration, the TRS yield became lower and the same applies to longer times and high catalyst concentration for both AL and DA pretreatment.

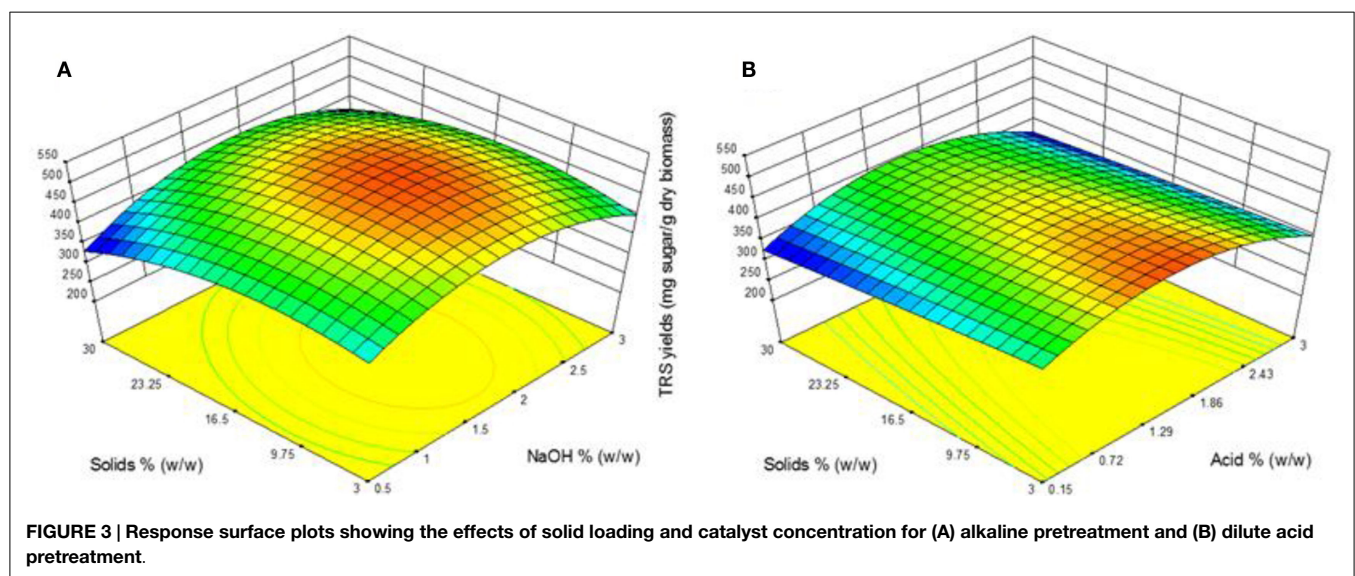
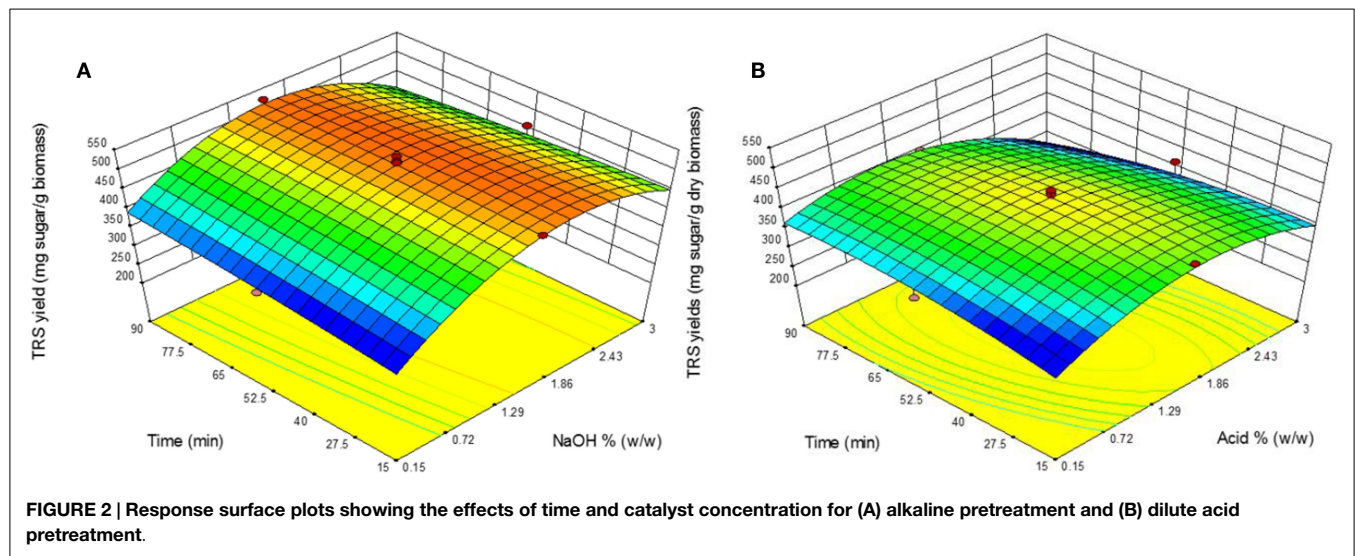
However, for AL pretreatment from 1.58 to 2.43% NaOH a TRS yield above ~460 mg sugar/g biomass is obtained within the study range of 15–90 min. In the other hand, in DA pretreatment a more distributed region is shown where the highest TRS yields was obtained at the central design points with a relatively shorter differences between the highest yield that occurred in run 7 (457 mg/g biomass) and an average of the central data points (433 mg/g biomass).

Effect of Pretreatment Catalyst Concentration and Solid Loading

The response surface plots presents the effect of catalyst concentration and solid loading on TRS yield of both AL and DA pretreatment is displayed in **Figure 3**. One area for AL pretreatment is clearly defined showing the highest TRS yield region in the middle range of both parameters. A TRS yield above 500 mg/g biomass is obtained in the range of 1.1–2.3% NaOH and solid loading between 4 and 20%. These results are supported with previous reports in AL pretreatment where using the same temperature conditions (121°C), moderate NaOH concentration (1%) and time (30–60 min), which achieved the highest TRS yield (Wang et al., 2010; Xu et al., 2010). During DA pretreatment a clear region where a TRS yield above 430 mg/g biomass was reached within the range of 0.7–2% acid and a solid loading of 3–15%. It is noticeable that such differences between the TRS yields were obtained from the highest experimental runs from both pretreatments at ~533 mg/g biomass from run 15 in AL and ~457 mg/g biomass from run 7 in DA. This differences are encounter from the objective of each pretreatment, which in the case of AL pretreatment is lignin removal whereas for DA pretreatment xylan removal is the main effect, as consequence a lower TRS yield should be obtained as there is lower xylan available as a substrate for the enzymes to be reacted into xylose causing a lower total TRS yield.

Optimization of Pretreatment Conditions

In both of the evaluated pretreatment processes (AL and DA), a lower catalyst concentration, shorter time and high solids loading if preferred to obtain an optimum TRS yield. The optimum catalyst concentration, retention time and solid loading were found to be for AL pretreatment of 1.87% NaOH concentration, 50.3 min and 13.1% solids loading, while DA pretreatment were 2.1% acid concentration, 33.8 min and 8.5% solids loading. For AL pretreatment, an 18% increase in NaOH concentration, 4% reduction in retention time and 20% reduction of solids loading, whereas for DA pretreatment, 33% increase in acid concentration, 35.6% reduction in retention time and 283% increase of solids

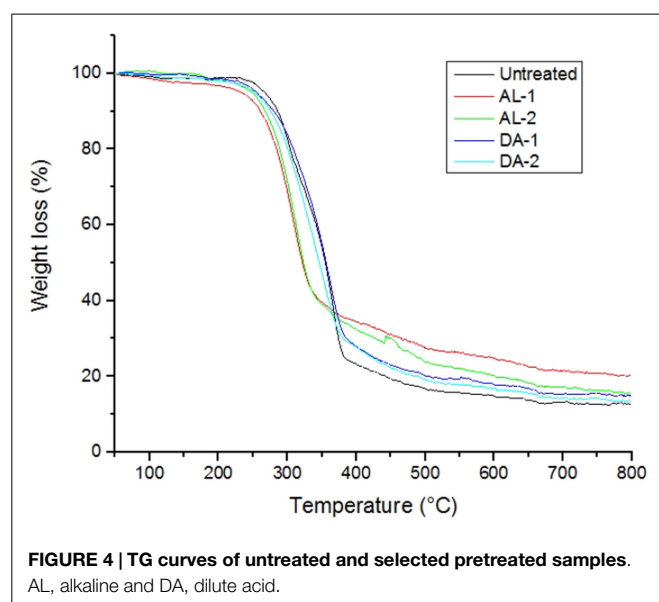


loading and when comparing the optimum conditions with the experimental conditions (Run 7, **Table 2**) that gave the highest yields.

Thermogravimetric and Differential Scanning Calorimetry Analysis

Untreated and selected pretreated AGB samples were thermogravimetrically analyzed to compare degradation characteristics in terms of pretreatment. Two samples were selected for TGA analysis for each pretreatment, named AL-1 and DA-1 corresponding to experimental run 8, in addition to AL-2 and DA-2 corresponding to experimental run 16 (one of the CCD points). **Figure 4** shows standards weight loss plots, while in **Figure 5** the differential TGA plots of the untreated and pretreated AGB samples are shown. All samples exhibit three decomposition regions with some initial weight loss from 50 to 125°C (mainly due to moisture evaporation). Up to 200°C, the samples presented thermal stability. The decomposition temperature (T_d) decrease for

both AL and AL pretreated samples as compared to the untreated AGB, shown in **Table 5**. In both of the analyzed pretreatment the lowest values correspond to AL-1 (run 8 sample). These results indicate that AL pretreatment reduced the activation energy that is needed to decompose the AGB in a higher extent than DA pretreatment by deconstructing the tight plant cell wall structures. AL-pretreated AGB samples obtained a lower T_d value when compared to an ionic liquid treated AGB from a recent report (310 vs. 347°C) (Perez-Pimienta et al., 2015). Thermal depolymerization of hemicelluloses and the cleavage of glycosidic linkages of cellulose occurs in the region of 220–300°C, while lignin decomposition extended to the whole temperature range, from 200 until 700°C, due to different activities of the chemical bonds present on its structure and the degradation of cellulose taken place between 275 and 400°C (Deepa et al., 2011). The final decomposition stage for all samples was completed above 400°C, where a weight loss due to thermolysis of carbon containing residues does take place (Fisher et al., 2002). DSC curves of untreated



AGB and selected samples from AL and DA (Figures S1–S3 in Supplementary Material) with two endothermic peaks observed and Table S1 in Supplementary Material summarizes those events. The first thermal is shown below 200°C with low energy between 5.3 and 13.9 J/g°C, where the untreated AGB present the onset temperature at 83°C (8.6 J/g°C), while the AL-4 (run 16 of AL pretreatment) achieved 13.9 J/g°C, whereas for DA the highest energy event was at 12.2 J/g°C with DA-1 (run 8) that employed a 3% acid loading. A similar peak was obtained with an IL-treated AGB sample where the untreated sample showed a dehydration peak at 89°C (Perez-Pimienta et al., 2015). In the other hand, the second thermal event presents a high energy peak for all samples with ΔH in the range of 120–627 J/g°C and temperature above 262 up to 415°C. AL pretreatment achieved its highest energy with run 16 (AL-4) with a peak at 335°C (627 J/g°C), whereas the evaluated DA-pretreated samples was with run 9 (AL-2) at 358°C and 296 J/g, so when compared to the untreated sample it is clear that a pretreated offers a reduction in terms of calorific value turning them into a more digestible biomass.

Scanning Electron and Confocal Fluorescence Microscopy

The SEM images of untreated and pretreated samples (run 16 sample for both AL and DA pretreatment) were taken at 500× (Figure 6). Untreated AGB (Figure 6A) presents an intact structure without degradation, otherwise AL pretreatments dissolves lignin disrupting the biomass, besides of the increase of pore quantity as can be observed in Figure 6B. Finally, DA pretreatment disrupts the lignocellulosic structure by mainly dissolving hemicellulose, hence, major microfibrillar cellulose structures remain (Figure 6C) and some lignin or lignin-carbohydrate complexes may be condensed on the surface of the cellulose fibers.

Elements content of untreated and pretreated AGB (run 16 from AL and DA pretreatment) are presented in Table 6. In the untreated AGB, C and O accounts for a 98.5% of the totals mass fraction remaining only 1.4% of Ca, these attributable to calcium

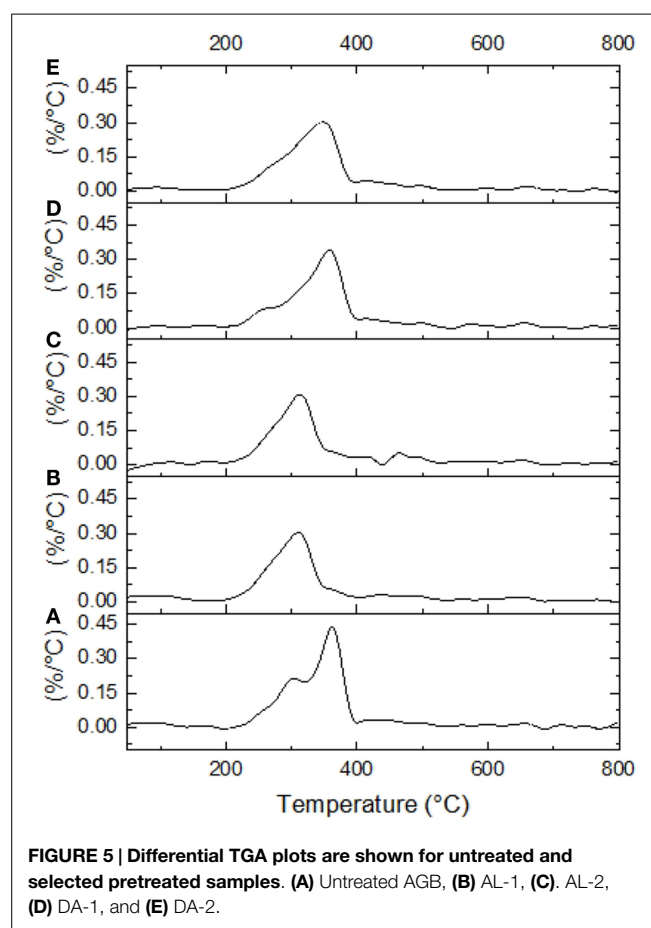


TABLE 5 | Decomposition T_d temperatures for untreated and pretreated AGB.

Property	Pretreatment				
	Untreated	AL-1	AL-2	DA-1	DA-2
T_d (°C)	366	310	317	360	353

oxalate (CaC_2O_4) crystals in considerable quantities along the surface of the plant cell wall as referred in a previous paper (Perez-Pimienta et al., 2015). In contrast, the DA-treated AGB the available Ca was removed during the process at these conditions (1.58% acid concentration, 52.5 min and 16.5% solids loading). Nonetheless, this Ca removal does not occurred in the AL-treated sample where a small amount of Na (1.2%) was found, possibly, as a result of some of the alkali was converted to irrecoverable salts and/or incorporated into the biomass.

Confocal fluorescence microscopy was used to investigate the surface morphologies of untreated and pretreated AGB (run 16 from AL and DA pretreatment) as presented in Figures 7A–F. When compared to the untreated AGB, only the DA-pretreated sample show a significant reduction in the fluorescence signal intensity in cell walls (lignin is represented with a green signal and cellulose with a blue signal), while the AL-treated sample presents only a slight reduction.

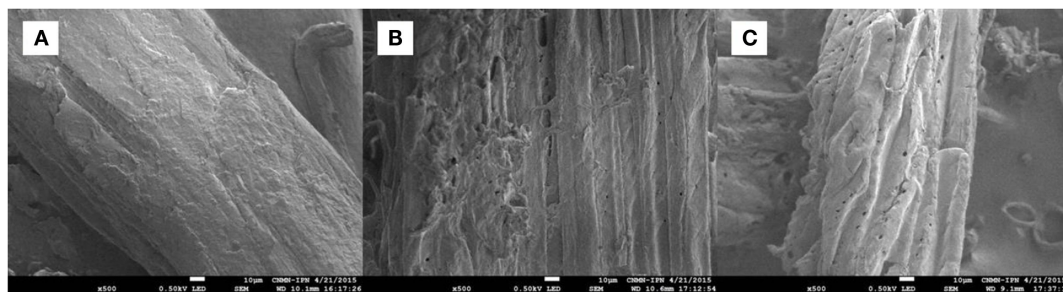


FIGURE 6 | SEM images of AGB samples: (A) untreated, (B) alkaline pretreated, and (C) dilute acid pretreated.

TABLE 6 | Elements content of untreated and pretreated agave bagasse measured by EDS spectroscopy.

Element	Untreated		AL		DA	
	Mass fraction (%)	Atomic mass fraction (%)	Mass fraction (%)	Atomic mass fraction (%)	Mass fraction (%)	Atomic mass fraction (%)
C	51.1 ± 0.9	58.6 ± 0.7	51.7 ± 1.7	59.4 ± 1.7	60.6 ± 5.6	67.2 ± 5.2
O	47.5 ± 0.5	40.9 ± 0.6	45.2 ± 2.2	39.0 ± 2.1	39.4 ± 5.6	32.8 ± 5.2
Ca	1.4 ± 0.4	0.5 ± 0.1	1.9 ± 0.9	1.2 ± 0.5	–	–
Na	–	–	1.2 ± 0.7	0.4 ± 0.3	–	–
Total	100.0	100.0	100.0	100.0	100.0	100.0

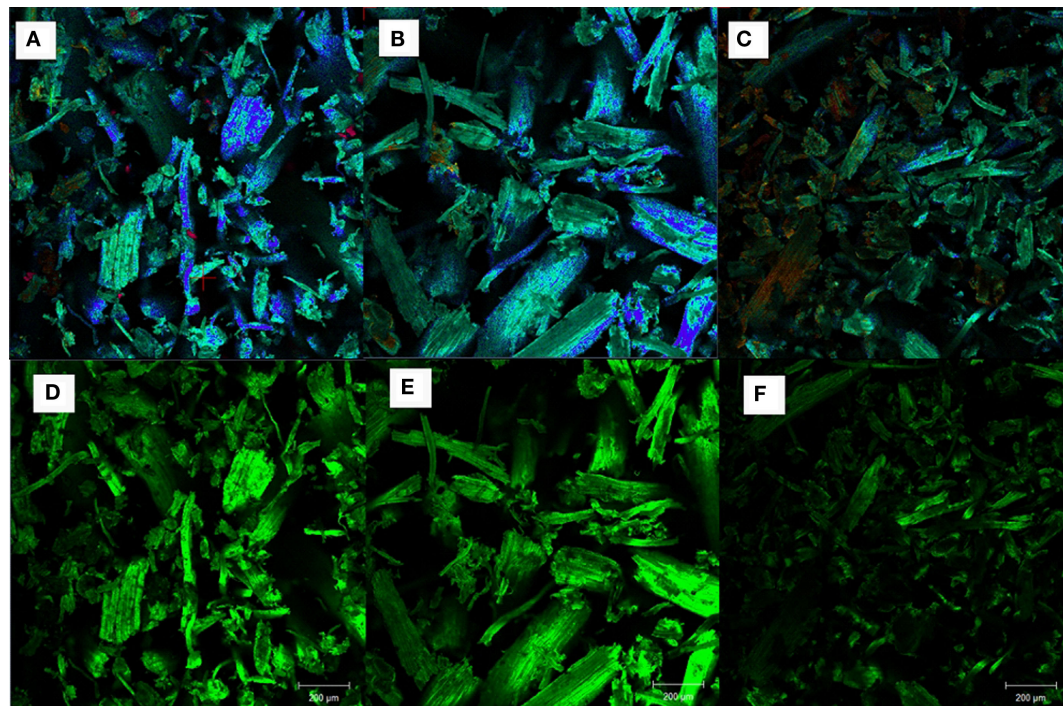


FIGURE 7 | Confocal fluorescence images of AGB samples: (A,D) untreated, (B,E) alkaline pretreated, and (C,F) dilute acid pretreated.

Effect of Pretreatment on Biomass Porosimetry

Pretreatment can affect the cellulose accessibility and is often accompanied by variation in the surface area. Surface area, pore volume, and pore average diameter were measured using the Brunauer–Emmett–Teller (BET) method by argon adsorption, which relates the gas pressures to the volume of gas adsorbed,

although might not be directly associated to enzyme accessibility since the size differences between argon molecules and enzymes (Li et al., 2013). **Table 7** summarizes surface area, pore volume and pore average diameter of untreated and run 16 (one of the CCD points from both AL and DA-pretreated AGB). When compared to the untreated samples an increment in the surface

TABLE 7 | Comparison of porosimetry parameters in untreated and pretreated AGB.

	Surface area (m ² /g)	Pore volume (cm ³ /g)	Pore average diameter (Å)
Untreated	0.6	0.0020	137.7
AL	0.9	0.0023	107.7
DA	1.1	0.0028	106.7

area is noticeable from 0.6 up to 1.1 m²/g. This is consistent with the changes in the SEM images upon AL and DA pretreatment described above. However, the pore volume of all samples (untreated and pretreated) presents a negligible difference close to 0.0008 cm³/g, whereas a reduction in the pore average diameter is obtained in the pretreated samples.

Conclusion

The effects of catalyst concentration, retention time and solids loading in terms of TRS yield of AL and DA pretreatment in AGB were investigated. This study demonstrated that AGB is a promising biofuel feedstock that can achieved high sugar yields using both DA and AL pretreatment. For both pretreatments, a model was generated with a high correlation obtained from actual

TRS data. Furthermore, the results indicate that TRS yield was enhanced by catalyst concentration and solid loading, but longer retention times does not. Both pretreatment increase porosity and surface area, but AL pretreatment achieved a lower decomposition temperature. Finally, RSM was also used to optimize the pretreatment conditions for maximum TRS yield. The optimum conditions were determined for AL pretreatment: 1.87% NaOH concentration, 50.3 min, and 13.1% solids loading, whereas DA pretreatment: 2.1% acid concentration, 33.8 min, and 8.5% solids loading. Finally, fuel synthesis studies should be performed in the sugars obtained using the best conditions for both pretreatments in order to obtain significant data for a scale-up process.

Acknowledgments

The authors gratefully thank CONACYT through the project 229711 and internal funding from Universidad Autónoma de Nayarit (Autonomous University of Nayarit).

Supplementary Material

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00146>

References

- Avci, A., Saha, B. C., Dien, B. S., Kennedy, G. J., and Cotta, M. A. (2013). Response surface optimization of corn stover pretreatment using dilute phosphoric acid for enzymatic hydrolysis and ethanol production. *Bioresour. Technol.* 130, 603–612. doi:10.1016/j.biortech.2012.12.104
- Avira, P., Tomás-Pejó, E., Ballesteros, M., and Negro, M. J. (2010). Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresour. Technol.* 101, 4851–4861. doi:10.1016/j.biortech.2009.11.093
- Caspeta, L., Caro-Bermúdez, M. A., Ponce-Noyola, T., and Martinez, A. (2014). Enzymatic hydrolysis at high-solids loading for the conversion of agave bagasse to fuel ethanol. *Appl. Energy* 113, 277–286. doi:10.1016/j.apenergy.2013.07.036
- Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2, 121–145. doi:10.1146/annurev-chembioeng-061010-114205
- da Costa Sousa, L., Chundawat, S. P. S., Balan, V., and Dale, B. E. (2009). 'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies. *Curr. Opin. Biotechnol.* 20, 339–347. doi:10.1016/j.copbio.2009.05.003
- Dadi, A. P., Varanasi, S., and Schall, C. A. (2006). Enhancement of cellulose saccharification kinetics using an ionic liquid pretreatment step. *Biotechnol. Bioeng.* 95, 904–910. doi:10.1002/bit.21047
- Davis, S. C., Dohleman, F. G., and Long, S. P. (2011). The global potential for agave as a biofuel feedstock. *GCB Bioenergy* 3, 68–78. doi:10.1111/j.1757-1707.2010.01077.x
- Deepa, B., Abraham, E., Cherian, B. M., Bismarck, A., Blaker, J. J., Pothan, L. A., et al. (2011). Structure, morphology and thermal characteristics of banana nano fibers obtained by steam explosion. *Bioresour. Technol.* 102, 1988–1997. doi:10.1016/j.biortech.2010.09.030
- Fisher, T., Hajaligol, M., Waymack, B., and Kellogg, D. (2002). Pyrolysis behavior and kinetics of biomass derived materials. *J. Anal. Appl. Pyrolysis* 62, 331–349. doi:10.1016/S0165-2370(01)00129-2
- Hernández-Salas, J. M., Villa-Ramírez, M. S., Veloz-Rendón, J. S., Rivera-Fernández, K. N., González-César, R. A., Plascencia-Espinosa, M. A., et al. (2009). Comparative hydrolysis and fermentation of sugarcane and agave bagasse. *Bioresour. Technol.* 100, 1238–1245. doi:10.1016/j.biortech.2006.09.062
- Kumar, P., Barrett, D., Delwiche, M. J., and Stroeve, P. (2009). Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Ind. Eng. Chem. Res.* 48, 3713–3729. doi:10.1021/ie801542g
- Li, C., Tanjore, D., He, W., Wong, J., Gardner, J. L., Sale, K. L., et al. (2013). Scale-up and evaluation of high solid ionic liquid pretreatment and enzymatic hydrolysis of switchgrass. *Biotechnol. Biofuels* 6, 154. doi:10.1186/1754-6834-6-154
- Mielenz, J. R., Rodriguez, M., Thompson, O. A., Yang, X., and Yin, H. (2015). Development of Agave as a dedicated biomass source: production of biofuels from whole plants. *Biotechnol. Biofuels* 8, 79. doi:10.1186/s13068-015-0261-8
- Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428. doi:10.1021/ac60147a030
- Modenbach, A. A., and Nokes, S. E. (2012). The use of high-solids loadings in biomass pretreatment – a review. *Biotechnol. Bioeng.* 109, 1430–1442. doi:10.1002/bit.24464
- Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., et al. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* 96, 673–686. doi:10.1016/j.biortech.2004.06.025
- Perez-Pimienta, J. A., Lopez-Ortega, M. G., Chavez Carvayar, J. A., Varanasi, P., Stavila, V., Cheng, G., et al. (2015). Characterization of agave bagasse as a function of ionic liquid pretreatment. *Biomass Bioenergy* 75, 180–188. doi:10.1016/j.biombioe.2015.02.026
- Perez-Pimienta, J. A., Lopez-Ortega, M. G., Varanasi, P., Stavila, V., Cheng, G., Singh, S., et al. (2013). Comparison of the impact of ionic liquid pretreatment on recalcitrance of agave bagasse and switchgrass. *Bioresour. Technol.* 127, 18–24. doi:10.1016/j.biortech.2012.09.124
- Sathitsuksanoh, N., Xu, B., Zhao, B., and Zhang, Y. H. P. (2013). Overcoming biomass recalcitrance by combining genetically modified switchgrass and cellulose solvent-based lignocellulose pretreatment. *PLoS ONE* 8:e73523. doi:10.1007/s10295-010-0853-z
- Saucedo-Luna, J., Castro-Montoya, A. J., Martinez-Pacheco, M. M., Sosa-Aguirre, C. R., and Campos-García, J. (2011). Efficient chemical and enzymatic saccharification of the lignocellulosic residue from Agave tequilana bagasse to produce ethanol by *Pichia caribbica*. *J. Ind. Microbiol. Biotechnol.* 38, 725–732. doi:10.1007/s10295-010-0853-z
- Tan, H. T., Lee, K. T., and Mohamed, A. R. (2011). Pretreatment of lignocellulosic plant biomass using a solvent-ionic liquid [BMIM]Cl for glucose recovery: An optimisation study using response surface methodology. *Carbohydr. Polymers* 83, 1862–1868. doi:10.1016/j.carbpol.2010.10.052

- Wang, Z., Keshwani, D. R., Redding, A. P., and Cheng, J. J. (2010). Sodium hydroxide pretreatment and enzymatic hydrolysis of coastal Bermuda grass. *Bioresour. Technol.* 101, 3583–3585. doi:10.1016/j.biortech.2009.12.097
- Xu, J., Cheng, J. J., Sharma-Sivappa, R. R., and Burns, J. (2010). Sodium hydroxide pretreatment of switchgrass for ethanol production. *Energy Fuels* 24, 2113–2119. doi:10.1021/ef9014718
- Zhang, C., Lei, X., Scott, T., Zhu, J. Y., and Li, K. (2014). Comparison of dilute acid and sulfite pretreatment for enzymatic saccharification of earlywood and latewood of Douglas fir. *Bioenergy Res.* 7, 362–370. doi:10.1007/s12155-013-9376-6

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Ávila-Lara, Camberos-Flores, Mendoza-Pérez, Messina-Fernández, Saldaña-Duran, Jimenez-Ruiz, Sánchez-Herrera and Pérez-Pimienta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evaluating lignocellulosic biomass, its derivatives, and downstream products with Raman spectroscopy

Jason S. Lupoi^{1,2}, Erica Gjersing^{1,2} and Mark F. Davis^{1,2}*

¹ Oak Ridge National Laboratory, BioEnergy Science Center, Oak Ridge, TN, USA

² National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO, USA

Edited by:

P. C. Abhilash, Banaras Hindu University, India

Reviewed by:

Zongbao K. Zhao, Dalian Institute of Chemical Physics, China

Yunqiao Pu, Georgia Institute of Technology, USA

*Correspondence:

Mark F. Davis, National Renewable Energy Laboratory, BioEnergy Science Center, 15013 Denver West Parkway, Golden, CO 80401-3305, USA

e-mail: mark.davis@nrel.gov

The creation of fuels, chemicals, and materials from plants can aid in replacing products fabricated from non-renewable energy sources. Before using biomass in downstream applications, it must be characterized to assess chemical traits, such as cellulose, lignin, or lignin monomer content, or the sugars released following an acid or enzymatic hydrolysis. The measurement of these traits allows researchers to gage the recalcitrance of the plants and develop efficient deconstruction strategies to maximize yields. Standard methods for assessing biomass phenotypes often have experimental protocols that limit their use for screening sizeable numbers of plant species. Raman spectroscopy, a non-destructive, non-invasive vibrational spectroscopy technique, is capable of providing qualitative, structural information and quantitative measurements. Applications of Raman spectroscopy have aided in alleviating the constraints of standard methods by coupling spectral data with multivariate analysis to construct models capable of predicting analytes. Hydrolysis and fermentation products, such as glucose and ethanol, can be quantified off-, at-, or on-line. Raman imaging has enabled researchers to develop a visual understanding of reactions, such as different pretreatment strategies, in real-time, while also providing integral chemical information. This review provides an overview of what Raman spectroscopy is, and how it has been applied to the analysis of whole lignocellulosic biomass, its derivatives, and downstream process monitoring.

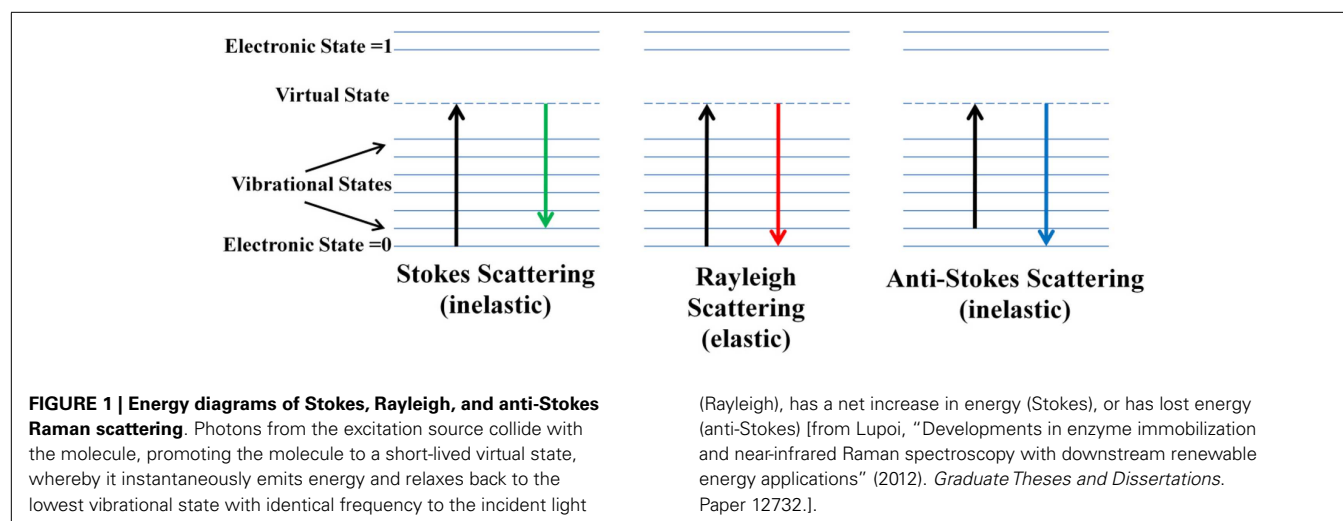
Keywords: Raman spectroscopy, high-throughput, lignin, glucose, xylose, process monitoring, cellulose, ethanol

INTRODUCTION

The production of fuels, chemicals, and materials from plants has offered an opportunity to supplant usage of products fashioned from non-renewable energy sources. Lignocellulosic biomass is predominantly composed of cellulose, non-cellulosic polysaccharides (NCPs), and lignin, and provides a useful starting feedstock for industrial processes. Before a specific plant can be considered for downstream applications, the chemical traits of the biomass must be characterized. These assessments include, but are not limited to, the compositional analysis of the plant's cellulose, NCP, and lignin contents, the ratio of syringyl (S), guaiacyl (G), and *p*-hydroxyphenol (H) lignin monomers, the release of simple sugars following an acid or enzymatic hydrolysis, and the cellulose crystallinity index. Many of these evaluations gage the recalcitrance of the plant cell wall, and enable researchers to develop appropriate pretreatment strategies to deconstruct the biomass (Blanch et al., 2011), or genetic strategies to synthesize a more ideal starting feedstock (Furtado et al., 2014). The standard methods developed for biomass characterization are beneficial for evaluating small sample sets, but specific experimental attributes limit their use for screening large arrays of prospective plants to isolate those possessing quintessential traits for biofuel and/or biomaterial production. These attributes include laborious sample preparation protocols [derivatization of samples in gas chromatography (GC) analysis and sample clean-up for liquid chromatography or

GC], use of toxic reagents that may require remediation (acetyl bromide, boron trifluoride etherate, trifluoroacetic acid, sulfuric acid), long analysis times [chromatography, nuclear magnetic resonance (NMR)], complex data analysis [pyrolysis GC/mass spectrometry (MS) analysis of lignin monomer content], and/or the destruction of the sample (pyrolysis, GC, solution state NMR). In order to circumvent some of these limitations, techniques have been developed that are non-destructive, require little to no sample preparation, and have increased throughput, allowing more plants to be assessed in less time and with reduced experimental costs (Lupoi et al., 2014b).

The phenomenon of Raman scatter was first envisaged theoretically in 1921 by Smekal, and was proven experimentally in 1928 by Raman and Krishnan, as well as Landsberg and Mandelstam (Smekal, 1923; Landsberg and Mandelstam, 1928; Raman and Krishnan, 1928). Raman spectroscopy is a vibrational spectroscopy technique in which the scattered photons, generated during the interaction between light and matter, are measured. While the light source C. V. Raman used was sunlight, modern applications of Raman spectroscopy employ ultraviolet (UV), visible, or near-infrared (NIR) lasers. The scattering produced can have an identical (elastic), higher (inelastic), or lower (inelastic) frequency than that of the excitation source [Figure 1; Lupoi (2012)]. These types of scattering are named Rayleigh, Stokes, and anti-Stokes, respectively (Carey, 1982; McCreery, 2000; Smith and Dent, 2005;



Popp, 2006). Rayleigh scattering is the most intense, and needs to be thoroughly removed from the optical beam path using specialized optics such as holographic notch filters (HNFs) (Smith and Dent, 2005; Dao, 2006). If not eliminated, Rayleigh scattering can lead to saturation of the detector, and can obscure Raman signal from Stokes scattering, a much weaker phenomenon, as only approximately one per one million photons generated lead to this type of inelastic scattering (Smith and Dent, 2005). Stokes scattering is the most common type measured using Raman spectroscopy, and results in an energy shift to higher vibrational levels. Anti-Stokes scattering results in a shift from a higher to lower vibrational levels, and is less common due to the lower probability of molecules populating higher vibrational levels at ambient conditions. An important feature of the Raman phenomenon is that, unlike in infrared (IR) spectroscopy, molecules are promoted to short-lived, virtual vibrational levels (Figure 1). Therefore, matching the excitation frequency to that necessary to promote molecules from the ground state to the first excited vibrational level is not requisite.

A molecule is considered "Raman active" if there is a change in the polarizability of the electron cloud during the interaction of the molecule with light. Vibrational modes including C–C, C=C, C–H, C–O, H–C–C, C–O–H, H–C–H, etc., can be expected in an archetypal Raman spectrum (Wiley and Atalla, 1987; Agarwal and Ralph, 1997; Larsen and Barsberg, 2010; Agarwal et al., 2011). As a rule of thumb, symmetric bonds will have the largest changes in polarizability and therefore the strongest Raman signals. Table 1 lists representative vibrational modes measured in biomass constituents, and their respective band assignments. In contrast to Raman, a change in dipole moment leads to molecules being "IR active" in IR spectroscopy. Therefore, asymmetric bonds have strong peaks in IR spectra. This difference in selection rules signifies that these two techniques provide complementary information. Vibrational modes that are Raman active will not be present or have small contributions in IR spectra, and vice versa. If a molecule has a center of symmetry, the principle of mutual exclusion states that the molecule will either be IR or Raman active. Some non-centrosymmetric molecules, such as those possessing C_1 symmetry, and hence no symmetry, can have both IR and

Raman active vibrational modes (Ingle and Crouch, 1988). Examples of these types of molecules include isopropyl alcohol, propylene glycol, and 2-butanol (National Institute of Standards and Technology (NIST), 2013). The diatomic nitric oxide is another molecule that, although it produces only one peak, gives rise to IR and Raman active modes, since there is both a change in dipole and polarizability (Smith and Dent, 2005). Another significant difference between the two techniques is the ability of Raman spectroscopy to be used for measuring aqueous and biological samples, whereas IR spectra are appreciably hindered by the presence of water. Lastly, Raman spectra are often less complex than IR spectra due to the diminished signals of overtone and combination vibrational modes, leading to more spectrally resolved peaks.

Typical Raman instruments are composed of the excitation source (i.e., lasers), beam-steering optics to focus the incident light onto the sample, collect the generated scattering, and guide the scattering to the entrance slit of a spectrometer (McCreery, 2000; Dao, 2006; Meyer et al., 2011). HNFs block Rayleigh scattering from entering the spectrometer. Dispersive instruments utilize gratings to diffract the scattering. The diffracted light is projected through the exit slit, and onto a detector. Various detectors, such as the charge-couple device or indium gallium arsenide NIR detector convert light into an electronic response, producing the Raman spectrum. The selection of the excitation wavelength is critical to obtaining quality Raman spectra, as the Raman intensity is directly proportional to the incident frequency to the fourth power. Therefore, when employing UV or visible lasers, the Raman signal, in theory, will be more intense. This is not often achieved in practice when using visible excitation to measure biomass, as the intrinsic fluorescence generated from plants can significantly conceal the Raman signal. Higher energy UV lasers require some precautions to be taken, such as the prevention of sample degradation, which can be achieved by equipping the instrument with a rotating stage, and the enactment of adequate safety strategies. Researchers have developed various techniques to combat the difficulty in obtaining Raman spectra from highly fluorescent molecules like lignin. These methods have typically included the employment of NIR excitation sources, such as 1064 nm neodymium-doped yttrium

Table 1 | Vibrational modes and band assignments measured in lignocellulosic biomass.

Vibrational mode cm ⁻¹	Biomass constituent	Assignment
311	Cellulose	CCO torsion (Schenzel and Fischer, 2001)
329	Cellulose	CCC, CO, CCO, ring deformation (Schenzel and Fischer, 2001)
331	Cellulose	Heavy atom bending (Wiley and Atalla, 1987)
337	Lignin	Ring deformation, OH torsion (Larsen and Barsberg, 2010)
344	Cellulose	Heavy atom stretching (Wiley and Atalla, 1987), CCC, CO, CCO, ring deformation (Schenzel and Fischer, 2001)
352	Cellulose	CCC, CO, CCO, ring deformation; cellulose II (Schenzel and Fischer, 2001)
357	Lignin	Aromatic ring substituents torsion (Larsen and Barsberg, 2010)
369	Lignin	Skeletal deformation (Agarwal and Atalla, 2010); COC symmetric, in-plane bend (Larsen and Barsberg, 2010)
373	Cellulose	CCC, CO, CCO, ring deformation; cellulose II (Schenzel and Fischer, 2001)
380	Cellulose	Heavy atom stretching (Wiley and Atalla, 1987), CCC, CO, CCO, ring deformation (Schenzel and Fischer, 2001)
418	Cellulose	CCC, CCO, ring deformation; cellulose II (Schenzel and Fischer, 2001)
437	Cellulose	Heavy atom stretching (Wiley and Atalla, 1987), CCC, CCO, ring deformation (Schenzel and Fischer, 2001)
458	Cellulose	Heavy atom stretching (Wiley and Atalla, 1987); CCC, CCO, ring deformation (Schenzel and Fischer, 2001)
468	Lignin	Ring deformation (Larsen and Barsberg, 2010)
503	NCSPs	CCO (Kacurikova et al., 1998)
519	Cellulose	Heavy atom stretching (Wiley and Atalla, 1987)
521	NCSPs	CCO (Kacurikova et al., 1998)
531	Lignin	Skeletal deformation (Agarwal and Atalla, 2010)
	NCSPs	CCO (Kacurikova et al., 1998)
553	NCSPs	CCO (Kacurikova et al., 1998)
559	Lignin	CCO and CCC in-plane bend (Larsen and Barsberg, 2010)
582	Lignin	Ring deformation (Larsen and Barsberg, 2010)
597	Lignin	Skeletal deformation (Agarwal and Atalla, 2010)
609	Cellulose	CCH torsion (Schenzel and Fischer, 2001)
637–644	Lignin	Ring and skeletal deformation (Agarwal, 1999; Larsen and Barsberg, 2010)
645	β-Pinene	Ring deformation (Schulz and Baranska, 2007)
652	1,8-Cineol (eucalyptol)	Ring deformation (Schulz and Baranska, 2007)
666	α-Pinene	Ring deformation (Schulz and Baranska, 2007)
712	Lignin	Ring deformation; CC stretch (Larsen and Barsberg, 2010)
780	Lignin	CO stretch (Larsen and Barsberg, 2010); lignin aromatic skeletal vibrations (Agarwal, 1999; Larsen and Barsberg, 2010)
793	Lignin	Out-of-plane CH bend; ring deformation; CO stretch (Larsen and Barsberg, 2010)
799	Lignin	Ring deformation; CO stretch (Larsen and Barsberg, 2010)
805	Lignin	CO stretch; aryl symmetric CH bend; CH out-of-plane bend (Larsen and Barsberg, 2010)
829	Lignin	CH out-of-plane bend (Larsen and Barsberg, 2010)
843	Lignin	Breathing mode (Larsen and Barsberg, 2010)
854	Pectin	(C–O–C) Skeletal mode of α-anomers (Schulz and Baranska, 2007)
885	Cellobiose	(C–O–C) Skeletal mode (Schulz and Baranska, 2007)
897	Cellulose, NCSPs	CH, ring (Kacurikova et al., 1998), C1–H (Schenzel and Fischer, 2001)
904	Cellulose, NCSPs	CH, ring (Kacurikova et al., 1998; Schenzel and Fischer, 2001)
910	Cellulose	HCC and HCO bending (Wiley and Atalla, 1987)
920–932	Lignin	CCH wag (Agarwal, 1999)

(Continued)

Table 1 | Continued

Vibrational mode cm^{-1}	Biomass constituent	Assignment
921	Lignin	Ring deformation; in-plane CC stretch; COC stretch (Larsen and Barsberg, 2010)
942	Lignin	Lignin CCH wag; aromatic skeletal vibrations (Larsen and Barsberg, 2010; Agarwal et al., 2011)
969	Cellulose	CC and CO stretching (Wiley and Atalla, 1987), CH_2 (Schenzel and Fischer, 2001)
977	Lignin	Lignin CCH and $-\text{HC}=\text{CH}-$ deformation; methyl wagging (Larsen and Barsberg, 2010; Agarwal et al., 2011)
	Cellulose	CO (Kacurikova et al., 1998)
995	Cellulose	CC and CO stretching (Wiley and Atalla, 1987)
997	Cellulose	CH_2 (Schenzel and Fischer, 2001)
1000–1010	Cellulose, NCSPs	CC and COH (Kacurikova et al., 1998)
1026	NCSPs	CC and COH (Kacurikova et al., 1998)
1033	Lignin	Lignin CH_3 wagging; CH_3 out-of-plane rock; aromatic skeletal vibrations, methoxy vibrations (Larsen and Barsberg, 2010; Agarwal et al., 2011)
1037	Cellulose	CC and CO stretching (Wiley and Atalla, 1987)
1043	Lignin	OC stretch; ring deformation; CH_3 wagging (Larsen and Barsberg, 2010)
1057	Cellulose	CC and CO stretching (Wiley and Atalla, 1987)
1089	NCSPs	COC, C–C, ring vibration (Kacurikova et al., 1998)
1095	Cellulose	CC and CO stretching (Wiley and Atalla, 1987), COC, glycosidic; ring breathing, symmetric (Schenzel and Fischer, 2001)
1101	NCSPs	COC, C–C, ring vibration (Kacurikova et al., 1998)
1106	NCSPs	COC, C–C, ring vibration (Kacurikova et al., 1998)
1115	Cellulose	COC, glycosidic; ring breathing, symmetric; cellulose II (Schenzel and Fischer, 2001)
1117	Lignin	Lignin methoxy vibrations; aryl CH bend (Larsen and Barsberg, 2010)
	Cellulose	CC and CO stretching (Wiley and Atalla, 1987)
1121	Cellulose	CC and CO stretching (Wiley and Atalla, 1987), COC, glycosidic; ring breathing, symmetric (Schenzel and Fischer, 2001)
1125	Cellulose, NCSPs	COC and C–C (Kacurikova et al., 1998)
1130–1136	Lignin	Coniferaldehyde/sinapaldehyde mode (Agarwal and Atalla, 2010)
1142	Cellulose	CC, CO ring breathing, asymmetric; cellulose II (Schenzel and Fischer, 2001)
1147	Lignin	Lignin methoxy vibrations; aromatic CCH bend (Larsen and Barsberg, 2010)
1151	Cellulose	Stretching and HCC and HCO bending (Wiley and Atalla, 1987), CC, CO ring breathing, asymmetric (Schenzel and Fischer, 2001)
1155	Carotenoids	C–C stretch (Schulz and Baranska, 2007)
1169	Lignin	Lignin hydroxyl COH bend; aromatic skeletal vibrations (Larsen and Barsberg, 2010)
1185	Lignin	Lignin methoxy vibrations; COH in-plane bend (Larsen and Barsberg, 2010)
1199	Lignin	In-plane CH stretch (Larsen and Barsberg, 2010)
1202	Lignin	Lignin methoxy vibrations (Larsen and Barsberg, 2010)
	Cellulose	CH_2 , HCC, HOC, COH (Schenzel and Fischer, 2001)
1214–1217	Lignin	Aryl-O of aryl OH and aryl- OCH_3 ; ring deformation (Agarwal, 1999; Larsen and Barsberg, 2010)
1256	Lignin	CO stretch (Larsen and Barsberg, 2010)
1262	Cellulose	CH_2 , HCC, HOC, COH (Schenzel and Fischer, 2001)
1268	Lignin	Lignin aromatic skeletal vibrations; methoxy vibrations (Agarwal, 1999; Agarwal et al., 2011)
1272	Lignin	Ring deformation; CO stretch (Larsen and Barsberg, 2010)
1275	Cellulose	HCC and HCO bending (Wiley and Atalla, 1987)
1288	Lignin	Ring deformation and in-plane CH and COH bend (Larsen and Barsberg, 2010)

(Continued)

Table 1 | Continued

Vibrational mode cm^{-1}	Biomass constituent	Assignment
1291	Cellulose	HCC and HCO bending (Wiley and Atalla, 1987), CH_2 , HCC, HCO, COH (Schenzel and Fischer, 2001)
1298	Lignin	CH, CC stretch; ring deformation (Larsen and Barsberg, 2010)
1312	NCSPs	CH, COH (Kacurikova et al., 1998)
1331	Lignin	Aliphatic OH stretch (Agarwal and Atalla, 2010)
	Cellulose	HCC and HCO bending (Wiley and Atalla, 1987)
1337	Cellulose	HCC, HCO, and HOC bending (Wiley and Atalla, 1987), CH_2 , HCC, HCO, COH (Schenzel and Fischer, 2001)
1365	NCSPs	CH, OH stretch (Kacurikova et al., 1998)
1373	NCSPs	CH_2 , HCC, HCO, COH; cellulose II (Schenzel and Fischer, 2001)
1372–1383	Lignin	Phenolic OH stretch
1376	NCSPs	CH, OH stretch (Kacurikova et al., 1998)
1378	Cellulose	HCC, HCO, and HOC bending (Wiley and Atalla, 1987; Schenzel and Fischer, 2001), CH_2 (Schenzel and Fischer, 2001)
1378–1390	Lignin	Phenolic OH (Agarwal, 1999)
1380	Lignin	Umbrella CH bend (Larsen and Barsberg, 2010)
1407	Cellulose	HCC, HCO, and HOC bending (Wiley and Atalla, 1987; Schenzel and Fischer, 2001), CH_2 (Schenzel and Fischer, 2001)
1427	Lignin	Lignin methoxy deformation, methyl bending, aromatic skeletal vibrations (Agarwal, 1999; Agarwal et al., 2011)
1455	Lignin	CH_3 scissoring; CH_3 out-of-plane bend; umbrella bend (Larsen and Barsberg, 2010)
1456	Cellulose	HCH and HOC bending (Wiley and Atalla, 1987; Schenzel and Fischer, 2001), CH_2 scissoring (Schenzel and Fischer, 2001)
1460	Lignin	Lignin methoxy deformation, CH_2 scissoring (Agarwal and Atalla, 2010)
1461	Cellulose	CH_2 scissoring (Schenzel and Fischer, 2001)
1465	Lignin	CH_3 scissoring; CH_3 out-of-phase bend (Larsen and Barsberg, 2010)
1475	Cellulose	HCH and HOC bending (Wiley and Atalla, 1987), CH_2 scissoring (Schenzel and Fischer, 2001)
1506–1514	Lignin	Aryl ring stretch, asymmetric (Agarwal, 1999)
1517–1521	Lignin	Asymmetric aryl ring stretch (Agarwal, 1999)
1528	Carotenoids	$\text{C}=\text{C}$ stretch (Schulz and Baranska, 2007)
1605	Lignin	Lignin aromatic skeletal vibrations (Agarwal, 1999; Larsen and Barsberg, 2010)
1632	Lignin	Lignin $\text{C}=\text{C}$ stretch of coniferaldehyde, sinapaldehyde, phenolic esters (Agarwal, 1999; Agarwal et al., 2011)
1643	α -Pinene	$\text{C}=\text{C}$ stretch (Schulz and Baranska, 2007)
1656	Lignin	Lignin $\text{C}=\text{C}$ stretch of coniferyl alcohol and sinapyl alcohol (Agarwal, 1999; Agarwal et al., 2011)
1659	β -Pinene	$\text{C}=\text{C}$ stretch (Schulz and Baranska, 2007)
1704	Lignin	Carbonyl stretch
1739	NCSPs	$\text{C}=\text{O}$ stretch (Schulz and Baranska, 2007)
1745	Pectin	$\text{C}=\text{O}$ stretch (Schulz and Baranska, 2007)
2842	Lignin	Out-of-plane CH symmetric stretch (Larsen and Barsberg, 2010)
2845	Lignin	CH stretch in OCH_3 , symmetric (Agarwal and Atalla, 2010)
2853	Cellulose	CH_2 symmetric stretch (Liang and Marchessault, 1959)
2859	Lignin	Out-of-plane symmetric CH stretch (Larsen and Barsberg, 2010)
2866	Cellulose	CH and CH_2 stretching (Wiley and Atalla, 1987)
2867	Lignin	Out-of-plane symmetric CH stretch (Larsen and Barsberg, 2010)
2889	Cellulose	CH and CH_2 stretching (Wiley and Atalla, 1987)

(Continued)

Table 1 | Continued

Vibrational mode cm^{-1}	Biomass constituent	Assignment
2917	NCSPs	CH stretch (Kacurikova et al., 1998)
2922	Lignin	Out-of-plane asymmetric CH stretch (Larsen and Barsberg, 2010)
2933	NCSPs	CH stretch (Kacurikova et al., 1998)
2939	Lignin	CH stretch in OCH_3 , asymmetric (Agarwal and Atalla, 2010; Larsen and Barsberg, 2010)
2943	Cellulose	CH and CH_2 stretching (Wiley and Atalla, 1987)
2963	Cellulose	CH and CH_2 stretching (Wiley and Atalla, 1987)
3005	Lignin	In-plane CH stretch (Larsen and Barsberg, 2010)
3014	Lignin	In-plane CH stretch (Larsen and Barsberg, 2010)
3039	Lignin	In-plane CH stretch (Larsen and Barsberg, 2010)
3062	Lignin	In-plane CH stretch (Larsen and Barsberg, 2010)
3071	Lignin	Aromatic CH (Agarwal and Atalla, 2010)
3286	Cellulose	OH stretch (Wiley and Atalla, 1987)
3335	Cellulose	OH stretch (Wiley and Atalla, 1987)
3363	Cellulose	OH stretch (Wiley and Atalla, 1987)
3402	Cellulose	OH stretch (Wiley and Atalla, 1987)

NCSPs, non-cellulosic structural polysaccharides.

orthovanadate or neodymium-doped yttrium aluminum garnet lasers (Agarwal et al., 2011; Meyer et al., 2011; Lupoi and Smith, 2012). NIR lasers, having the longest wavelength, lead to diminished spectral intensities. Conversely, since fluorescence emission maxima occur at lower wavelength regions, the employment of NIR excitation can result in significantly reduced spectral background. As an example, the use of a 785 nm laser, juxtaposed to a 1064 nm laser, will produce 3.8-times more scattering (Meyer et al., 2011). The analysis of a lignin sample using both excitations, however, revealed a background that was 160-times higher when employing the higher frequency 785 nm laser (Meyer et al., 2011). Most of these applications utilizing NIR lasers have been Fourier-transform Raman (FT-Raman) spectroscopy experiments. However, instrumental advances, such as better detectors for NIR wavelengths, have enabled NIR, dispersive Raman spectroscopy to provide a lower cost alternative to FT-Raman systems (Chase and Talmi, 1991; Lewis et al., 1993; Barbillat and Da Silva, 1997). Other instrumental methods like coherent anti-Stokes Raman scattering (CARS) and stimulated Raman scattering (SRS) spectroscopies have also provided fluorescence free Raman spectra (Saar et al., 2010; Zeng et al., 2012; Pohling et al., 2014).

Due to the complex composition of biomass, Raman spectra should be prudently interpreted. There can be significant spectral overlap between vibrational modes, challenging a routine spectral assignment of peaks. Cellulose and hemicellulose are structurally similar, and therefore, exhibit comparable Raman spectra. Subtle differences due exist, however, and quantitation may require the use of minor, rather than the most intense peaks (Shih et al., 2011). Hemicelluloses, due to their disorder and complexity, typically result in broader Raman bands than cellulose (Gierlinger and Schwanninger, 2006). Raman vibrational modes of

cellulose are strongly affected by crystallinity and fiber orientation, enabling studies of cellulose polymorphs (Schenzel and Fischer, 2001). The dominant lignin vibrational mode near 1600 cm^{-1} is assigned to ring breathing, and therefore, is comprised of any phenyl-containing molecules, like flavonoids. If a biomass sample has a high extractable content, i.e., herbaceous feedstocks, the 1600 cm^{-1} peak will include contributions from lignin and other extractable molecules (Lupoi and Smith, 2012). Studies on lignin, therefore, require the efficient removal of extraneous species. Additionally, the 1600 cm^{-1} lignin peak contains overlapping signals from S, G, and H lignin monomers, complicating quantitative or semi-quantitative analyses between different biomass species (Lupoi and Smith, 2012). If the ratio of the monomers is known and does not significantly change between samples, and the samples have been exhaustively extracted, the 1600 cm^{-1} mode may be useful for evaluating relative lignin contents within feedstocks.

DISPERSIVE RAMAN SPECTROSCOPY

As previously mentioned, NIR dispersive Raman spectroscopy can provide a suitable, less costly alternative to FT-Raman spectroscopy. Despite this, there are relatively few instances of researchers using this instrumental configuration (Roder and Sixta, 2005; Shih and Smith, 2009; Li et al., 2010, 2011, 2013; Meyer et al., 2011; Shih et al., 2011; Zakzeski et al., 2011; Lupoi and Smith, 2012; Ewanick et al., 2013; Gray et al., 2013; Azimvand, 2014; Iversen et al., 2014). A comparison between 785 and 1064 nm excitation sources revealed the latter to provide better signal-to-noise (S/N) when measuring hydrolytic lignin using home-built Raman spectrometers (Figure 2) (Meyer et al., 2011). The spectrum generated using the 785 nm laser exhibited a broad, featureless fluorescence background (Figure 3). The fluorescence

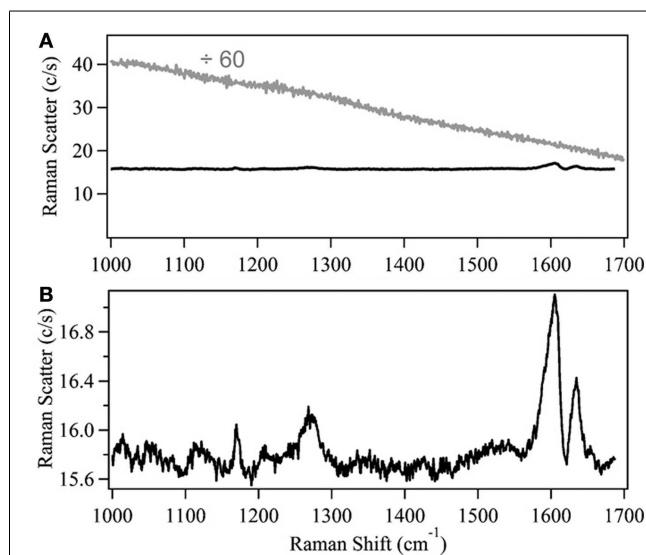
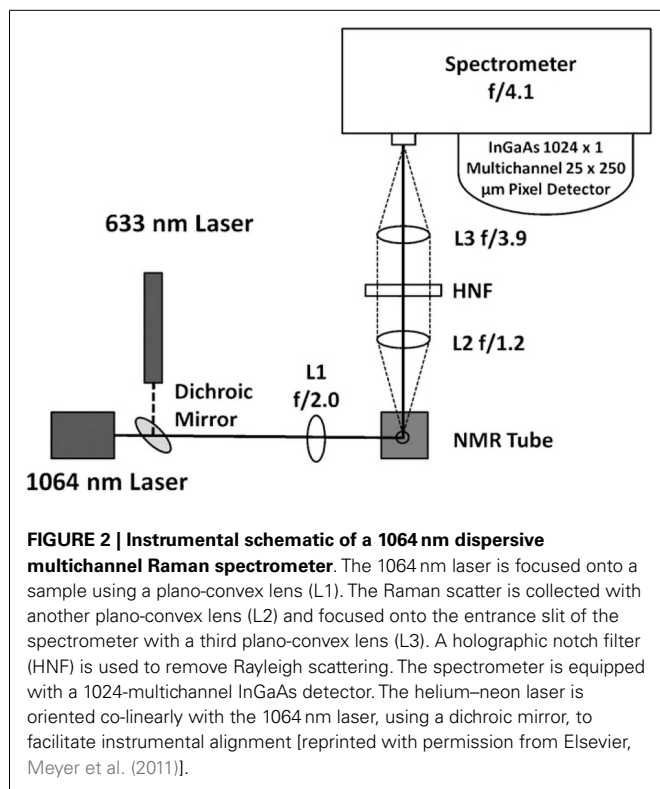


FIGURE 3 | Comparison of 785 and 1064 nm excitation wavelengths to evaluate lignin. (A) Background-subtracted Raman spectra of 50 mg/mL lignin, dissolved in methanol, obtained using a dispersive 785 nm (gray) or 1064 nm (black) spectrometer. The 785 nm excitation spectrum has been divided by 60. (B) The 1064 nm excitation lignin spectrum, plotted on a smaller scale to elucidate spectral features [reprinted with permission from Elsevier, Meyer et al. (2011)].

emission peak maximum is expected to be in the visible region of the electromagnetic spectrum. When excited with the 785 nm light, however, a low intensity peak was detected that resembled

the background measured in the Raman spectrum. Although the intensities of the peaks generated using the 1064 nm laser were relatively weak, the fluorescence was virtually eliminated (**Figure 3**). This instrumental configuration also provided higher S/N when compared to a commercial FT-Raman spectrometer using acquisition times greater than 15 s. The same system was used to develop a principal component regression (PCR) model to predict the S and G lignin content of a diverse assortment of feedstocks, including *Miscanthus*, switchgrass, poplar, and pine (Lupoi and Smith, 2012). The model was constructed from Raman spectral data conjoined with thioacidolysis/GCMS S and G lignin percentages.

The quantitation of glucose, xylose, and ethanol in complex matrices illustrated other novel applications of NIR, dispersive Raman spectroscopy (Shih and Smith, 2009; Shih et al., 2011). Raman methods were juxtaposed to those obtained using UV/visible (UV/VIS) spectrophotometry and headspace-GCMS. The UV/VIS methods required longer sample preparation and incubation steps. The GCMS analysis required the samples to be preheated to promote ethanol into the headspace, and had an experimental run time over 10 min per sample. The Raman measurements required relatively no sample preparation, and the spectral data was obtained using a 200 s acquisition time for glucose and xylose, and 100 s for ethanol. Another interesting feature of this work was the demonstration of the ability to simultaneously quantify glucose and xylose in hydrolyzate liquor using a multi-peak curve fit, with detection limits of 3 ± 2 and 1 ± 1 mg mL⁻¹ for glucose and xylose respectively (**Figure 4**). The authors also evaluated the effects of various biomass pretreatment strategies on the ability to measure glucose. Soaking the biomass in aqueous ammonia or extracting using an aqueous ethanol solution resulted in lower detection limits. An acid pretreatment did not lower the detection limit, indicating that it was likely lignin and/or extractives like non-lignin phenolics that caused the higher spectral background, and thus elevated detection limits. These results clearly demonstrate the competence of Raman spectroscopy to measure hydrolysis and fermentation products rapidly and accurately.

In addition to evaluating samples after the reaction has concluded (off-line), Raman spectroscopy can provide a valuable on-line, process monitoring tool, such as during the fermentation of glucose to ethanol. A fiber optic probe can be inserted directly into the reaction slurry. When glucose solutions were used as the starting feedstock, a partial least squares (PLS) model that married the Raman spectral data to standard HPLC ethanol and glucose measurements revealed correlation coefficients (R^2) of 0.984 for ethanol and 0.92 for glucose (**Figure 5**) and good root mean standard errors of cross validation (RMSECV = 0.41, ethanol; 0.53, glucose) given the concentration range evaluated (Ewanick et al., 2013). When switchgrass hydrolyzate liquor was used as the fermentation feedstock, the measurement of glucose was significantly hindered. The hampered ability to quantify glucose resulted from its low concentration as well as the complex, heterogeneous nature of the hydrolyzate, which likely contained lignin-derived phenolics. The ability to measure the spectra of and use PLS to predict the concentration of ethanol was not impeded ($R^2 = 0.935$, RMSECV = 0.60). A Raman spectrometer equipped with a 993 nm laser and a fiber optic probe enabled the real-time

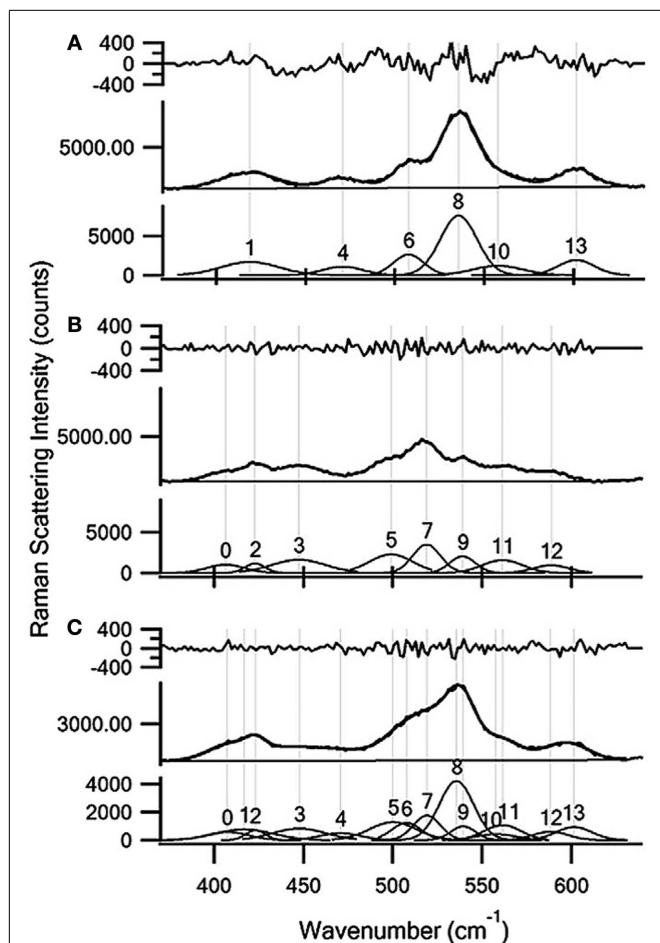


FIGURE 4 | Multi-peak curve fitting results for (A) 100 mg/mL xylose; (B) 120 mg/mL glucose; and (C) a 60 mg/mL glucose and 50 mg/mL xylose mixture in a soaking in aqueous ammonia hydrolysis broth. The residual between the multi-peak fit and the experimental data is shown in the top panel. The residual is of the same order of magnitude as the spectral noise. The experimental spectra are shown in the middle panel (thicker line) and the fit results are shown in the bottom panel (thinner line) [reprinted with permission from Elsevier, Shih et al. (2011)].

study of the formation of a complex assortment of products generated during a simultaneous saccharification and fermentation reaction (Gray et al., 2013). A simple univariate calibration using the 883 cm^{-1} vibrational mode allowed the quantitation of ethanol. The calibration was validated using a separate set of fermentation samples, and exhibited a $R^2 = 0.996$, and a standard error of prediction (SEP) of 0.604. Multivariate PLS calibration models were generated for total starch, dextrins, maltotriose, maltose, glucose, and ethanol using HPLC standard measurements. The percentage error (defined as the SEP/modeling concentration range) was quite low for ethanol (2.1%), starch (2.5%), and dextrin (4.7%) when the calibration sets were broken up into low and high concentration series. The error was approximately two to seven times higher when only one calibration set was employed for these analytes. The percentage errors of glucose, maltose, and maltotriose were 12% or higher. On-line fermentation monitoring has

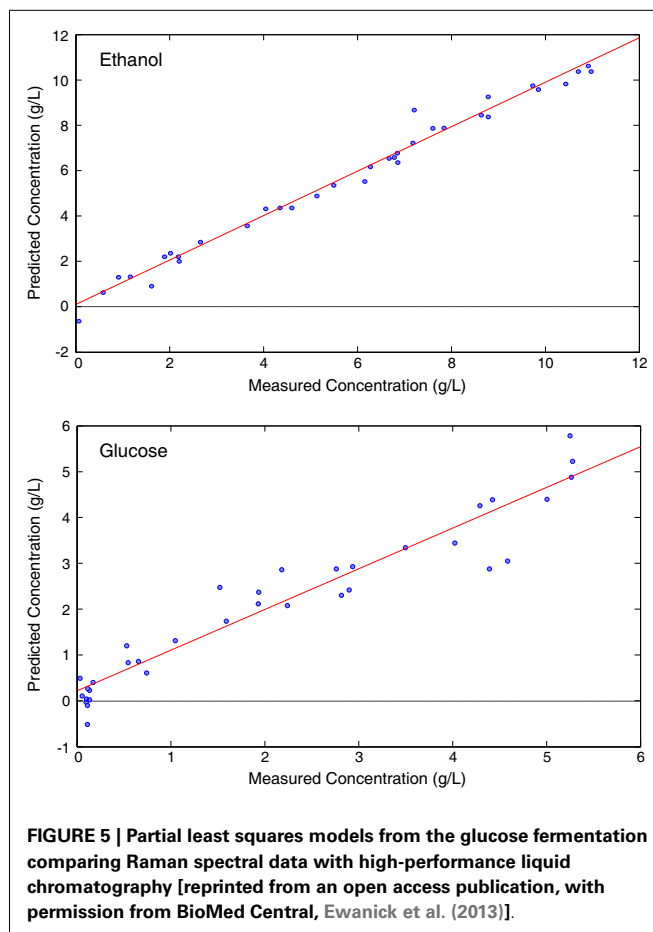


FIGURE 5 | Partial least squares models from the glucose fermentation comparing Raman spectral data with high-performance liquid chromatography [reprinted from an open access publication, with permission from BioMed Central, Ewanick et al. (2013)].

been further illustrated using a similar instrumental configuration for the estimation of ethanol, glucose, and yeast concentrations (Iversen et al., 2014). Increasing concentrations of yeast were found to decrease the intensities of ethanol and glucose peaks caused by Mie scattering from the cells. The attenuation of the Raman signal was corrected using the 1627 cm^{-1} water band as an internal standard to adjust for the scattering from cell particulates. Once the spectra were corrected using the developed quadratic equations for each analyte, a simple linear regression allowed the quantitation of glucose and ethanol with high correlation ($R^2 = 0.99$, ethanol; 0.995 , glucose). This method also enabled the estimation of the yeast concentration.

As previously discussed, visible Raman excitation sources are not commonly employed, due to the intrinsic fluorescence of biomass; however, there have been some recent applications of visible Raman spectroscopy. A frequency doubled 1064 nm Nd:YAG green laser (532 nm for analysis) was used in an interesting study of laser-induced fluorescence (LIF) (Lähdetie et al., 2013). A variety of model compounds representing typical lignin sub-structures were evaluated, including erol, bioerol, dibenzodioxocin, 4-O-methylated bioerol, two synthesized phenolic compounds, and dehydrodivanillin-5-5'. Erol and dibenzodioxocin were easily measured with a relatively flat baseline. Bioerol and 4-O-methylated bioerol revealed broad fluorescence backgrounds containing relatively no Raman bands. The synthesized molecules

could be measured with only moderate spectral contributions from fluorescence; however, measurement of dehydrodivanillin-5-5' resulted in the suppression of Raman modes by a fluorescence background. The authors conclude that the 5-5' linkage is likely a strong source of LIF. Molecules that did not possess a conjugated link between two phenolic moieties did not exhibit fluorescence in the Raman spectra. Dibenzodioxocin, although it possesses the 5-5' linkage, did not display a strong fluorescent background, which the authors deduce likely stems from the molecule's rigid octagonal ring. Raman spectra from spruce wood and thermomechanical pulp (TMP), using the 532 nm laser, showed fluorescent backgrounds, however, the characteristic cellulose and lignin peaks were clearly discernible. When chemically treated pulps [kraft, enzymatic mild acidolysis lignin (EMAL), and milled wood lignin (MWL)] were analyzed, however, LIF was more pronounced. While EMAL and MWL isolation procedures are considered to be mild, retaining the native lignin structure, the analysis of these samples was distinctly different than lignin in wood. The authors hypothesized that the lack of a strong enough LIF background to prevent analysis of spruce likely arises from lignin being bound to the polymer matrix, preventing a release of fluorescence emission. Since EMAL and MWL are no longer connected to the polymeric network, a more malleable conformation results, which could trigger the increased fluorescence background. Other analyses using visible laser sources for Raman spectroscopy of biomass include the analysis of carbonaceous plant materials like bio-char (Ochoa et al., 2014; Tsaneva et al., 2014), and how changes in the cellulose crystallinity of delignified hybrid poplar samples affected the enzymatic hydrolysis yields (Laureano-Perez et al., 2006).

FOURIER-TRANSFORM RAMAN SPECTROSCOPY

Fourier-transform Raman spectroscopy has been the most commonly used instrumental configuration for the analysis of biomass (Agarwal and Atalla, 1993; Sene et al., 1994; Agarwal and Ralph, 1997; Ona et al., 1997; Takayama et al., 1997; Kacurikova et al., 1998; Ona et al., 1998a; Ona et al., 1998b,c; Ona et al., 2000; Schenzel and Fischer, 2001; Sivakesava et al., 2001a; Sivakesava et al., 2001b; Kihara et al., 2002; Proniewicz et al., 2002; Agarwal et al., 2003; Ona et al., 2003; Cao et al., 2004; Vester et al., 2004; Agarwal and Kawai, 2005; Schenzel et al., 2005; Keown et al., 2007; Schulz and Baranska, 2007; Agarwal and Ralph, 2008; Keown et al., 2008; Schenzel et al., 2009; Agarwal and Atalla, 2010; Agarwal et al., 2010; Larsen and Barsberg, 2010; Agarwal, 2011; Agarwal et al., 2011; Chundawat et al., 2011; Larsen and Barsberg, 2011; Sun et al., 2012; Agarwal et al., 2013; Kim et al., 2013; Lupoi et al., 2014a; Wójciak et al., 2014; Lupoi et al., 2015). A recent study surveyed three high-throughput vibrational spectrometers (NIR, FTIR, and FT-Raman) to evaluate which was best suited for developing PLS models for predicting lignin S/G ratios (Lupoi et al., 2014a). Pyrolysis-molecular beam MS (pyMBMS) data from 245 diverse *Acacia* and eucalypt (*Eucalyptus* and *Corymbia*), encompassing 17 different biomass species, was coupled with NIR, FTIR, and Raman spectral data to build one global model. Iterations of different spectral processing techniques were conducted to see which permitted the most robust, accurate PLS model(s). The 245 samples were split into randomly generated 195-sample calibration and 50-sample

validation sets. Additionally, the metrics used for evaluating each model were the result of three, independent, randomized models for each type of spectral transformation. The low error in the calibration and validation statistics indicated that these models were highly robust, as in most cases, it did not matter which samples were in the calibration or validation sets since every combination employed led to similar metrics. The best models (Table 2), based on RMSEP, were constructed using first-derivative, seven-point smoothed, Raman spectra with an extended multiplicative scatter correction (EMSC) (RMSEP = 0.13) and FTIR spectra that had been transformed using EMSC, first, and then the second derivative with 15-point smoothing (RMSEP = 0.13). In a follow-up study, the best Raman model was used to predict the lignin S/G ratio of 269 unknown *Acacia* and eucalypt samples (Lupoi et al., 2015). The calibration and validation sets used to generate the model were recombined to provide a larger data set, enabling more accurate predictions. The Raman predicted S/G ratios displayed no statistical differences from the pyMBMS measured results for all but one of the biomass species (Table 3). Additionally, the plant samples were ranked to illustrate which had the lowest and highest S/G ratios.

Lignin S/G ratios of Eucalyptus, sorghum, switchgrass, maize, and *Arabidopsis* were evaluated using the deconvolution of FT-Raman spectra into peaks identified as representative of S or G lignin monomers (Sun et al., 2012). The specific vibrational modes unique to the different biomass constituents were determined through the measurement of cellulose, xylan, and various model compounds, such as coniferaldehyde, sinapic acid, creosol, 5-methylpyrogallol trimethyl ether, sinapinaldehyde, and sinapyl alcohol. Spectrally resolved peaks corresponding to S or G lignin derivatives were then applied to the biomass samples. The ratios of the resolved S and G peaks were determined and compared to pyGCMS results. The ratios calculated using Raman spectroscopy were consistently higher than those measured using pyGCMS, which could be due to the presence of polysaccharide vibrational modes overlapping with spectral regions designated for each monomer. The deconvolution process itself also contributed to some false peaks such as an artificial S band for pine, a plant known to contain no real S components. Nonetheless, a calibration curve generated using the pyGCMS and Raman calculated ratios resulted in a reasonable correlation ($R^2 = 0.983$). *Arabidopsis* mutants were used to validate the regression model, resulting in a better correlation with the pyGCMS S/G ratios.

When analyzing lignocellulosic materials with Raman spectroscopy, a phenomenon termed "self-absorption" must be considered (Agarwal and Kawai, 2005). Self-absorption occurs when scattered photons are re-absorbed back into the analyte, resulting in an attenuation of the scattered light reaching the detector. This can be visually identified in a Raman spectrum by the decrease in intensity of a vibrational mode where the molecule absorbs light. An analysis of cellulose filter paper, spruce TMP, and MWL illustrated that most of the spectral suppression occurred at the 2895 cm^{-1} C-H peak of the filter paper and TMP (Agarwal and Kawai, 2005). Evaluation of the spectra pointed to cellulose and water as the main contributors of self-absorption, while lignin's involvement was unmeasured. FT-Raman spectroscopy enabled the analysis of the structure of MWLs produced from hard- and softwoods

Table 2 | Comparison of partial least squares models using vibrational spectroscopy and pyrolysis-molecular beam mass spectrometry [reprinted from Lupoi et al. (2014a)].

Method	SEL validation ^{a,g}	SEP ^{b,g}	RMSEP ^{c,g}	r-Val ^d	R ² Val ^e	Outliers ^f
Raman 2nd deriv. (19pt) + SNV 32 scans	0.05	0.14	0.13	0.89 ± 0.04	0.79 ± 0.08	1
Raman 1st deriv. (7pt) + EMSC 32 scans	0.05	0.13	0.13	0.91 ± 0.02	0.83 ± 0.04	1
Raman EMSC + 2nd deriv. (15pt) 96 scans	0.05	0.14	0.15	0.90 ± 0.02	0.81 ± 0.04	0
Raman 2nd deriv. (15pt) + SNV 96 scans	0.06	0.17	0.16	0.86 ± 0.02	0.74 ± 0.04	0
MIR EMSC + 2nd deriv. (15pt)	0.05	0.14	0.13	0.87 ± 0.06	0.8 ± 0.1	1
MIR 2nd deriv. (17pt) + MSC	0.05	0.14	0.14	0.91 ± 0.01	0.83 ± 0.01	1
MIR 2nd deriv. (17pt) + SNV	0.05	0.15	0.15	0.87 ± 0.02	0.76 ± 0.03	1
NIR EMSC + 2nd deriv. (25pt)	0.06	0.19	0.20	0.79 ± 0.01	0.62 ± 0.01	0
NIR 2nd deriv. (25pt) + MSC	0.06	0.18	0.18	0.82 ± 0.04	0.67 ± 0.07	1
NIR 2nd deriv. (25pt) + SNV	0.06	0.22	0.21	0.80 ± 0.04	0.65 ± 0.07	1

^aStandard error of the laboratory for the validation data.^bStandard error of prediction.^cRoot mean standard error prediction.^dCorrelation coefficient for the validation set.^ePearson coefficient of determination for validation.^fNumber of outliers removed from validation models.^gAverage errors of three randomly generated models using data provided. Models were not statistically different.

The numbers listed parenthetically reflect the degree of Savitzky–Golay spectral smoothing.

Statistical values are the average of three independent models.

and chemically treated black spruce (Agarwal et al., 2011). The Raman spectra revealed distinct changes when differentiating the untreated to pretreated samples. Acetylation and methylation produced sizeable changes in aliphatic C–H vibrational modes, and also resulted in the formation of several new peaks.

The viability of FT-Raman spectroscopy for monitoring a bioethanol process has also been explored (Sivakesava et al., 2001b). Glucose, ethanol, and optical cell density were evaluated during ethanol fermentation. Raman spectra were coupled with HPLC results for the construction of PLS and PCR models. Although the predictions of glucose and ethanol were acceptable, the cell density modeling proved to be more erroneous due to the weak scattering generated from the cultures. Another study analyzed glucose, lactic acid, and cell density, at-line, during a lactic acid fermentation process (Sivakesava et al., 2001a). PLS models generated using IR, NIR, and Raman spectral data were contrasted, with the Raman models having the second lowest SEP in glucose prediction. The Raman SEP of lactic acid and cell density predictions ranked third between the three instruments. The authors attribute this lack of accuracy to the fact that glucose, lactic acid, and proteins have weaker Raman signals compared to IR spectroscopy.

RESONANCE RAMAN SPECTROSCOPY

Resonance Raman (RR) spectroscopy is achieved when a molecule has an electronic absorption that overlaps with the excitation source wavelength, resulting in the promotion of the molecule to a real, rather than a virtual, electronic state. In complex analytes such as biomass, molecules resonating with the excitation source will be selectively enhanced. For example, lignin has an electronic

absorption in the UV region, leading to increased lignin spectral intensities when UV lasers are employed. This resonance allows lignin to be preferentially studied while polysaccharides generate limited spectral response. An advantage of evaluating lignin with UVRR, as previously discussed, is that the lignin can be measured *in situ*. This allows a more pragmatic analysis of lignin structure, since the techniques commonly employed to extract or isolate lignin from plants are known to alter the lignin. Another benefit to using UV lasers is that fluorescence, ubiquitous in visible Raman spectroscopy, and still a hindrance at some shorter NIR wavelengths, is not problematic. The analysis of lignin model compounds using UVRR enabled S, G, and H lignin markers to be characterized (Saariaho et al., 2003, 2005). The use of a tunable argon laser allowed three different excitation wavelengths to be evaluated: 229, 244, and 257 nm (Saariaho et al., 2003). The lignin S, G, and H markers were preferentially enhanced based upon which excitation wavelength was used. H lignin structures showed the strongest enhancement when 244 nm was employed, while G moieties were more intense when 257 nm was used. The spectra generated from S functionalities were essentially indistinguishable when using either 244 or 257 nm. A follow-up study utilized PLS to determine the specific wavelengths correspondent to each type of lignin monomer, as well as condensed structures, conjugated C = C and C = O, and stilbenes (Saariaho et al., 2005). The authors note that using multivariate analysis in this fashion can aid in qualitatively interpreting complex spectra of polymeric lignin, since the UVRR spectra typically have broad peaks. The evaluation of the PLS model loadings plots allowed the identification of important vibrational modes corresponding to the different model compound classes.

Table 3 | Comparison of pyrolysis-molecular beam mass spectrometry measured to Raman spectroscopy predicted syringyl/guaiacyl ratios^a.

Plant species	No. of samples	Reference range	pyMBMS S/G average	No. of samples	Prediction range	Raman S/G average	Predicted vs. reference comparisons (<i>p</i> -values)
<i>Acacia microbotrya</i>	5	1.2–1.5	1.3 ± 0.1	10	0.9–1.5	1.3 ± 0.2	0.83
<i>A. saligna</i>	4	1.4–1.9	1.7 ± 0.2	11	1.2–2.0	1.7 ± 0.2	0.69
<i>Corymbia citriodora</i> subsp. <i>citriodora</i>	17	2.1–2.8	2.4 ± 0.2	44	2.0–2.7	2.3 ± 0.1	0.61
<i>Corymbia</i> hybrids	47	1.6–2.8	2.2 ± 0.2	–	–	–	NA
<i>C. torelliana</i>	56	1.8–2.4	2.1 ± 0.1	–	–	–	NA
<i>C. citriodora</i> subsp. <i>variegata</i>	39	2.0–3.2	2.5 ± 0.3	61	2.2–2.7	2.5 ± 0.1	0.65
<i>Eucalyptus argophloia</i>	5	1.9–2.2	2.1 ± 0.1	5	1.7–2.0	1.8 ± 0.1	0.03
<i>E. cladocalyx</i>	3	2.3–2.6	2.5 ± 0.2	2	2.1, 2.4	2.2 ± 0.2	0.40
<i>E. cloeziana</i>	7	1.7–2.3	1.9 ± 0.2	15	1.7–2.4	2.1 ± 0.2	0.31
<i>E. crebra</i>	4	1.4–2.1	1.6 ± 0.4	6	1.2–2.1	1.8 ± 0.3	0.59
<i>E. dunnii</i>	4	2.2–2.8	2.5 ± 0.3	11	2.2–2.5	2.4 ± 0.1	0.36
<i>E. globulus</i>	11	2.3–3.0	2.6 ± 0.2	19	2.0–2.8	2.5 ± 0.2	0.15
<i>E. grandis</i>	2	1.9, 2.2	2.0 ± 0.2	13	1.9–2.4	2.2 ± 0.1	0.23
<i>E. kochii</i>	5	1.9–2.3	2.2 ± 0.2	10	1.7–2.5	2.2 ± 0.2	0.76
<i>E. longirostrata</i>	8	2.1–2.4	2.2 ± 0.1	7	2.0–2.3	2.2 ± 0.1	0.86
<i>E. loxophleba</i>	7	2.1–2.6	2.4 ± 0.1	23	2.2–2.7	2.4 ± 0.1	0.22
<i>E. moluccana</i>	5	2.0–2.5	2.2 ± 0.2	11	1.8–2.5	2.2 ± 0.2	0.91
<i>E. occidentalis</i>	6	2.1–2.5	2.4 ± 0.2	9	2.2–2.6	2.4 ± 0.1	0.55
<i>E. polybractea</i>	8	2.0–2.7	2.3 ± 0.2	12	1.9–2.5	2.2 ± 0.1	0.40

^aData compiled from Lupoi et al. (2014a), “High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using multivariate analysis: a comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development,” *Biotechnology for Biofuels*, Volume 7, p. 93 and Lupoi et al. (2015), “High-throughput prediction of *Acacia* and *eucalypt* lignin syringyl/guaiacyl content using FT-Raman spectroscopy and partial least squares modeling” *Bioenergy Research*, open access.

The functional groups of lignin contribute to its chemical properties and its valorization potential. Phenolic moieties, one of the principal functionalities in lignin, define the reactivity and solubilization of lignin (Zakis, 1994). The ionization of phenolic species in alkaline media results in a concomitant shift in the vibrational modes of lignin in pulps and lignin model compounds (Warsta et al., 2012). Shifts from 8 to 35 cm^{−1} were measured when the pH was increased from 6 to 12. In general, as the pH became more alkaline, a shift to lower wavenumbers was detected. When wood pulps were analyzed, a less pronounced shift resulted, since the pulps have less phenolic functionalities than model compounds. When non-phenolic 3, 4-dimethoxytoluene was measured, no shift was detected, indicative that the shifting occurred due to ionization of the phenolic group. Increases in pH also resulted in augmented band intensities; however, the band intensity was still directly proportional to analyte concentration, as exemplified by the construction of a calibration curve for guaiacol. While these bands were detected at more neutral pH levels, the enhancement of these bands at strongly basic pH provided a more detailed structural analysis. The authors suggest that the shifting of the aromatic band near 1600 cm^{−1} from increasing the alkalinity of the matrix may aid in determining the amount of free phenolic groups (for example, an 11 cm^{−1} shift can be expected if *all* of the phenylpropanoid functionalities have a free phenolic group).

Ultraviolet resonance Raman enabled the analysis of extractable lipophilic and hydrophilic components from Scots pine wood resin (Nuopponen et al., 2004b,c). The authors employed a tunable argon laser set to one of three different excitation wavelengths: 229, 244, or 257 nm. The level of the enhancement for different structures depended on the particular laser wavelength employed. Molecules such as resin (dehydroabietic, abietic, and pimaric type) and fatty acids, sitosterol, and sitosterol acetate were evaluated as standards, in hexane extracts from the biomass, and in solid wood samples. Double-bond moieties, such as those found in alkenes, were resonantly enhanced using the UV laser wavelengths. When the 257 nm wavelength was used, compounds with isolated double bonds provided the most information, while the 229 nm wavelength was more useful for analyzing conjugated resin acids. Additionally, the 257 nm laser was best suited for studying sapwood hexane extracts, while either the 229 or the 244 nm lasers could be employed for evaluating heartwood extracts. The measurement of solid wood revealed a vibrational mode at 1650 cm^{−1}, indicative of unsaturated wood resin constituents. For the hydrophilic extractables, only the 244 and 257 nm wavelengths were used. Aromatic and unsaturated moieties of pinosylvin and chrysin were found to be resonantly enhanced. Wavelength selection had a minimal effect on chrysin analysis. The heartwood acetone/water extract included pinosylvin plus resin and fatty acid markers. The sapwood extract

contained oleophilic structures of the resin and fatty acids, as well as some guaiacyl modes. The measurement of Scots pine knot-wood unveiled an abundant resin contribution, illustrating that the resin was more resonantly enhanced than lignin. These two analyses have demonstrated the competence of UVRR to selectively analyze extractable compounds. Although extractives are a smaller proportion of biomass compared to polysaccharides and lignin, they have significant impacts on plant properties and may also present a source of bio-based chemicals.

A UVRR method was established for quantifying lignin in bleached hardwood kraft pulps (Jaaskelainen et al., 2005). Lignin quantification techniques typically are developed using unbleached biomass, and therefore are not readily transferable to bleached samples. A strong linear correlation ($R^2 = 0.987$) was calculated when the 1604 cm^{-1} peak was normalized to the 1093 cm^{-1} cellulose peak, and plotted against increasing lignin concentration. A 244 nm excitation wavelength provided more accurate lignin content measurements, since the use of 257 nm resulted in more fluctuations in the spectral baseline. The Raman measured lignin contents were compared with kappa numbers measured using a standard method, and were found to linearly correlate. Other applications of UVRR include the degradation of lignin following a chemical treatment such as bleaching (Halttunen et al., 2001; Mononen et al., 2005; Jaaskelainen et al., 2006; Lähdele et al., 2009) or steam treatment (Nuopponen et al., 2004a), the changes in TMP after laccase treatments (Lähdele et al., 2009), photodegradation using an UV laser (Pandey and Vuorinen, 2008), and an evaluation of 25 diverse tropical hardwoods using UVRR spectral data and principal component analysis (PCA) (Nuopponen et al., 2006).

Resonance Raman spectroscopy using visible excitation sources with Kerr-gated fluorescence rejection has enabled structural analyses of lignin that were previously unattainable (Barsberg et al., 2005, 2006). A Kerr-gate is a device consisting of two polarizers and a Kerr medium (carbon disulfide, in this instrument). When closed, the polarizers blocked scattered photons from reaching the detector. The Kerr-gate provided a time-window of 4 ps to collect Raman spectra free from fluorescence, a phenomenon occurring on a nanosecond timeframe. Once the Raman data had been acquired, the Kerr-gate was switched to the closed position, thereby blocking fluorescence. Syringyl moieties were resonantly enhanced when a 400 nm laser was used, whereas the use of 500 nm light caused a reduction in selectivity. The effects of laccase plus various mediators on beech lignin were studied using both excitation wavelengths and RR difference spectra. In a follow-up study, the authors successfully measured lignin radicals produced enzymatically using laccase (Barsberg et al., 2006). A 1570 cm^{-1} band was measured in dry wood, regardless of the type of biomass was analyzed. When wet beech was evaluated, a lignin radical peak at 1606 cm^{-1} was detected. Density functional theory was used to correlate the experimental results with the predicted vibrational modes of lignin radicals, and indicated that the radicals were formed from syringyl and guaiacyl moieties in beech and spruce, respectively. RR difference spectra were calculated to subtract spectral contributions from the main lignin peak near 1600 cm^{-1} . The radicals could only be detected when the 500 nm light was used to generate Raman scatter.

Resonance Raman spectroscopy coupled with Kerr-gated fluorescence suppression allowed the measurement of strongly fluorescent chemical pulps using 400 nm light (Saariaho et al., 2004). Although these pulps are not typically assessable, due to lignin fluorescence, the use of the Kerr-gate permitted a 250-fold reduction in the fluorescence background, enabling much weaker Raman bands to be detected. Chromophoric vibrational modes at 1605 and 1655 cm^{-1} were measured in peroxide-bleached pulps, while only the 1605 cm^{-1} was identified in biomass treated with chlorine dioxide. When a 257 nm laser was used to evaluate the pulps, the intensity of the aromatic lignin peak was approximately 20-times higher than the main cellulose mode. The square root of the ratio of the 1605 cm^{-1} vibrational mode to the 1098 cm^{-1} peak correlated linearly with brightness percentage, as measured using a standard method. The authors concluded that while UV excitation preferentially probed lignin, visible lasers allowed the detection of chromophoric lignin structures. Lignin remaining in chemically treated pulps could be quantified using RR spectroscopy, although the detection limit can be lowered when UV lasers are employed.

RAMAN IMAGING

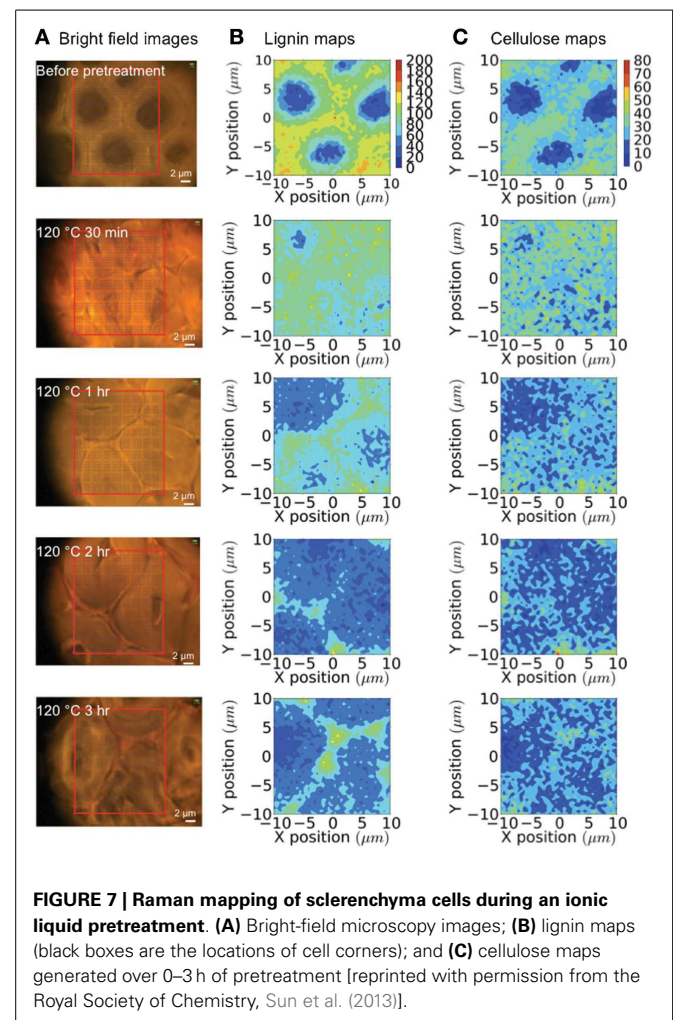
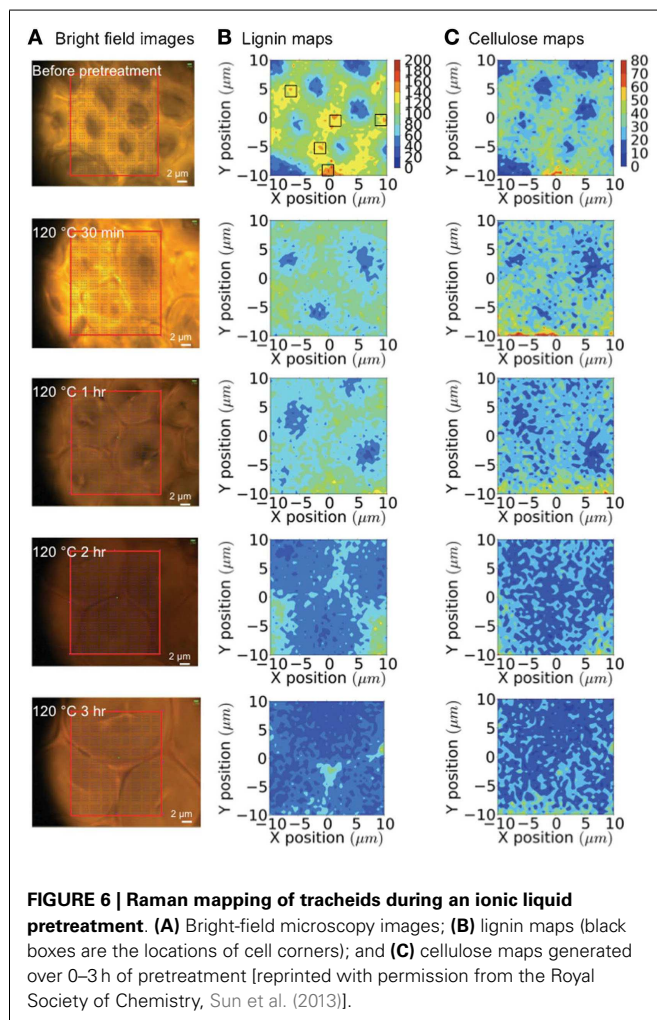
Raman imaging techniques have enabled the visual examination of biomass cell and cell wall structure, and the evaluation of real-time changes in the morphology and chemical content of plants, such as after different pretreatment strategies. These experiments have provided a glimpse into the chemistry of plants before and after treatments, permitting researchers to identify the biomass modification approaches best suited for reducing recalcitrance and increasing yields from downstream conversion into simple sugars. The laser can be focused to small spot sizes, enabling minute areas of interest to be evaluated. Instrumental advances have allowed the rapid acquisition of images with short integration times, preventing the photodegradation of the sample. Another advantage of Raman imaging, juxtaposed to other imaging techniques, is that no staining or embedding of the sample is required. Raman spectra are collected from the sample, as the instrument passages the sample to defined locations using a set step-size, resulting in a plethora of structural and chemical data that can be daunting to analyze. Multivariate analysis, coupled with imaging techniques, has enabled enhanced data mining for valuable information.

Raman microspectroscopy has been used to evaluate how a room temperature pretreatment with the ionic liquid (IL) 1-*n*-ethyl-3-methylimidazolium acetate modified the cell walls of poplar (Lucas et al., 2011). A 785 nm diode laser was used to collect spectral data from $50\text{ }\mu\text{m}$ poplar sections. Raman spectra from untreated poplar revealed the characteristic vibrational modes from cellulose, hemicellulose, and lignin. When the wood was swollen with water, the same peaks were identified; however, the intensities differed from the untreated samples. The intensity ratio of the 1460 cm^{-1} cellulose peak to the 1605 cm^{-1} lignin peak decreased, which signified diminutions in the cellulose-abundant S2 sub-layer compared to the hydrophobic, lignin-rich compound middle lamella (CML) region. The authors conclude that the swelling must be pushing the fibers apart, and progressing into more amorphous cellulose regions since crystalline cellulose fibers are recalcitrant to water penetration. The Raman spectra of the IL treated poplar samples depicted strong signals from the IL itself.

When the samples were washed with water prior to analysis, the spectra showed no traces of IL vibrational modes, and resembled the water-swollen poplar Raman spectrum, leading to the conclusion that both the water and IL treatments led to similar overall cell wall compositions. Confocal Raman spectroscopy using a 785 nm diode laser enabled an evaluation of tissue-specific changes when pretreating corn stover with the IL 1-ethyl-3-methylimidazolium acetate (Sun et al., 2013). A temporal study was conducted to gauge the lignin and cellulose remaining in the plant cell walls during the IL pretreatment at 120°C using 0, 30 min, 1, 2, and 3-h time points. To assess the changes brought about by the IL treatment, tracheids, sclerenchyma, and parenchyma cell structures were probed (see **Figures 6** and **7**). Before pretreating the corn stover, cellulose and lignin concentrations were highest in the cell corners (CCs) and CML portions of the three cell structures and in the secondary walls of the sclerenchyma and parenchyma cell types. The lignin content was measured to decrease rapidly during the IL treatment, while no preferential cellulose dissolution was detected. The IL pretreatment is known to cause swelling of the secondary wall, but not of the CML. Accordingly, more significant swelling was observed in tracheid and sclerenchyma cells than parenchyma cells, which are composed of primary cell

walls. Although tracheids contained higher lignin concentrations and thicker walls than parenchyma cells, the lignin dissolution occurred more rapidly in the tracheid cells. Confocal Raman microscopy was also employed to evaluate normal and tension wood sections from poplar (Gierlinger and Schwanninger, 2006). The allocation of cell wall components was calculated following the integration of distinct vibrational modes. The Raman images of normal wood illustrated higher lignin concentrations in the CCs and the CML, and increased cellulose content in the S2 layer of parenchyma ray cells and two lesser layers located on each side of the CML, presumed to be S1. A higher fluorescence background was measured for CCs and the CML, which is expected due to the greater lignin concentrations in these regions. Analysis of tension wood samples revealed lignin to be localized in CCs and the CML, while no lignin was detected in the gelatinous, or G-layer. Signals from lignin increased, however, in the lumen. Aromatic compounds were measured to coalesce along an inner region of the G-layer, and were also found deeper in the G-layer, toward the CCs of the S2 layer.

Many applications of Raman imaging utilize NIR excitation sources. Visible excitation, however, has been demonstrated as offering a higher energy source for obtaining Raman images. A



novel, polarized 633 nm laser was used to attain images of black spruce cross-sections (Agarwal, 2006). Fluorescence was efficiently blocked by acquiring data in confocal mode using a 100 μm pinhole. Lignin concentrations were highest in the CCs, concurrent with other studies, but were not profoundly different in the CML and secondary wall. Coniferaldehyde and conifer alcohol distribution, using the 1650 cm^{-1} , was measured to correspond with lignin. Cellulose localization followed an opposite pattern to lignin distribution (high S2 and low CC and CML concentrations). A confocal Raman microscope, equipped with a 532 nm laser and a 100 μm pinhole, was used to characterize black cottonwood (Perera et al., 2011). Given the heterogeneity of the sample, the abundance of spectral information, and spectrally unresolved vibrational modes, the authors developed a new analysis strategy to aid in determining the structural characteristics and chemical composition of the wood. The method encompasses three main phases: spectral preprocessing, stepwise clustering, and estimation of spectral profiles of pure components and their respective weights. The spectral processing included wavelet analysis to remove noise, second-derivative transformations to remove contributions from fluorescence, and PCA to reduce the amount of variables as well as reduce noise from the data matrix. Stepwise clustering was achieved using *k*-means clustering to classify the samples according to a preordained number of groups. The image can then be reconstructed using the cluster groupings, facilitating the identification of diverse sub-layers within the cell wall. The last step involves determining which factors are important in contributing to the distinct localization of different cell wall components in the images. A technique called spectral entropy minimization methodology allowed the pure components spectra to be captured. Estimated pure polysaccharide and lignin spectra were generated. Pure cellulose and hemicellulose components could not be generated due to the structural, and therefore, spectral resemblance between the two polysaccharides. The lignin spectrum included regions typically assigned to lignin monomers, permitting an *in situ* analysis of monolignol composition. The authors note that this is not possible with routine data processing techniques, since the lignin monomers have significant spectral overlap with carbohydrate vibrational modes. The partitioning of lignin and carbohydrates in the images was determined by subtracting first the pure lignin and then the pure carbohydrate spectra from the average spectra determined for each cluster. The image analysis procedure can be extended to other types of spectral data such as IR, MS, or fluorescence. In a follow-up study, this method was employed to evaluate the S and G lignin contents of *Arabidopsis*, *Miscanthus*, and poplar. Spectral distinctions between the three plants were clearly discernible in the estimated lignin spectra, indicative of differences in S, G, and H contents. The *Miscanthus* spectrum was less intricate than the dicots, which, the authors deduce is illustrative of *Miscanthus* having a higher percentage of non-condensed lignin. Lignin S/G ratios were calculated to be 0.5 ± 0.08 , 0.6 ± 0.1 , and 1.9 ± 0.2 for *Miscanthus*, *Arabidopsis*, and poplar, respectively. The S/G ratios within different cell wall structures could also be calculated (0.8 ± 0.1 for *Miscanthus* xylary fiber cells, 0.6 ± 0.1 for *Miscanthus* interfascicular cells of basal stems). A transgenic poplar sample, in which the monolignol biosynthesis gene encoding for 4-coumarate-CoA

ligase was suppressed, revealed reduced total lignin contents and decreased S/G ratios. These examples demonstrate the power of Raman spectroscopy coupled with chemometric techniques to exhaustively extract obscured information from the spectra.

Coherent anti-Stokes Raman scattering microscopy can be used to obtain images of biomass devoid of fluorescence (Zeng et al., 2012; Pohling et al., 2014). In CARS, multiple lasers interact with the analytes, termed pump, probe, or Stokes excitation sources. These lasers are used to generate the anti-Stokes photons. When the frequency difference between the pump and Stokes lasers is tuned to coincide with a specific Raman vibrational mode, the signal is enhanced. CARS intensities are stronger than those obtained using spontaneous Raman spectroscopy, leading to increased sensitivity, and shorter acquisition times. Spectral assignments in CARS spectra are identical to those assignments made using traditional Raman spectra. Wood samples of birch, oak, and spruce were evaluated using CARS microscopy (Pohling et al., 2014). Standards of pure cellulose, xylan, and lignin were measured to establish indicative marker bands. Through the use of spectrally broad lasers, the CARS protocol can probe multiple vibrations (MCARS). Using a technique called the maximum entropy method, Raman spectra could be extracted from the MCARS spectral data, revealing spectra that resembled standard spectra collected with spontaneous Raman spectroscopy, minus fluorescence contributions. Transverse and longitudinally oriented cuts of the wood samples illustrated the cell wall structure and composition. Cellulose, hemicellulose, and lignin were localized allowing the assignment of cellulose-rich secondary walls, and lignin-rich intercellular space. The longitudinally cut images showed polarization dependence. Stronger cellulose signals were detected using horizontal polarization, while more intense lignin peaks were measured using vertical polarization. A semi-quantitative assessment of cellulose, hemicellulose, lignin, and water illustrated the significance of polarization in the longitudinal plant sections, as there was greater disparity in the results when measured with horizontal or vertical polarization.

Coherent anti-Stokes Raman scattering, however, experiences an electronic background that can alter spectral data, obscuring the quantitation of analytes from CARS imaging techniques (Li et al., 2005). SRS microscopy provides orders-of-magnitude higher spectral signals, eliminating the effects of the higher background (Freudiger et al., 2008). The SRS phenomenon is similar to that observed in CARS. Two lasers are overlapped and focused onto the analyte. When the difference frequency of the two lasers resonates with a vibrational mode in the sample, the rate at which photons migrate to higher vibrational levels is enhanced due to stimulated photon excitation. Energy transfers only occur when in resonance with a molecule's fundamental vibrational mode(s). Although the signals are weak and obscured in the background produced from the laser, the laser noise can be eliminated by using a high-frequency ($> 1\text{ MHz}$) amplitude modulation/lock-in detection procedure (Saar et al., 2010). Like CARS, the assignments of vibrational modes are equivalent to those generated from spontaneous Raman scattering. Also analogous to spontaneous Raman spectroscopy, signals produced using SRS are linearly dependent on analyte concentration. SRS microscopy was employed to evaluate the real-time processing of corn stover (Saar et al., 2010). The images were acquired in approximately 3 s, whereas the same

image would have required nearly 2 h using spontaneous Raman scattering. Cellulose and lignin localization were generated in the images, without using labels or staining, by tuning the frequency difference of the two lasers to the well-known vibrational modes of each biopolymer. The validity of this technique was confirmed by comparing with common staining techniques, such as phloroglucinol for lignin detection. The vessel, tracheid, and fiber cells revealed significant lignification compared to phloem cells. The cellulose content was more uniformly dispersed throughout the cells, juxtaposed to lignin. Areas of higher and lower cellulose and lignin concentration could be detected, an observation that is more challenging in CARS, due to the inability to separate the signal from the higher background. The authors used this method to monitor the delignification of corn stover using a sodium chlorite treatment. An eightfold decrease in lignin content was measured while the cellulose content remained relatively unchanged. Analysis of the SRS images provided information on where lignin was preferentially removed from the corn stover during the bleaching process. The bleaching rates were fastest for the lignin contained in the phloem and CCs. Parenchyma, tracheid, vessel, and fiber cells demonstrated similar delignification patterns, signifying similar accessibilities of lignin to sodium chlorite.

Other Raman imaging applications include an analysis of the structural changes in polyaromatic molecules and non-aromatic moieties following the carbonization of Japanese cedar, cotton cellulose, and lignin at 500–1000°C (Ishimaru et al., 2007), studies on deformation properties of native and regenerated celluloses (Hamad, 2008), the monitoring of structural and chemical changes in *Miscanthus x giganteus* following a sodium hydroxide treatment (Chu et al., 2010), the localization of cellulose and lignin in corn stover and *Eucalyptus globulus* (Sun et al., 2010), the *in situ* detection of a single carotenoid crystal (Baranska et al., 2011), and the characterization of cellulose nanocrystal (CNC)-polypropylene composites, and determine the spatial distribution of the CNC in the filaments (Agarwal et al., 2012).

CONCLUSION

As the search for ideal wild-type or transgenic biofuel and biomaterial feedstocks progresses, methods that rapidly and accurately screen large arrays of different plants are becoming essential. Raman spectroscopy, in its diverse configurations, has proven to be a viable asset to these qualitative and quantitative studies. As instrumental innovations evolve, such as field-portable devices, measurements of the feedstocks can be conducted in their natural environments, reducing the need for time-consuming sampling protocols. The construction of robust, multivariate predictive models coupled to Raman spectral data will increase experimental throughput, thereby narrowing the pool of potential plants suitable for downstream renewable energy applications. Raman imaging techniques have empowered researchers to evaluate deconstruction strategies in real-time, providing both fundamental insights into how specific reagents affect the morphology of the biomass, and also the ability to nominate or exclude the pretreatment method based on the efficiency of rendering the cell wall-less recalcitrant based on end-product yields. The extent of endeavors explored for the characterization of

lignocellulosic biomass using Raman spectroscopy continues to escalate. Future advancements in instrumentation, multivariate analysis modeling, and the revolutionary ways in which Raman spectroscopy is utilized will continue to proffer researchers a versatile, non-destructive, non-invasive, user-friendly, high-throughput analytical tool.

REFERENCES

- Agarwal, U. P. (2011). "Lignin quantitation by FT-Raman spectroscopy," in *Proceedings 16th International Symposium on Wood, Fiber and Pulp Chemistry*, 170–173.
- Agarwal, U. P. (1999). "An overview of Raman spectroscopy as applied to lignocellulosic materials," in *Advances in Lignocellulosics Characterization*, Chap. 9. ed. D. S. Argyropoulos (Atlanta, GA: TAPPI Press), 201–225.
- Agarwal, U. P. (2006). Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (*Picea mariana*). *Planta* 224, 1141–1153. doi:10.1007/s00425-006-0295-z
- Agarwal, U. P., and Atalla, R. H. (1993). Raman spectroscopic evidence for coniferyl alcohol structures in bleached and sulfonated mechanical pulps. *ACS Symp. Ser.* 531, 26–44. doi:10.1021/bk-1992-0531.ch002
- Agarwal, U. P., and Atalla, R. H. (2010). "Vibrational spectroscopy," in *Lignin and Lignans: Advances in Chemistry*, eds C. Heitner, D. R. Dimmel, and J. A. Schmidt (Boca Raton, FL: CRC Press), 103–136.
- Agarwal, U. P., and Kawai, N. (2005). "Self-absorption" phenomenon in near-infrared Fourier transform Raman spectroscopy of cellulosic and lignocellulosic materials. *Appl. Spectrosc.* 59, 385–388. doi:10.1366/0003702053585327
- Agarwal, U. P., Mcsweeney, J. D., and Ralph, S. A. (2011). FT-Raman investigation of milled-wood lignins: softwood, hardwood, and chemically modified black spruce lignins. *J. Wood Chem. Technol.* 31, 324–344. doi:10.1080/02773813.2011.562338
- Agarwal, U. P., and Ralph, S. A. (1997). FT-Raman spectroscopy of wood: identifying contributions of lignin and carbohydrate polymers in the spectrum of black spruce (*Picea mariana*). *Appl. Spectrosc.* 51, 1648–1655. doi:10.1366/0003702971939316
- Agarwal, U. P., and Ralph, S. A. (2008). Determination of ethylenic residues in wood and TMP of spruce by FT-Raman spectroscopy. *Holzforschung* 62, 667–675. doi:10.1515/HE.2008.112
- Agarwal, U. P., Reiner, R. R., and Ralph, S. A. (2013). Estimation of cellulose crystallinity of lignocelluloses using near-IR FT-Raman spectroscopy and comparison of the Raman and Segal-WAXS methods. *J. Agric. Food Chem.* 61, 103–113. doi:10.1021/jf304465k
- Agarwal, U. P., Reiner, R. S., and Ralph, S. A. (2010). Cellulose I crystallinity determination using FT-Raman spectroscopy: univariate and multivariate methods. *Cellulose (Dordrecht, Neth.)* 17, 721–733. doi:10.1021/jf304465k
- Agarwal, U. P., Sabo, R., Reiner, R. S., Clemons, C. M., and Rudie, A. W. (2012). Spatially resolved characterization of cellulose nanocrystal-polypropylene composite by confocal Raman microscopy. *Appl. Spectrosc.* 66, 750–756. doi:10.1366/11-06563
- Agarwal, U. P., Weinstock, I. A., and Atalla, R. H. (2003). FT-Raman spectroscopy for direct measurement of lignin concentrations in kraft pulps. *Tappi J.* 2, 22–26.
- Azimvand, J. (2014). Studying of optical breakdown of DIOXANE lignin as a result of structural changes in the functional groups carbonyl, ether C-O and PHE-NOLIC hydroxyl with the use of Raman and FT-IR spectroscopy. *Adv. Nat. Appl. Sci.* 8, 94–101.
- Baranska, M., Baranski, R., Grzebelus, E., and Roman, M. (2011). In situ detection of a single carotenoid crystal in a plant cell using Raman microspectroscopy. *Vib. Spectrosc.* 56, 166–169. doi:10.1016/j.vibspec.2011.02.003
- Barbillat, J., and Da Silva, E. (1997). Near infra-red Raman spectroscopy with dispersive instruments and multichannel detection. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 53, 2411–2422. doi:10.1016/S1386-1425(97)00181-9
- Barsberg, S., Matousek, P., and Towrie, M. (2005). Structural analysis of lignin by resonance Raman spectroscopy. *Macromol. Biosci.* 5, 743–752. doi:10.1002/mabi.200500042
- Barsberg, S., Matousek, P., Towrie, M., Joergensen, H., and Felby, C. (2006). Lignin radicals in the plant cell wall probed by Kerr-gated resonance Raman spectroscopy. *Biophys. J.* 90, 2978–2986. doi:10.1529/biophysj.105.070391
- Blanch, H. W., Simmons, B. A., and Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnol. J.* 6, 1086–1102. doi:10.1002/biot.201000180

- Cao, Y., Lu, Y., and Huang, Y. (2004). NIR FT-Raman study of biomass (*Triticum aestivum*) treated with cellulase. *J. Mol. Struct.* 693, 87–93. doi:10.1016/j.molstruc.2004.02.017
- Carey, P. R. (1982). "Principles of Raman spectroscopy," in *Biochemical Applications of Raman and Resonance Raman Spectroscopies* (New York: Academic Press), 11–47.
- Chase, B., and Talmi, Y. (1991). The use of a near-infrared array detector for Raman spectroscopy beyond one micron. *Appl. Spectrosc.* 45, 929–931. doi:10.1366/0003702914336156
- Chu, L.-Q., Masyuko, R., Sweedler, J. V., and Bohn, P. W. (2010). Base-induced delignification of *Miscanthus x giganteus* studied by three-dimensional confocal Raman imaging. *Bioresour. Technol.* 101, 4919–4925. doi:10.1016/j.biortech.2009.10.096
- Chundawat, S. P. S., Donohoe, B. S., Sousa, L. D. C., Elder, T., Agarwal, U. P., Lu, F., et al. (2011). Multi-scale visualization and characterization of lignocellulosic plant cell wall deconstruction during thermochemical pretreatment. *Energy Environ. Sci.* 4, 973–984. doi:10.1039/c0ee00574f
- Dao, N. Q. (2006). "Dispersive Raman spectroscopy, current instrumental designs," in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers (Hoboken, NJ: John Wiley & Sons), 13024–13058.
- Ewanick, S. M., Thompson, W. J., Marquardt, B. J., and Bura, R. (2013). Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. *Biotechnol. Biofuels* 6, 28–28. doi:10.1186/1754-6834-6-28
- Freudiger, C. W., Min, W., Saar, B. G., Lu, S., Holtom, G. R., He, C., et al. (2008). Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* 322, 1857–1861. doi:10.1126/science.1165758
- Furtado, A., Lupoi, J. S., Hoang, N. V., Healey, A., Singh, S., Simmons, B. A., et al. (2014). Modifying plants for biofuel and biomaterial production. *Plant Biotechnol. J.* 12, 1246–1258. doi:10.1111/pbi.12300
- Gierlinger, N., and Schwanninger, M. (2006). Chemical imaging of poplar wood cell walls by confocal Raman microscopy. *Plant Physiol.* 140, 1246–1254. doi:10.1104/pp.105.066993
- Gray, S. R., Peretti, S. W., and Lamb, H. H. (2013). Real-time monitoring of high-gravity corn mash fermentation using in situ Raman spectroscopy. *Biotechnol. Bioeng.* 110, 1654–1662. doi:10.1002/bit.24849
- Halttunen, M., Vyorykka, J., Hortling, B., Tamminen, T., Batchelder, D., Zimmermann, A., et al. (2001). Study of residual lignin in pulp by UV resonance Raman spectroscopy. *Holzforchung* 55, 631–638. doi:10.1515/HF.2001.103
- Hamad, W. Y. (2008). "Studies of deformation processes in cellulose using Raman microscopy," in *Characterization of Lignocellulosic Materials*, Chap. 7, ed. T. Q. Hu (Oxford: Blackwell Publishing), 121–137. doi:10.1002/9781444305425.ch7
- Ingle, J. D. Jr., and Crouch, S. R. (1988). "Raman Spectroscopy" in *Spectrochemical Analysis*. Upper Saddle River, NJ: Prentice Hall, 499–512.
- Ishimaru, K., Hata, T., Bronsveld, P., Meier, D., and Imamura, Y. (2007). Spectroscopic analysis of carbonization behavior of wood, cellulose and lignin. *J. Mater. Sci.* 42, 122–129. doi:10.1007/s10853-006-1042-3
- Iversen, J., Berg, R., and Ahning, B. (2014). Quantitative monitoring of yeast fermentation using Raman spectroscopy. *Anal. Bioanal. Chem.* 406, 4911–4919. doi:10.1007/s00216-014-7897-2
- Jaaskelainen, A.-S., Saariaho, A.-M., and Vuorinen, T. (2005). Quantification of lignin and hexenuronic acid in bleached hardwood kraft pulps: a new calibration method for UVR spectroscopy and evaluation of the conventional methods. *J. Wood Chem. Technol.* 25, 51–65. doi:10.1081/WCT-200058239
- Jaaskelainen, A.-S., Saariaho, A.-M., Vyorykka, J., Vuorinen, T., Matousek, P., and Parker, A. W. (2006). Application of UV-Vis and resonance Raman spectroscopy to study bleaching and photoyellowing of thermomechanical pulps. *Holzforchung* 60, 231–238. doi:10.1515/HF.2006.038
- Kacurikova, M., Wellner, N., Ebringerova, A., Hromidkova, Z., Wilson, R. H., and Belton, P. S. (1998). Characterization of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies. *Food Hydrocoll.* 13, 35–41. doi:10.1016/S0268-005X(98)00067-8
- Keown, D. M., Li, X., Hayashi, J.-I., and Li, C.-Z. (2007). Characterization of the structural features of char from the pyrolysis of cane trash using Fourier transform-Raman spectroscopy. *Energy Fuels* 21, 1816–1821. doi:10.1021/ef070049r
- Keown, D. M., Li, X., Hayashi, J.-I., and Li, C.-Z. (2008). Evolution of biomass char structure during oxidation in O₂ as revealed with FT-Raman spectroscopy. *Fuel Process. Technol.* 89, 1429–1435. doi:10.1016/j.fuproc.2008.07.002
- Kihara, M., Takayama, M., Wariishi, H., and Tanaka, H. (2002). Determination of the carbonyl groups in native lignin utilizing Fourier transform Raman spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 58A, 2213–2221. doi:10.1016/S1386-1425(01)00693-X
- Kim, S. H., Lee, C. M., and Kaffe, K. (2013). Characterization of crystalline cellulose in biomass: basic principles, applications, and limitations of XRD, NMR, IR, Raman, and SFG. *Korean J. Chem. Eng.* 30, 2127–2141. doi:10.1007/s11814-013-0162-0
- Lähdetie, A., Liitiae, T., Tamminen, T., and Jaaskelainen, A.-S. (2009). Reflectance UV-vis and UV resonance Raman spectroscopy in characterization of kraft pulps. *Bioresources* 4, 1600–1619.
- Lähdetie, A., Liitiae, T., Tamminen, T., Pere, J., and Jaaskelainen, A.-S. (2009). Activation of thermomechanical pulp by laccases as studied by UV-Vis, UV resonance Raman and FTIR spectroscopy. *Holzforchung* 63, 745–750. doi:10.1515/HF.2009.125
- Lähdetie, A., Nousiainen, P., Sipilä, J., Tamminen, T., and Jaaskelainen, A.-S. (2013). Laser-induced fluorescence (LIF) of lignin and lignin model compounds in Raman spectroscopy. *Holzforchung* 67, 531–538. doi:10.1515/hf-2012-0177
- Landsberg, G., and Mandelstam, L. (1928). Eine neue Erscheinung bei der Lichtstreuung in Kristallen. *Naturwissenschaften* 16, 557–558.
- Larsen, K. L., and Barsberg, S. (2010). Theoretical and Raman spectroscopic studies of phenolic lignin model monomers. *J. Phys. Chem. B* 114, 8009–8021. doi:10.1021/jp1028239
- Larsen, K. L., and Barsberg, S. (2011). Environmental effects on the lignin model monomer, vanillyl alcohol, studied by Raman spectroscopy. *J. Phys. Chem. B* 115, 11470–11480. doi:10.1021/jp203910h
- Laureano-Perez, L., Dale, B. E., O'dwyer, J. P., and Holtzapple, M. (2006). Statistical correlation of spectroscopic analysis and enzymatic hydrolysis of poplar samples. *Biotechnol. Prog.* 22, 835–841. doi:10.1021/bp050284x
- Lewis, E. N., Treado, P. J., and Levin, I. W. (1993). A miniaturized, no-moving-parts Raman spectrometer. *Appl. Spectrosc.* 47, 539–543. doi:10.1366/0003702934067144
- Li, C., Cheng, G., Balan, V., Kent, M. S., Ong, M., Chundawat, S. P. S., et al. (2011). Influence of physico-chemical changes on enzymatic digestibility of ionic liquid and AFEX pretreated corn stover. *Bioresour. Technol.* 102, 6928–6936. doi:10.1016/j.biortech.2011.04.005
- Li, C., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V., Auer, M., et al. (2010). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. *Bioresour. Technol.* 101, 4900–4906. doi:10.1016/j.biortech.2009.10.066
- Li, C., Sun, L., Simmons, B. A., and Singh, S. (2013). Comparing the recalcitrance of eucalyptus, pine, and switchgrass using ionic liquid and dilute acid pretreatments. *Bioenergy Res.* 6, 14–23. doi:10.1007/s12155-012-9220-4
- Li, L., Wang, H., and Cheng, J.-X. (2005). Quantitative coherent anti-stokes Raman scattering imaging of lipid distribution in coexisting domains. *Biophys. J.* 89, 3480–3490. doi:10.1529/biophysj.105.065607
- Liang, C. Y., and Marchessault, R. H. (1959). Infrared spectra of crystalline polysaccharides. I. Hydrogen bonds in native celluloses. *J. Polym. Sci.* 37, 385–395. doi:10.1002/pol.1959.1203713209
- Lucas, M., Wagner, G. L., Nishiyama, Y., Hanson, L., Samayam, I. P., Schall, C. A., et al. (2011). Reversible swelling of the cell wall of poplar biomass by ionic liquid at room temperature. *Bioresour. Technol.* 102, 4518–4523. doi:10.1016/j.biortech.2010.12.087
- Lupoi, J. S., Healey, A., Singh, S., Sykes, R. W., Davis, M. F., Lee, D. J., et al. (2015). High-throughput prediction of *Acacia* and eucalypt lignin syringyl/guaiacyl content using FT-Raman spectroscopy and partial least squares modeling. *Bioenerg. Res.* doi:10.1007/s12155-015-9578-1
- Lupoi, J. S., Singh, S., Davis, M., Lee, D. J., Shepherd, M., Simmons, B. A., et al. (2014a). High-throughput prediction of eucalypt lignin syringyl/guaiacyl content using multivariate analysis: a comparison between mid-infrared, near-infrared, and Raman spectroscopies for model development. *Biotechnol. Biofuels* 7, 93. doi:10.1186/1754-6834-7-93
- Lupoi, J. S., Singh, S., Simmons, B. A., and Henry, R. J. (2014b). Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high-throughput techniques. *Bioenerg. Res.* 7, 1–23. doi:10.1007/s12155-013-9352-1
- Lupoi, J. S., and Smith, E. A. (2012). Characterization of woody and herbaceous biomasses lignin composition with 1064 nm dispersive multichannel Raman spectroscopy. *Appl. Spectrosc.* 66, 903–910. doi:10.1366/12-06621

- Lupoi, J.S. (2012). *Developments in Enzyme Immobilization and Near-Infrared Raman Spectroscopy with Downstream Renewable Energy Applications*. Ph.D., Iowa State University.
- McCreery, R. L. (2000). *Raman Spectroscopy for Chemical Analysis*. New York/Chichester/Weinheim/Brisbane/Singapore/Toronto: Wiley Interscience, 448.
- Meyer, M. W., Lupoi, J. S., and Smith, E. A. (2011). 1064 nm dispersive multichannel Raman spectroscopy for the analysis of plant lignin. *Anal. Chim. Acta* 706, 164–170. doi:10.1016/j.aca.2011.08.031
- Mononen, K., Jaaskelainen, A.-S., Alvilä, L., Pakkanen, T. T., and Vuorinen, T. (2005). Chemical changes in silver birch (*Betula pendula* Roth) wood caused by hydrogen peroxide bleaching and monitored by color measurement (CIELab) and UV-vis, FTIR and UVR spectroscopy. *Holzforschung* 59, 381–388. doi:10.1515/HF.2005.010
- National Institute of Standards and Technology (NIST). (2013). *Computational Chemistry Comparison and Benchmark Database*. Available from: cccbdb.nist.gov/pglist.asp
- Nuopponen, M., Vuorinen, T., Jaamsae, S., and Viitaniemi, P. (2004a). Thermal modifications in softwood studied by FT-IR and UV resonance Raman spectroscopies. *J. Wood Chem. Technol.* 24, 13–26. doi:10.1081/WCT-120035941
- Nuopponen, M., Willför, S., Jaaskelainen, A. S., Sundberg, A., and Vuorinen, T. (2004b). A UV resonance Raman (UVR) spectroscopic study on the extractable compounds of Scots pine (*Pinus sylvestris*) wood part I: lipophilic compounds. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 60A, 2953–2961. doi:10.1016/j.saa.2004.02.008
- Nuopponen, M., Willför, S., Jaaskelainen, A. S., and Vuorinen, T. (2004c). A UV resonance Raman (UVR) spectroscopic study on the extractable compounds in Scots pine (*Pinus sylvestris*) wood. Part II. Hydrophilic compounds. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 60, 2963–2968. doi:10.1016/j.saa.2004.02.007
- Nuopponen, M. H., Wikberg, H. I., Birch, G. M., Jaaskelainen, A.-S., Maunu, S. L., Vuorinen, T., et al. (2006). Characterization of 25 tropical hardwoods with Fourier transform infrared, ultraviolet resonance Raman, and 13C-NMR cross-polarization/magic-angle spinning spectroscopy. *J. Appl. Polym. Sci.* 102, 810–819. doi:10.1002/app.24143
- Ochoa, A., Aramburu, B., Ibanez, M., Valle, B., Bilbao, J., Gayubo, A. G., et al. (2014). Compositional insights and valorization pathways for carbonaceous material deposited during bio-oil thermal treatment. *ChemSusChem* 7, 2597–2608. doi:10.1002/cssc.201402276
- Ona, T., Sonoda, T., Ito, K., Shibata, M., Katayama, T., Kato, T., et al. (1998a). Non-destructive determination of lignin syringyl/guaiacyl monomeric composition in native wood by Fourier transform Raman spectroscopy. *J. Wood Chem. Technol.* 18, 43–51. doi:10.1080/02773819809349586
- Ona, T., Sonoda, T., Ito, K., Shibata, M., Kato, T., and Ootake, Y. (1998b). Determination of wood basic density by Fourier transform Raman spectroscopy. *J. Wood Chem. Technol.* 18, 367–379. doi:10.1080/02773819809349586
- Ona, T., Sonoda, T., Ito, K., Shibata, M., Kato, T., and Ootake, Y. (1998c). Non-destructive determination of hemicellulosic neutral sugar composition in native wood by Fourier transform Raman spectroscopy. *J. Wood Chem. Technol.* 18, 27–41. doi:10.1080/02773819809349586
- Ona, T., Sonoda, T., Ito, K., Shibata, M., Kato, T., and Ootake, Y. (1997). Non-destructive determination of wood constituents by Fourier-transform Raman spectroscopy. *J. Wood Chem. Technol.* 17, 399–417. doi:10.1080/02773819708003141
- Ona, T., Sonoda, T., Ito, K., Shibata, M., Kato, T., Ootake, Y., et al. (2000). Rapid prediction of native wood pulp properties by Fourier-transform Raman spectroscopy. *J. Pulp Pap. Sci.* 26, 43–47.
- Ona, T., Sonoda, T., Ohshima, J., Yokota, S., and Yoshizawa, N. (2003). A rapid quantitative method to assess eucalyptus wood properties for kraft pulp production by FT-Raman spectroscopy. *J. Pulp Pap. Sci.* 29, 6–10.
- Pandey, K. K., and Vuorinen, T. (2008). UV resonance Raman spectroscopic study of photodegradation of hardwood and softwood lignins by UV laser. *Holzforschung* 62, 183–188. doi:10.1515/HF.2008.046
- Perera, P. N., Schmidt, M., Schuck, P. J., and Adams, P. D. (2011). Blind image analysis for the compositional and structural characterization of plant cell walls. *Anal. Chim. Acta* 702, 172–177. doi:10.1016/j.aca.2011.06.021
- Pohling, C., Brackmann, C., Duarte, A., Backup, T., Enejder, A., and Motzkus, M. (2014). Chemical imaging of lignocellulosic biomass by CARS microscopy. *J. Biophotonics* 7, 126–134. doi:10.1002/jbio.201300052
- Popp, J. A. K. W. (2006). “Raman scattering, fundamentals,” in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers (Hoboken, NJ: John Wiley & Sons), 13104–13139.
- Proniewicz, L. M., Paluszkiwicz, C., Weselucha-Birczynska, A., Baranski, A., and Dutka, D. (2002). FT-IR and FT-Raman study of hydrothermally degraded groundwood-containing paper. *J. Mol. Struct.* 614, 345–353. doi:10.1016/S0022-2860(02)00275-2
- Raman, C. V., and Krishnan, K. S. (1928). A new type of secondary radiation. *Nature* 121, 501–502. doi:10.1038/121501c0
- Roder, T., and Sixta, H. (2005). Confocal Raman spectroscopy – applications on wood, pulp, and cellulose fibres. *Macromol. Symp.* 223, 57–66. doi:10.1002/masy.200550504
- Saar, B. G., Zeng, Y., Freudiger, C. W., Liu, Y.-S., Himmel, M. E., Xie, X. S., et al. (2010). Label-free, real-time monitoring of biomass processing with stimulated Raman scattering microscopy. *Angew. Chem. Int. Ed. Engl.* 49, 5476–5479. doi:10.1002/anie.201000900
- Saariaho, A.-M., Argyropoulos, D. S., Jaaskelainen, A.-S., and Vuorinen, T. (2005). Development of the partial least squares models for the interpretation of the UV resonance Raman spectra of lignin model compounds. *Vib. Spectrosc.* 37, 111–121. doi:10.1016/j.vibspec.2004.08.001
- Saariaho, A.-M., Jaaskelainen, A.-S., Nuopponen, M., and Vuorinen, T. (2003). Ultra-violet resonance Raman spectroscopy in lignin analysis: determination of characteristic vibrations of p-hydroxyphenyl, guaiacyl, and syringyl lignin structures. *Appl. Spectrosc.* 57, 58–66. doi:10.1366/00037020321165214
- Saariaho, A.-M., Jaaskelainen, A.-S., Matousek, P., Towrie, M., Parker, A. W., and Vuorinen, T. (2004). Resonance Raman spectroscopy of highly fluorescing lignin containing chemical pulps: suppression of fluorescence with an optical Kerr gate. *Holzforschung* 58, 82–90. doi:10.1515/HF.2004.011
- Schenzel, K., Almloef, H., and Germgard, U. (2009). Quantitative analysis of the transformation process of cellulose I → cellulose II using NIR FT Raman spectroscopy and chemometric methods. *Cellulose (Dordrecht, Neth.)* 16, 407–415. doi:10.1007/s10570-009-9286-0
- Schenzel, K., and Fischer, S. (2001). NIR FT Raman spectroscopy – a rapid analytical tool for detecting the transformation of cellulose polymorphs. *Cellulose (Dordrecht, Neth.)* 8, 49–57. doi:10.1023/A:1016616920539
- Schenzel, K., Fischer, S., and Brendler, E. (2005). New method for determining the degree of cellulose I crystallinity by means of FT Raman spectroscopy. *Cellulose (Dordrecht, Neth.)* 12, 223–231. doi:10.1007/s10570-004-3885-6
- Schulz, H., and Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* 43, 13–25. doi:10.1016/j.vibspec.2006.06.001
- Sene, C. F. B., McCann, M. C., Wilson, R. H., and Grinter, R. (1994). Fourier-transform Raman and Fourier-transform infrared spectroscopy. An investigation of five higher plant cell walls and their components. *Plant Physiol.* 106, 1623–1631.
- Shih, C.-J., Lupoi, J. S., and Smith, E. A. (2011). Raman spectroscopy measurements of glucose and xylose in hydrolysate: role of corn stover pretreatment and enzyme composition. *Bioresour. Technol.* 102, 5169–5176. doi:10.1016/j.biortech.2011.01.043
- Shih, C.-J., and Smith, E. A. (2009). Determination of glucose and ethanol after enzymatic hydrolysis and fermentation of biomass using Raman spectroscopy. *Anal. Chim. Acta* 653, 200–206. doi:10.1016/j.aca.2009.09.012
- Sivakesava, S., Irudayaraj, J., and Ali, D. (2001a). Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques. *Process Biochem.* 37, 371–378. doi:10.1016/S0032-9592(01)00223-0
- Sivakesava, S., Irudayaraj, J., and Demirci, A. (2001b). Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. *J. Ind. Microbiol. Biotechnol.* 26, 185–190. doi:10.1038/sj.jim.7000124
- Smekal, A. (1923). Zur Quantentheorie der dispersion. *Naturwissenschaften* 11, 873–875. doi:10.1007/BF01576902
- Smith, W., and Dent, G. (2005). *Modern Raman Spectroscopy*. Chichester: John Wiley & Sons.
- Sun, L., Li, C., Xue, Z., Simmons, B. A., and Singh, S. (2013). Unveiling high-resolution, tissue specific dynamic changes in corn stover during ionic liquid pretreatment. *RSC Adv.* 3, 2017–2027. doi:10.1039/C2RA20706K
- Sun, L., Simmons, B. A., and Singh, S. (2010). Understanding tissue specific compositions of bioenergy feedstocks through hyperspectral Raman imaging. *Biotechnol. Bioeng.* 108, 286–295. doi:10.1002/bit.22931

- Sun, L., Varanasi, P., Yang, F., Loque, D., Simmons, B. A., and Singh, S. (2012). Rapid determination of syringyl:guaiacyl ratios using FT-Raman spectroscopy. *Biotechnol. Bioeng.* 109, 647–656. doi:10.1002/bit.24348
- Takayama, M., Johjima, T., Yamanaka, T., Wariishi, H., and Tanaka, H. (1997). Fourier-transform Raman assignment of guaiacyl and syringyl marker bands for lignin determination. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 53A, 1621–1628. doi:10.1016/S1386-1425(97)00100-5
- Tsaneva, V. N., Kwapinski, W., Teng, X., and Glowacki, B. A. (2014). Assessment of the structural evolution of carbons from microwave plasma natural gas reforming and biomass pyrolysis using Raman spectroscopy. *Carbon N. Y.* 80, 617–628. doi:10.1016/j.carbon.2014.09.005
- Vester, J., Felby, C., Nielsen, O. F., and Barsberg, S. (2004). Fourier transform Raman difference spectroscopy for detection of lignin oxidation products in thermomechanical pulp. *Appl. Spectrosc.* 58, 404–409. doi:10.1366/000370204773580239
- Warsta, E., Lahdetie, A., Jaaskelainen, A.-S., and Vuorinen, T. (2012). Effect of pH on lignin analysis by Raman spectroscopy. *Holzforschung* 66, 451–457. doi:10.1515/hf.2011.176
- Wiley, J. H., and Atalla, R. H. (1987). Band assignments in the Raman spectra of celluloses. *Carbohydr. Res.* 160, 113–129. doi:10.1016/0008-6215(87)80306-3
- Wójciak, A., Kasprzyk, H., Sikorska, E., Krawczyk, A., Sikorski, M., and Weselucha-Birczynska, A. (2014). FT-Raman, FT-infrared and NIR spectroscopic characterization of oxygen-delignified kraft pulp treated with hydrogen peroxide under acidic and alkaline conditions. *Vib. Spectrosc.* 71, 62–69. doi:10.1016/j.vibspec.2014.01.007
- Zakis, G. F. (1994). *Functional Analysis of Lignins and Their Derivatives*. Atlanta, GA: Tappi Press.
- Zakzeski, J., Bruijninx, P. C. A., and Weckhuysen, B. M. (2011). In situ spectroscopic investigation of the cobalt-catalyzed oxidation of lignin model compounds in ionic liquids. *Green Chem.* 13, 671–680. doi:10.1039/c0gc00437e
- Zeng, Y., Himmel, M. E., and Ding, S.-Y. (2012). Coherent Raman microscopy analysis of plant cell walls. *Methods Mol. Biol.* 908, 49–60. doi:10.1007/978-1-61779-956-3_5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 February 2015; paper pending published: 19 March 2015; accepted: 27 March 2015; published online: 20 April 2015.

Citation: Lupoi JS, Gjersing E and Davis MF (2015) Evaluating lignocellulosic biomass, its derivatives, and downstream products with Raman spectroscopy. *Front. Bioeng. Biotechnol.* 3:50. doi: 10.3389/fbioe.2015.00050

This article was submitted to *Bioenergy and Biofuels*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Lupoi, Gjersing and Davis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Standard flow liquid chromatography for shotgun proteomics in bioenergy research

Susana M. González Fernández-Niño¹, A. Michelle Smith-Moritz¹, Leanne Jade G. Chan¹, Paul D. Adams^{1,2}, Joshua L. Heazlewood^{1,3} and Christopher J. Petzold^{1*}

¹ Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

² Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

³ Australian Research Council Centre of Excellence in Plant Cell Walls, School of Botany, The University of Melbourne, Melbourne, VIC, Australia

Edited by:

Robert Henry, The University of Queensland, Australia

Reviewed by:

Qaisar Mahmood, COMSATS Institute of Information Technology, Pakistan

Yu-Shen Cheng, University of California Davis, USA

*Correspondence:

Christopher J. Petzold, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, 5885 Hollis Street, 4th Floor, Emeryville, CA 94608, USA
e-mail: cjpetzold@lbl.gov

Over the past 10 years, the bioenergy field has realized significant achievements that have encouraged many follow on efforts centered on biosynthetic production of fuel-like compounds. Key to the success of these efforts has been transformational developments in feedstock characterization and metabolic engineering of biofuel-producing microbes. Lagging far behind these advancements are analytical methods to characterize and quantify systems of interest to the bioenergy field. In particular, the utilization of proteomics, while valuable for identifying novel enzymes and diagnosing problems associated with biofuel-producing microbes, is limited by a lack of robustness and limited throughput. Nano-flow liquid chromatography coupled to high-mass accuracy, high-resolution mass spectrometers has become the dominant approach for the analysis of complex proteomic samples, yet such assays still require dedicated experts for data acquisition, analysis, and instrument upkeep. The recent adoption of standard flow chromatography (ca. 0.5 mL/min) for targeted proteomics has highlighted the robust nature and increased throughput of this approach for sample analysis. Consequently, we assessed the applicability of standard flow liquid chromatography for shotgun proteomics using samples from *Escherichia coli* and *Arabidopsis thaliana*, organisms commonly used as model systems for lignocellulosic biofuels research. Employing 120 min gradients with standard flow chromatography, we were able to routinely identify nearly 800 proteins from *E. coli* samples; while for samples from *Arabidopsis*, over 1,000 proteins could be reliably identified. An examination of identified peptides indicated that the method was suitable for reproducible applications in shotgun proteomics. Standard flow liquid chromatography for shotgun proteomics provides a robust approach for the analysis of complex samples. To the best of our knowledge, this study represents the first attempt to validate the standard flow approach for shotgun proteomics.

Keywords: proteomics, standard flow chromatography, biofuels, mass spectrometry

INTRODUCTION

Advances in biofuels research focusing on feedstock characterization and engineering (Persil-Cetinkol et al., 2012; DeMartini et al., 2013; Shen et al., 2013; Eudes et al., 2014) as well as the genetic manipulation of microbes (Alper et al., 2006; Tyo et al., 2007; Keasling, 2008; Lee et al., 2008) have progressed significantly in the last few years. Unfortunately, analytical capabilities required to efficiently monitor and assess these changes are lagging. Furthermore, many modern bioanalytical techniques are focused toward medical and health related research, which have significantly different priorities and requirements for success. Most biotechnology research challenges are not constrained by sensitivity or resolution of an assay, rather they depend on accurate identification and quantitation of target molecules for a large number of samples. Consequently, an important component for biotechnological research is sample throughput supported by a robust analytical platform. Recent advances in proteomics and metabolomics have

focused on liquid chromatography-mass spectrometry (LC-MS) methods by increasing their sensitivity to aid discovery-based research efforts. This is most evident with the development of nano-LC couple to high-resolution mass spectrometers; yet, this technology is yet to mature into a robust platform capable of consistently analyzing hundreds of samples per week. Consequently, alternate technologies capable of answering the questions necessary for biotechnology progress are needed. Recently, we published a high throughput targeted proteomic toolkit based on standard flow chromatography coupled to mass spectrometry (MS) to help address these issues for *Escherichia coli*; however, a significant amount of methods development was necessary. This work prompted us to assess the utility of standard flow liquid chromatography (LC) for shotgun proteomic methods related to biotechnology.

The discipline of proteomics has been dominated by nano-flow LC coupled to MS since its early development over 20 years ago

(Emmett and Caprioli, 1994; Gatlin et al., 1998). The adoption of nano-flow LC for protein identification was driven by the substantial increases in sensitivity and detection capabilities of nano-flow (ca. 500 nL/min) over capillary (ca. 50 μ L/min) and standard flow (ca. 0.5 mL/min) chromatography. Typically, shotgun proteomic studies utilize nano-flow chromatography methods due to limited amounts of sample and to obtain optimal ionization efficiency for sensitive detection of peptides but at the cost of ease of use and system robustness (Gapeev et al., 2009). These issues are particularly problematic for biotechnology research that depends heavily on high sample throughput. Recently, ultra-high performance LC coupled to triple quadrupole mass spectrometers has been shown to yield comparable sensitivity and better analytical metrics (coefficient of variation, dynamic range) than nano-flow LC for MRM-based analysis of biomarker proteins when the amount of sample is adjusted to the column size (Percy et al., 2012). Additionally, the well-established robust operation of standard flow chromatography makes this instrumentation attractive for applications that rely on high sample throughput, consistent results, and less system downtime (Swartz, 2005). For applications where sample abundance is not limited, one of the greatest concerns with ultra-high pressure liquid chromatography-mass spectrometry (UHPLC-MS) workflows is the loss of sensitivity, relative to nano-LC-MS workflows, leading to datasets of insufficient depth to answer questions of interest.

The adoption of a UHPLC-MS workflow for sample delivery requires both efficient ionization and instrument speed to handle both the sample delivery rate and reduced elution times for peptides. The past decade has seen significant developments and advances in instrumentation associated with proteomics-based MS. Advances in reversed phase C18 chromatography columns yield greater separation efficiency and more stable retention times. The current generation of instrumentation is faster, more sensitive, and is better able to deal with the dynamic range inherent in biological samples. Moreover, the development and adoption of off-axis nebulizers for sample delivery when using electrospray ionization significantly reduces contamination of capillaries and skimmers (Banerjee and Mazumdar, 2012). Collectively, these improvements have enabled the development of a UHPLC-MS workflow for MRM-based analysis of biomarker proteins (Percy et al., 2012). Consequently, we were interested in assessing the capacity of this workflow in shotgun proteomic experiments. The work described here details the results of an analysis of a prokaryote (*E. coli*) and an eukaryote (*Arabidopsis thaliana*) whole cell proteomes on an Agilent UHPLC-QTOFMS system, but would be generally applicable for any current generation of tandem mass spectrometer being utilized for shotgun proteomics.

MATERIALS AND METHODS

PROTEIN EXTRACTION

Protein was extracted using standard techniques with analytical reagents where suitable. For *E. coli* DH5 α samples, cell lysis and protein precipitation was accomplished using a chloroform/methanol precipitation. A 100 μ L aliquot of cells was transferred to a 1.7 mL tube, followed by the addition of 400 μ L of methanol, 100 μ L of chloroform, and 300 μ L of water, with mixing by vortex after each addition. Following centrifugation at

21,000 $\times g$ for phase separation, the methanol and water layer was removed and 300 μ L of methanol was added. The tube was briefly vortexed to dislodge the protein pellet, then centrifuged at 21,000 $\times g$ for 2 min. The chloroform and methanol layer was removed and the protein pellet was dried for 5 min in a vacuum concentrator. The protein pellet was re-suspended in 100 mM (NH₄)HCO₃ with 20% methanol, reduced with 5 mM TCEP [Tris(2-carboxyethyl)phosphine hydrochloride] for 30 min at room temperature, treated with 10 mM iodoacetamide (IAA) for 30 min in the dark at room temperature, and digested with trypsin (1:50 w/w) overnight at 37°C. Aliquots of 40 μ g were taken for analysis by LC-MS/MS. For *A. thaliana* (L.) Heynh. (ecotype Landsberg erecta), protein was extracted from a previously described heterotrophic cell culture (Ito et al., 2011). A total of 1 g plant material (fresh weight) was used for the isolation of total protein. The plant material was harvested and frozen with liquid nitrogen in an Eppendorf tube with two small steel balls. The protein extraction was performed by the addition of 0.4 mL of fresh disruption buffer [125 mM Tris-HCl, 7% (w/v) SDS, and 10% β -mercaptoethanol], followed by vortex for 10 min at room temperature. The samples were centrifuged at 10,000 $\times g$ for 5 min at 4°C and the supernatant separated into two 2 mL tubes. Samples were further extracted in 800 μ L methanol and mixed, then 200 μ L chloroform added and mixed, and finally 500 μ L of ddH₂O added and vortexed (30 sec each time). Samples were centrifuged for 5 min at 10,000 $\times g$ at 4°C, the aqueous phase removed, and 500 μ L of methanol added. Samples were vortexed for 30 s and centrifuged for 10 min at 9,000 $\times g$ at 4°C. The supernatant was discarded and the pellet air-dried. The dried pellet was suspended in 200 μ L of re-suspension buffer [3M urea, 50 mM (NH₄)HCO₃, and 5 mM dithiothreitol, pH 8], and incubated with IAA at a final concentration 10 mM for 30 min in the dark. Prior to analysis by MS, 40 μ g of extracted protein was digested with trypsin (1:10 w/w) overnight at 37°C. Peptides were desalted using C18 Micro SpinColumns (Harvard Apparatus) as previously outlined (Parsons et al., 2013). Eluted peptides were re-suspended in 2% acetonitrile, 0.1% formic acid prior to analysis by LC-MS/MS.

STANDARD FLOW MASS SPECTROMETRY

All samples were analyzed on an Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies) coupled to an Agilent 1290 UHPLC system. Peptide samples were loaded onto a Sigma-Aldrich Ascentis Peptides ES-C18 column (2.1 mm \times 100 mm, 2.7 μ m particle size, operated at 60°C) via an Infinity Autosampler (Agilent Technologies) with Buffer A (2% acetonitrile, 0.1% formic acid) flowing at 0.400 mL/min. Peptides were eluted into the mass spectrometer via a gradient with initial starting condition of 5% buffer B (98% acetonitrile, 0.1% formic acid). For analysis of all samples, buffer B was increased to 35% over 120 min. Buffer B was then increased to 50% over 5 min, then up to 90% over 1 min, and held for 7 min at a flow rate of 0.6 mL/min, followed by a ramp back down to 5% B over 1 min where it was held for 6 min to re-equilibrate the column to original conditions. Peptides were introduced to the mass spectrometer from the LC by using a Jet Stream source (Agilent Technologies) operating in positive-ion mode (3,500 V). Source parameters employed gas temp (250°C), drying gas (14 L/min), nebulizer (35 psig), sheath

gas temp (250°C), sheath gas flow (11 L/min), VCap (3,500 V), fragmentor (180 V), OCT 1 RF Vpp (750 V). The data were acquired with Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.05.00 (Build 5.0.5042.2) operating in Auto MS/MS mode whereby the 20 most intense ions (charge states, 2–5) within 300–1,400 m/z mass range above a threshold of 1,500 counts were selected for MS/MS analysis. MS/MS spectra (100–1,700 m/z) were collected with the quadrupole set to “Medium” resolution and were acquired until 45,000 total counts were collected or for a maximum accumulation time of 333 ms. Former parent ions were excluded for 0.1 min following MS/MS acquisition.

LC-MS/MS DATA ANALYSIS AND INTEGRATION

The acquired data were exported as .mgf files using the Export as MGF function of the MassHunter Workstation Software, Qualitative Analysis (Version B.05.00 Build 5.0.519.13 Service Pack 1, Agilent Technologies) using the following settings: peak Filters (MS/MS), the absolute height (≥ 20 counts), relative height ($\geq 0.100\%$ of largest peak), maximum number of peaks (300) by height; for charge state (MS/MS), the peak spacing tolerance (0.0025 m/z plus 7.0 ppm), isotope model (peptides), charge state limit assigned to (5) maximum. Resultant data files were interrogated with the Mascot search engine version 2.3.02 (Matrix Science) with a peptide tolerance of ± 50 ppm and MS/MS tolerance of ± 0.1 Da; fixed modifications Carbamidomethyl (C); variable modifications Oxidation (M); up to one missed cleavage for trypsin; peptide charge 2+, 3+, and 4+; and the instrument type was set to ESI-QUAD-TOF. Searches were performed against either an *E. coli* (strain K12) dataset obtained from UniProt (Magrane and Consortium, 2011) or the latest release of the *A. thaliana* dataset comprising TAIR10 obtained from The Arabidopsis Information Resource (Lamesch et al., 2012). Both databases incorporated proteins comprising the common Repository of Adventitious Proteins (cRAP v2012.01.01 from The Global Proteome Machine). The *E. coli* database comprised 4,429 sequences (1,398,775 residues) while the *Arabidopsis* database comprised 35,508 sequences (14,522,421 residues). Protein and peptide matches identified after interrogation of MS/MS data by Mascot were filtered and validated using Scaffold v4.3.0 (Proteome Software Inc.). Peptide identifications were accepted if they could be established at $>95.0\%$ probability by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at $>95.0\%$ probability and contained at least 1 identified peptide (at 95% and greater). Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). This resulted in false discovery rates of 0.9 (*E. coli*) and 0.3% (*Arabidopsis*) for protein and 0.37 (*E. coli*) and 1.66% (*Arabidopsis*) for peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

STATISTICAL DATA ANALYSIS

Statistical analysis was performed using the statistical toolbox in MATLAB v2009b (Mathworks). The following analysis was performed on data from both *Arabidopsis* and *E. coli*. After data

filtering using Scaffold v4.3.0 (outlined above), peptide data were exported as .csv files and only peptides derived from the same protein identified across all replicates were considered. The Mascot ions score and total ion current values were only used for the best matching peptide (based on ion score) in each replicate. The total spectrum count for each peptide from each replicate was calculated manually. The coefficient of variation (CV: SD/mean) of total spectrum counts, Mascot ions score, and total ion current for each of the common peptides across the replicates was calculated to determine the variation across the technical replicates. The histograms of the CV were plotted to determine distribution and general extreme value fit algorithm was used to determine fit and mean of the distribution. For principal component analysis (PCA), the total spectrum count, Mascot ions score, and total ion current for all the common peptides of each technical replicate were mean centered. PCA was done by eigenvalue decomposition of the data covariance, resulting in a set of linearly uncorrelated variables. The principal component scores were then plotted to identify groupings.

RESULTS AND DISCUSSION

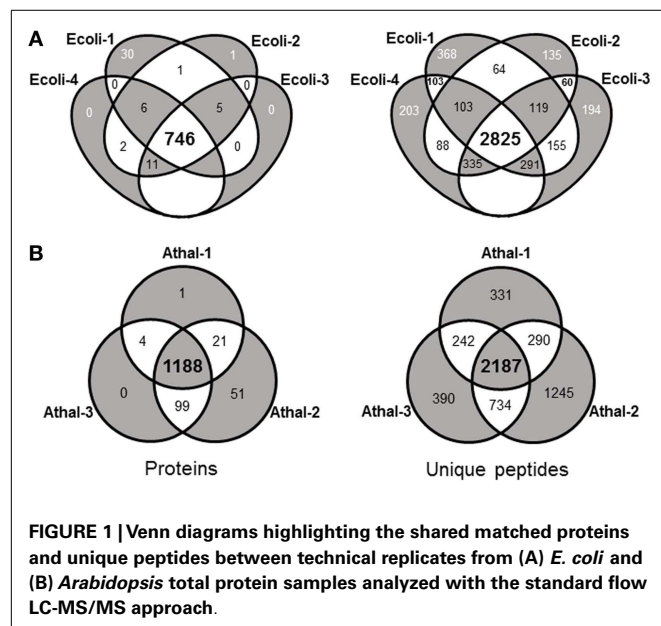
The initial setup and tuning of the standard flow LC-MS/MS parameters was undertaken using 20 fmol aliquots of trypsin digested bovine serum albumin (BSA). The system settings were deemed adequate once identification of BSA was comparable (i.e., unique peptides, total peptides, and coverage) to that achieved using previously benchmarked nano-flow LC-MS/MS approaches on a variety of instruments over the past 5 years.

APPLICATION OF STANDARD FLOW LC-MS/MS WITH PROKARYOTIC SAMPLES

Initial experiments were performed on *E. coli*, a well-characterized organism with minimal proteomic complexity. The *E. coli* genome of the widely utilized laboratory strain K-12 was completed nearly 20 years ago (Blattner et al., 1997). The genome has undergone multiple revisions and is estimated to codes for over 4,300 proteins (Magrane and Consortium, 2011). Shake flasks of *E. coli* DH5 α cultures were grown aerobically at 37°C on Luria Broth (LB) medium supplemented with 1% glucose. Total protein was extracted and digested overnight with trypsin at 37°C. The analysis procedure for *E. coli* samples was developed to take advantage of the speed and robustness of a standard flow analysis and as such, a method was developed that incorporated a 120 min gradient. A total of four biological replicates each equivalent to 40 μ g of peptide were analyzed by a standard flow UHPLC-QTOFMS operating with typical shotgun proteomics data acquisition parameters. The samples yielded on average $31,938 \pm 989$ (SE) MS/MS spectra (Table 1). Each of these MS/MS datasets was used to query the *E. coli* (K12) protein database from UniProt (Magrane and Consortium, 2011) using the Mascot search engine (Matrix Science) with resultant protein matches filtered using Scaffold (Proteome Software). In total, 802 proteins were identified from the four replicates with an average of 786 ± 4.5 (SE) per sample (Table 1 and Table S1 in Supplementary Material). This represents about 20% of the protein coding capacity of *E. coli*. On average, $14,639 \pm 484$ (SE) MS/MS spectra were successfully matched to a peptide, corresponding to around 46% of the total queries for

Table 1 | Values obtained from *E. coli* (Ec) and *Arabidopsis* (At) samples analyzed using the standard flow technique.

	Ec 1	Ec 2	Ec 3	Ec 4	At 1	At 2	At 3
Queries	31,452	29,348	33,638	33,315	26,135	37,388	28,203
Spectra	14,677	13,265	15,439	15,173	6,038	9,712	6,846
Spectra matched (%)	47	45	46	46	23	26	24
Unique peptides	4,028	3,729	4,154	4,123	3,050	6,643	3,553
Total proteins	788	772	791	791	1,214	1,359	1,291



each sample (Table 1). This would be regarded as a high conversion rate indicating that the acquired MS/MS was of sufficient quality to confidently assign nearly 50% of the queries using the approach. The consistency of the standard flow approach was highlighted by the total number of proteins assigned in each sample. A total of 768 proteins were identified in three out of four replicates, while 746 proteins (93%) were identified collectively in all four samples. This compares well to the total number of proteins identified in all samples (802). Few proteins were identified uniquely in a single replicate except for replicate Ecoli-1, where 30 unique proteins were identified (Figure 1A). Consequently, we sought to examine whether this setup was adequately dealing with the sample complexity given the flow rate and an approximate peptide elution time of around 6 s (Figure S1 in Supplementary Material). Nearly 75% of the matched MS/MS spectra were redundant, with an average of 4,009 unique peptides per sample (Table 1). Only 2,825 unique spectra (54%) were shared between all four samples; however, the new unique peptides (2,393) accounted for only 56 new proteins (Figure 1A). The approach consistently identifies the major proteins in a sample and a majority of unique spectra assigned between replicates are derived from previously identified proteins. This indicates that the QTOFMS has the capacity to handle the standard flow UHPLC setup in shotgun mode even at these higher flow rates and reduced peptide elution times.

APPLICATION OF STANDARD FLOW LC-MS/MS TO A COMPLEX EUKARYOTIC SAMPLE

Next, we were interested in assessing the suitability of the standard flow proteomics platform on a complex eukaryotic proteome. The genome of the reference plant *Arabidopsis* was completed over a decade ago (*Arabidopsis* Genome Initiative, 2000) and now represents one of the most highly curated and annotated genomes in biology. As a consequence, *Arabidopsis* is being utilized as a proving ground for plant synthetic biology approaches, many of which have focused on biomass manipulation for efficient biofuels production (Eudes et al., 2012, 2015). The most recent proteome release (TAIR10) from The *Arabidopsis* Information Resource (Lamesch et al., 2012) comprises over 27,000 loci and over 35,000 distinct protein products. Protein extractions were performed from a 7-day old *Arabidopsis* cell cultures that have previously been extensively employed for proteomic assays (Parsons et al., 2012). A total protein sample was analyzed in triplicate by UHPLC-QTOFMS using the standard flow proteomics platform. We analyzed the equivalent of ca. 40 µg of digested total protein over a 120 min gradient with MS conditions identical to those used with the *E. coli* samples. An average of $30,575 \pm 3,458$ (SE) MS/MS spectra were collected over the 2-hour run from the three replicates (Table 1). These numbers are similar to those obtained using the *E. coli* samples and may reflect the upper capacity of the method given the increased peptide complexity that would be expected from a eukaryotic sample.

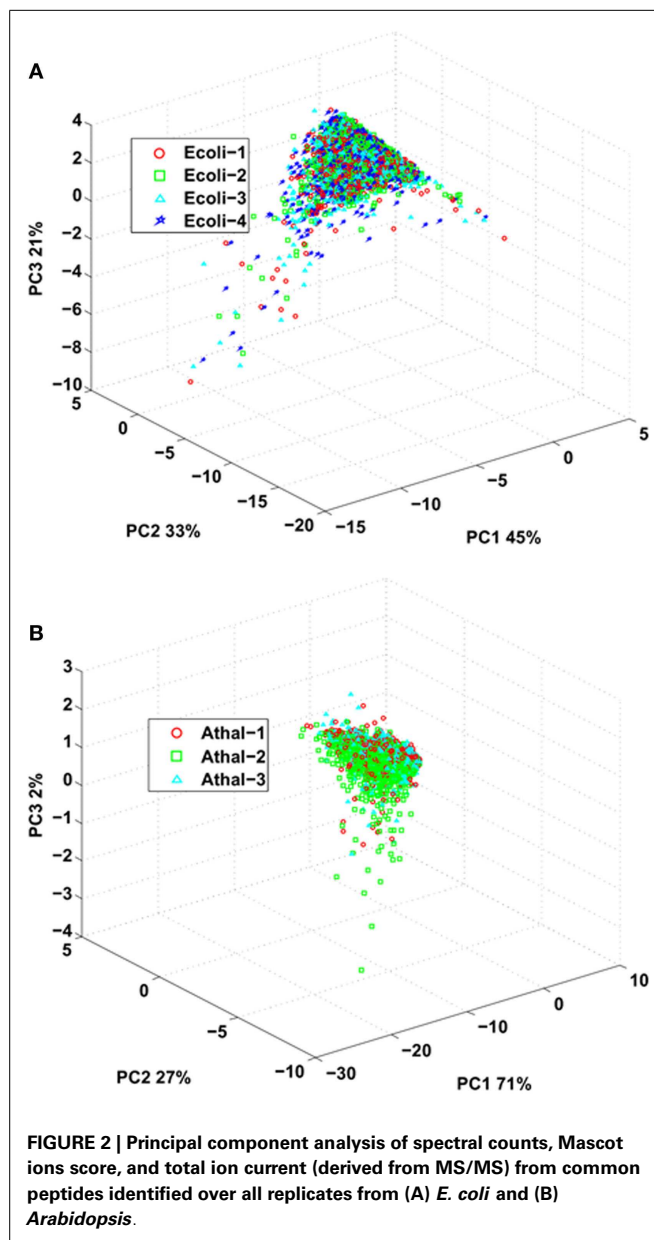
These *Arabidopsis* datasets were each used to interrogate the most recent *Arabidopsis* protein dataset using the Mascot search engine and matches filtered and integrated using Scaffold. This resulted in an average of $7,532 \pm 1,115$ (SE) matched spectra from the three replicates and corresponds to about 25% of the total spectra obtained (Table 1). This corresponds to about half the conversion rate observed for the *E. coli* samples and likely reflects the quality of the MS/MS spectra given the overall ion intensities are likely on average considerably lower due to the increased size of the proteome. The average number of proteins identified over the three samples was $1,288 \pm 73$ (SE) with a total of 1,364 unique proteins identified in all three replicates (Table S2 in Supplementary Material). The number of proteins consistently identified by all three replicates was 1,188 proteins or 87% of the total number (Figure 1B). The minor variation in identifications between these technical replicates demonstrates the reproducibility of the standard flow approach for shotgun proteomics as the method was capable of consistently identify the same proteins in each of the replicates. In an attempt to understand whether the approach is adequately dealing with the increased complexity of this eukaryotic sample, we examined the proportion of unique

matched peptides identified in each replicate. The 1,364 unique proteins identified in these samples were matched using 5,419 unique peptides, while the 1,188 proteins identified in all three replicates only required 2,187 (40.4% of the total) unique peptides (**Figure 1B**). These results indicate that again, as found with the *E. coli* samples, the majority of peptides exclusive to a given replicate are derived from proteins that have already yielded a high scoring peptide match. It is conceivable that the complexity of the eukaryotic sample contributes to many more co-eluting peptides during the 120 min gradient than the *E. coli* sample, resulting in unique peptides selected for MS/MS by the data-dependent acquisition method in each replicate (**Figure 1B**). Although similar results with regard to unique peptides were also observed with the *E. coli* samples, on average, an *E. coli* protein was identified by 6.8 unique peptides (768 proteins from 5,218 unique peptides) compared to the *Arabidopsis* proteins at 4.0 unique peptides each (1,364 proteins from 5,419 unique peptides). Nearly twice as many proteins were identified in the *Arabidopsis* samples (1,364 proteins compared to 768 proteins), which would be expected given the differences in coding capacity between these species. However, considerably more identified proteins would have been expected given proteome sizes, indicating that the standard flow approach has a more limited capacity to deal with the eukaryotic sample due to issues, such as increased sample complexity at any given point in time, lower overall ion intensity, and dynamic range limitations (as discussed above) of this QTOFMS system.

ASSESSING THE PERFORMANCE OF STANDARD FLOW LC-MS/MS

Last, we investigated the reproducibility of the standard flow UHPLC-QTOFMS by comparing parameters of the identified peptides between the replicates for both *E. coli* and *Arabidopsis* samples. The peptide ions score obtained from Mascot after data interrogation can be indicative of the quality of the fragmentation spectra (Perkins et al., 1999), total ion current obtained from the MS/MS spectra can provide information about the intensity of an eluted peptide (Asara et al., 2008) while total spectral count can provide an indication of peptide intensity and spectral complexity (Lundgren et al., 2010). These values can be used as a proxy to assess sample limitations and reproducibility by the LC-MS system. PCA was employed to ascertain whether there were any differences between ion scores, total ion current, and spectral count for peptides identified in all the replicates for either *E. coli* or *Arabidopsis* samples (**Figure 2**). The analysis demonstrated that none of the replicates for either *E. coli* or *Arabidopsis* could be separated by principal component scores for these attributes, indicating that majority of the differences for either data set can be largely attributed to biological heterogeneity of the sample (**Figure 2**).

The similarities in ion score, total ion current, and spectral counts can also be observed when analyzing the distributions of CV for the each identified peptides across the replicates (Figure S2 in Supplementary Material). The variations in Mascot ion score and total ion current (derived from MS/MS spectra) for the peptides identified over the replicates were similar for both *E. coli* and *Arabidopsis*. This indicates that from sample to sample, identical ions performed similarly with regard to the intensity of matched



peptides (total ion current) as measured by the mass spectrometer. This is further supported by the small variation in Mascot ion scores with ion intensity having a relationship to the quality of the MS/MS spectra and subsequent spectral matching procedures. The variation in total spectrum count was more pronounced between the identified peptides of *E. coli* (0.67) and *Arabidopsis* (0.32), with samples of lower complexity (i.e., *E. coli*) having a larger variation in spectral counts for a peptide across replicates. This could be due to the higher repeat sampling rate that likely occurred during analysis of *E. coli* samples due to the reduced number of distinct ions/peptides in the sample. This conclusion is supported when looking at the average number of spectral counts for a given identified peptide; for *E. coli*, it was 3.74 spectra per peptide while for *Arabidopsis* it was 2.70. The lower average number of spectral

counts for a peptide from *Arabidopsis* is likely indicative of the increased sample complexity. Collectively, these data indicate that between the replicates, the quality and intensities of peptides identified across the replicates was similar, indicating that the standard flow approach was not significantly impeding the performance of the mass spectrometer to acquire tandem mass spectra.

CONCLUSION

The recent progress of bioenergy research has relied heavily on transformational developments in feedstock characterization and metabolic engineering. Yet, omics methods to characterize and quantify systems of interest have mainly been adapted from the health and clinical fields that have very different research needs (i.e., high sensitivity, deep proteome coverage). We report the application of a standard flow LC-MS/MS approach that is suitable for large numbers of shotgun proteomic experiments where sample abundance is not limiting. The setup is capable of undertaking a rapid analysis of low complexity samples as well as handling highly complex samples by employing extended analysis times. Although its application requires instrumentation with the ability to deal with increased flow rates and shorter peptide elution times, it is apparent that the current generation of tandem mass spectrometers is capable of handling these parameters. While traditional nano-LC-MS/MS approaches are likely to continue to dominate shotgun analyses as they produce a greater number of protein identifications and have the ability to deal with increased sample complexity and dynamic range, the robust nature and simplicity of standard flow coupled to MS makes this approach an attractive alternative for applications where sample throughput and reproducibility are important factors.

AUTHOR CONTRIBUTIONS

JH and CP conceived and advised in all aspects of the study. PA supervised all aspects of the study. SG and LC performed experiments. AS analyzed the data. JH and CP interpreted the data and wrote the manuscript. All authors discussed and commented on the manuscript.

ACKNOWLEDGMENTS

This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00044/abstract>

REFERENCES

- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R., and Stephanopoulos, G. (2006). Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314, 1565–1568. doi:10.1126/science.1131969
- Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi:10.1038/35048692
- Asara, J. M., Christofk, H. R., Freemark, L. M., and Cantley, L. C. (2008). A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* 8, 994–999. doi:10.1002/pmic.200700426
- Banerjee, S., and Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* 2012, 282574. doi:10.1155/2012/282574
- Blattner, F. R., Plunkett, G. III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462. doi:10.1126/science.277.5331.1453
- DeMartini, J. D., Pattathil, S., Miller, J. S., Li, H. J., Hahn, M. G., and Wyman, C. E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy Environ. Sci.* 6, 898–909. doi:10.1039/c3ee23801f
- Emmett, M. R., and Caprioli, R. M. (1994). Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. *J. Am. Soc. Mass Spectrom.* 5, 605–613. doi:10.1016/1044-0305(94)85001-1
- Eudes, A., George, A., Mukerjee, P., Kim, J. S., Pollet, B., Benke, P. I., et al. (2012). Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.* 10, 609–620. doi:10.1111/j.1467-7652.2012.00692.x
- Eudes, A., Liang, Y., Mitra, P., and Loque, D. (2014). Lignin bioengineering. *Curr. Opin. Biotechnol.* 26, 189–198. doi:10.1016/j.copbio.2014.01.002
- Eudes, A., Sathitsuksanoh, N., Baidoo, E. E. K., George, A., Liang, Y., Yang, F., et al. (2015). Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol. J.* doi:10.1111/pbi.12310
- Gapeev, A., Berton, A., and Fabris, D. (2009). Current-controlled nanospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* 20, 1334–1341. doi:10.1016/j.jasms.2009.03.007
- Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J., and Yates, J. R. III (1998). Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal. Biochem.* 263, 93–101. doi:10.1006/abio.1998.2809
- Ito, J., Bath, T. S., Petzold, C. J., Redding-Johanson, A. M., Mukhopadhyay, A., Verboom, R., et al. (2011). Analysis of the *Arabidopsis* cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J. Proteome Res.* 10, 1571–1582. doi:10.1021/pr1009433
- Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* 3, 64–76. doi:10.1021/cb7002434
- Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392. doi:10.1021/ac025747h
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi:10.1093/nar/gkr1090
- Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., and Keasling, J. D. (2008). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* 19, 556–563. doi:10.1016/j.copbio.2008.10.014
- Lundgren, D. H., Hwang, S. I., Wu, L. F., and Han, D. K. (2010). Role of spectral counting in quantitative proteomics. *Expert. Rev. Proteomics* 7, 39–53. doi:10.1586/epr.09.69
- Magrane, M., and Consortium, U. (2011). UniProt knowledgebase: a hub of integrated protein data. *Database(Oxford)* 2011, bar009. doi:10.1093/database/bar009
- Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658. doi:10.1021/ac0341261
- Parsons, H. T., Christiansen, K., Knierim, B., Carroll, A., Ito, J., Bath, T. S., et al. (2012). Isolation and proteomic characterization of the *Arabidopsis* Golgi defines functional and novel targets involved in plant cell wall biosynthesis. *Plant Physiol.* 159, 12–26. doi:10.1104/pp.111.193151
- Parsons, H. T., González Fernández-Niño, S. M., and Heazlewood, J. L. (2013). “Separation of the plant Golgi apparatus and endoplasmic reticulum by free-flow electrophoresis,” in *Plant Proteomics: Methods and Protocols, Second Edition*, eds J. V. Jorrin Novo, S. Komatsu, W. Weckwerth, and S. Weinkoop (New York: Humana Press), 527–539.
- Percy, A. J., Chambers, A. G., Yang, J., Domanski, D., and Borchers, C. H. (2012). Comparison of standard- and nano-flow liquid chromatography platforms for

- MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 404, 1089–1101. doi:10.1007/s00216-012-6010-y
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
- Persil-Cetinkol, Ö, Smith-Moritz, A. M., Cheng, G., Lao, J., George, A., Hong, K., et al. (2012). Structural and chemical characterization of hardwood from tree species with applications as bioenergy feedstocks. *PLoS One* 7:e52820. doi:10.1371/journal.pone.0052820
- Shen, H., Poovaiah, C. R., Ziebell, A., Tschaplinski, T. J., Pattathil, S., Gjersing, E., et al. (2013). Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol. Biofuels* 6, 71. doi:10.1186/1754-6834-6-71
- Swartz, M. E. (2005). UPLCTM: an introduction and review. *J. Liq. Chromatogr. Relat. Technol.* 28, 1253–1263. doi:10.1081/JLC-200053046
- Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007). Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol.* 25, 132–137. doi:10.1016/j.tibtech.2007.01.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 February 2015; accepted: 18 March 2015; published online: 01 April 2015.

Citation: González Fernández-Niño SM, Smith-Moritz AM, Chan LJG, Adams PD, Heazlewood JL and Petzold CJ (2015) Standard flow liquid chromatography for shotgun proteomics in bioenergy research. *Front. Bioeng. Biotechnol.* 3:44. doi: 10.3389/fbioe.2015.00044

This article was submitted to *Bioenergy and Biofuels*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 González Fernández-Niño, Smith-Moritz, Chan, Adams, Heazlewood and Petzold. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Development of a high throughput platform for screening glycoside hydrolases based on Oxime-NIMS

OPEN ACCESS

Edited by:

Bo Hu,
University of Minnesota, USA

Reviewed by:

Chiranjeevi Thulluri,
Jawaharlal Nehru Technological
University Hyderabad, India
Jiwei Zhang,
University of Minnesota, USA

*Correspondence:

Kai Deng
kdeng@sandia.gov;
Trent R. Northen
trnorthen@lbl.gov

[†]Present address:

Noppadon Sathitsuksanoh,
University of Louisville, Louisville, KY,
USA

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Bioengineering
and Biotechnology

Received: 10 August 2015

Accepted: 21 September 2015

Published: 13 October 2015

Citation:

Deng K, Guenther JM, Gao J,
Bowen BP, Tran H, Reyes-Ortiz V,
Cheng X, Sathitsuksanoh N, Heins R,
Takasuka TE, Bergeman LF,
Geertz-Hansen H, Deutsch S,
Loqué D, Sale KL, Simmons BA,
Adams PD, Singh AK, Fox BG and
Northen TR (2015) Development of a
high throughput platform for
screening glycoside hydrolases
based on Oxime-NIMS.
Front. Bioeng. Biotechnol. 3:153.
doi: 10.3389/fbioe.2015.00153

Kai Deng^{1,2*}, Joel M. Guenther^{1,2}, Jian Gao³, Benjamin P. Bowen³, Huu Tran^{1,2},
Vimalier Reyes-Ortiz^{1,3}, Xiaoliang Cheng^{1,3}, Noppadon Sathitsuksanoh^{1,3†},
Richard Heins^{1,2}, Taichi E. Takasuka⁴, Lai F. Bergeman⁴, Henrik Geertz-Hansen¹,
Samuel Deutsch^{3,5}, Dominique Loqué^{1,3}, Kenneth L. Sale^{1,2}, Blake A. Simmons^{1,2},
Paul D. Adams^{1,3,6}, Anup K. Singh^{1,2}, Brian G. Fox^{4,7} and Trent R. Northen^{1,3*}

¹ US Department of Energy Joint BioEnergy Institute, Emeryville, CA, USA, ² Sandia National Laboratories, Livermore, CA, USA, ³ Lawrence Berkeley National Laboratory, Berkeley, CA, USA, ⁴ US Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI, USA, ⁵ Joint Genome Institute, Walnut Creek, CA, USA, ⁶ Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA, ⁷ Department of Biochemistry, University of Wisconsin, Madison, WI, USA

Cost-effective hydrolysis of biomass into sugars for biofuel production requires high-performance low-cost glycoside hydrolase (GH) cocktails that are active under demanding process conditions. Improving the performance of GH cocktails depends on knowledge of many critical parameters, including individual enzyme stabilities, optimal reaction conditions, kinetics, and specificity of reaction. With this information, rate- and/or yield-limiting reactions can be potentially improved through substitution, synergistic complementation, or protein engineering. Given the wide range of substrates and methods used for GH characterization, it is difficult to compare results across a myriad of approaches to identify high performance and synergistic combinations of enzymes. Here, we describe a platform for systematic screening of GH activities using automatic biomass handling, bioconjugate chemistry, robotic liquid handling, and nanostructure-initiator mass spectrometry (NIMS). Twelve well-characterized substrates spanning the types of glycosidic linkages found in plant cell walls are included in the experimental workflow. To test the application of this platform and substrate panel, we studied the reactivity of three engineered cellulases and their synergy of combination across a range of reaction conditions and enzyme concentrations. We anticipate that large-scale screening using the standardized platform and substrates will generate critical datasets to enable direct comparison of enzyme activities for cocktail design.

Keywords: cellulase, NIMS, oxime bioconjugation, high throughput screening, enzyme assays

Abbreviations: AFEX-SG, ammonia fiber expansion pretreated switchgrass; CBM, carbohydrate binding module; CelE, broad specificity GH family 5 (GH5) domain from *C. thermocellum* Cthe_0797; DA-SG, diluted acid pretreated switchgrass; GH, glycoside hydrolase; IL-SG, ionic liquid pretreated switchgrass; NIMS, nanostructure-initiator mass spectrometry UT-SG, untreated switchgrass.

INTRODUCTION

Lignocellulosic biomass (Carroll and Somerville, 2009) is a renewable source of energy, capable of providing the nation with clean, renewable transportation fuels. To convert biomass into biofuels, one key step is enzymatic saccharification, which is known to be inefficient and expensive (Klein-Marcuschamer et al., 2012). Thus, low-cost, robust, high-performance enzymes or enzyme cocktails are needed to reduce the overall cost of biofuel production. There are many sources of inedible plant biomass that can serve as feedstocks for biofuel production, including agricultural wastes [corn stover, switchgrass (SG), wood trimming], municipal solid wastes, and emerging bioenergy crops. However, these various feedstocks have different glycan composition, bond linkages, and individual sugar contents, which complicates the development of cost-effective saccharification approaches. Moreover, the content and structure of glycans and lignin from the same biomass may respond in different ways to the prerequisite pretreatments (Li et al., 2010). These variations contribute to the observation that there is no universal enzyme or enzyme cocktail for all substrates and biofuels processes.

Given these process constraints, it is important to characterize saccharification enzymes against a wide range of biomass compositions, pretreatments, and processing conditions to generate data that can help enable customized optimal enzyme cocktails for saccharification (Banerjee et al., 2010; Walton et al., 2011). Given the large number of potential combinations of substrate-pretreatment reaction conditions, it is desirable to have a reliable high throughput enzyme assay methods and standardized panels of substrates and conditions.

Currently, several high throughput glycosyl hydrolyze (GH) assays are available to characterize enzyme activity. The majority of these are based on detection of colorimetric or fluorescent products. For example, the 2,4-dinitrosalicylic acid (DNS) reducing sugar assay (Decker et al., 2009) can provide rapid analysis of large enzyme libraries. However, it is a non-specific method that can only provide total reducing sugar content. Fluorescence-based enzyme assays using surrogate substrates, e.g., 4-methylumbelliferyl- β -glucopyranoside (van Tilbeurgh et al., 1982) are also available for the evaluation of β glycosidases, are also available. Surrogate substrate methods require preparation of each substrate for each enzyme type, which may be laborious, and the reactivity may be biased versus the native substrate. Moreover, care is required to avoid interference from background absorbance/fluorescence.

To address some of these limitations, we developed a mass spectrometry-based enzyme assay platform called Nimzyme (Northern et al., 2008; Reindl et al., 2011; Greving et al., 2012). The first-generation Nimzyme platform was based on soluble model substrates, which were synthesized chemically (Deng et al., 2012). This approach provided the specificity, sensitivity, and high throughput needed to screen a GH1 library of 175 enzymes. Several high-performance enzymes were identified with desired bioprocessing conditions (70°C, 20% ionic liquid) (Heins et al., 2014). Another manuscript in this volume reports use of the first-generation Nimzyme platform for numerical analysis of reactions with cellotetraose-nanostructure-initiator mass spectrometry

(NIMS) (Deng et al., 2014). Results of this work demonstrate diagnostic behaviors of several classes of GH enzymes.

However, the use surrogate substrates does not allow interrogation of the myriad of bonding types present in plant biomass or the three-dimensional arrangements of these bonds present on the plant biomass, and so does not represent a fully realistic approach for the study of enzyme reactions.

To overcome this limitation, we developed an oxime-Nimzyme probe (Deng et al., 2014) to directly study enzyme hydrolysis of plant biomass. In this approach, soluble oligosaccharide products are captured in a stable oxime linkage and then delivered to the NIMS chip for subsequent analysis, while inclusion of ^{13}C -labeled monosaccharide standards (glucose and xylose) allows quantitation of the derivatized glycans. Besides the diagnostic detection of solubilized products, this next-generation Nimzyme approach also allows quantitative studies of the time-dependence of product formation, and dissection of individual apparent rates for reactions of individual enzymes with plant biomass.

To further advance application of the Nimzyme approach, here we report the development of a process platform and a panel of 12 diverse glycan substrates. These substrates were selected to represent the diversity of plant glycosidic bond linkages and the sugar compositions that are relevant to biofuel production. We use this panel of substrates to characterize three previously described enzymes from *Clostridium thermocellum* that have been engineered to perform outside of their natural cellulosomal location. These are fusions of the catalytic domains of CelA (gene locus Cthe_0269, GH8 endoglucanase family), CelR (gene locus Cthe_0578, GH9 cellotetrahydrolase), and CelE (gene locus Cthe_0797, GH5 endoglucanase family) to the CBM3a domain from scaffoldin protein CipA. The CBM3a domain used in this work comes from the CipA scaffoldin protein from *C. thermocellum*. CBM3a is a well-studied carbohydrate-binding module helps to promote binding of the enzyme onto the polysaccharide, thus plays a key role in promoting the efficient hydrolysis of cellulose (Yaniv et al., 2012). We envision that the standardization and automation of GH assays enabled by this approach will be valuable in providing large datasets of GH performance needed to select enzymes for improved biomass deconstruction.

MATERIALS AND METHODS

Materials

1,4- β -D-cellotetraose ~95% was purchased from Megazyme (Ireland) Cat. No. O-CTE100, Lot No. 130604; 1,4- β -D-xylotetraose ~95% was purchased from Megazyme (Ireland), Cat. No. O-XTE, Lot No. 120204; 1,4- β -D-mannotetraose ~95% was purchased from Megazyme (Ireland), Cat. No. O-MTE, Lot No. 111004; Arabinoxylan was purchased from Megazyme (Ireland), Cat. No. P-WAXYI, Lot No. 120801a; Carob galactomannan was purchased from Megazyme (Ireland) Cat. No. P-GALML, Lot No. 10501b; beechwood xylan was purchased from Sigma-Aldrich (St. Louis, MO, USA), Cat. No. X4252-100G, Lot No. BCBL2915V.

Switchgrass has emerged as a potential bioenergy crop because it is perennial, resource-efficient, and requires

low-inputs for maintenance (Keshwani and Cheng, 2009). Untreated switchgrass (UT-SG), also named as Putnam SG, was obtained from Daniel Putnam at UC Davis; diluted acid pretreated switchgrass (DA-SG) was obtained by mixing Putnam SG with 1% sulfuric acid at 190°C for 0.5 min at NREL. Ammonium fiber expansion pretreated switchgrass (AFEX-SG) was prepared at Michigan State University. Ionic liquid pretreated switchgrass (IL-SG) was prepared at 140°C for 3 h with 15% solid loading in [EMIM][OAc] (1-ethyl-3-methylimidazolium acetate) at JBEI. All four SG substrates (IL-SG, AFEX-SG, DA-SG, and UT-SG) were milled by Thomas Wiley Mill (Model 3383 L1) for 20 min and sieved (passage through a 20-mesh sieve and retention by an 80-mesh sieve). The particle size was found in the range of 200–450 μm. Avicel PH-101 cellulose was requested from FMC Biopolymer (Philadelphia, PA, USA), Lot No. P112824596. Phosphoric acid swollen cellulose (PASC) was prepared from Avicel PH-101 cellulose (FMC Biopolymer) with 85% phosphoric acid as reported (Zhang et al., 2006).

Synthesis

The synthesis of *O*-alkyloxyamine fluorous tag has been reported previously (Deng et al., 2014).

Enzymes

The catalytic domains of the enzymes studied were obtained from the following gene loci: CelA (Cthe_0269); CelR (Cthe_0578); and CelE (Cthe_0797). Each of these catalytic domains was fused to the CBM3a domain from the scaffoldin CipA (Cthe_3077). Additional information on these genes can be found at Uniprot (Apweiler et al., 2011). All genes were prepared by PCR using *C. thermocellum* ATCC 27405 genomic DNA as template and cloned into the *Escherichia coli* expression vector pEC_CBM3a to create enzyme_CBM3a fusion proteins, e.g., CelAcc_CBM3a. The vector pEC_CBM3a is a hybrid of pEU_HSBC_CBM3a and pVP65K (Takasuka et al., 2014; Aceti et al., 2015) that yields fusion proteins having an N-terminal enzyme catalytic domain fused by an ~40 aa linker sequence to the CBM3a domain from Cthe_3077. Methods for PCR amplification, capture, and sequence verification of protein coding sequences, transformation into *E. coli* 10G competent cells (Lucigen, Middleton, WI, USA) for DNA manipulations and *E. coli* B834 for protein expression were as previously reported (Takasuka et al., 2014). Additional details of the properties and methods for use of pEU is described elsewhere (Aceti et al., 2015).

Enzyme Plate Construction

Three enzymes (CelAcc-CBM3a, CelRcc-CBM3a, and CelEcc-CBM3a) were chosen to construct an enzyme plate with varied enzyme concentrations (microgram per microliter) shown in Table 1. Each enzyme was present in triplicate as shown for CelAcc-CBM3a (columns A, B, C), CelRcc-CBM3a (columns D, E, F), and CelEcc-CBM3a (columns G, H, I). Some enzyme combinations were also included to investigate the potential synergy among these three enzymes as shown in columns J, K, L.

TABLE 1 | Enzyme plate construction.

Unit	CelAcc-CBM3a			CelRcc-CBM3a			CelEcc-CBM3a			CelAcc-CBM3a + CelRcc-CBM3a		CelAcc-CBM3a + CelEcc-CBM3a	
	A	B	C	D	E	F	G	H	I	J	K	L	
μg/μL	1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.005 + 0.005	0.005 + 0.005	0.005 + 0.005	
μg/μL	2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.025 + 0.025	0.025 + 0.025	0.025 + 0.025	
μg/μL	3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.05 + 0.05	0.05 + 0.05	0.05 + 0.05	
μg/μL	4	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.125 + 0.125	0.125 + 0.125	0.125 + 0.125	
μg/μL	5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.25 + 0.25	0.25 + 0.25	0.25 + 0.25	
μg/μL	6	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25	0.625 + 0.625	0.625 + 0.625	0.625 + 0.625	
μg/μL	7	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	1.25 + 1.25	1.25 + 1.25	1.25 + 1.25	
μg/μL	8	5	5	5	5	5	5	5	5	2.5 + 2.5	2.5 + 2.5	2.5 + 2.5	

Procedures for Handling Insoluble Substrates with Labman Solids-Handling Robot

To minimize static electricity, all plastic labware was prophylactically treated using a Tabletop Ionizing Transport System Model IT-7000 (Electrostatics Incorporated, 90610-03610, Harleysville, PA, USA). Commercially sourced substrates – Avicel PH-101, arabinoxylan, carob galactomannan, and beechwood xylan – were received as fine powders and aliquoted without further processing. The Labman (North Yorkshire, UK) solids-handling robot at JBEI was used to aliquot insoluble substrates (Table 2, numbers 4 through 12) from 2-mL Sarstedt vials (72.694.007) into 340- μ L 96-well thermal cycler plates (plate: VWR 82006-636, holder: Axygen R-96-PCR-FY) with a target mass of 2 ± 0.25 mg per well. Feeder vibration power was set to 40% for all substrates and feeding durations were optimized for each to minimize overfeeding. Average feeding durations ranged from 3 to 5 h per 96-well plate depending on the substrate. After aliquoting, substrate plates were sealed using peelable heat seal (Agilent 24210-001) using an Agilent PlateLoc heat sealer set to 175°C for 2.5 s and stored at 4°C.

Procedures for Handling Insoluble Substrates with Biomek FX Robot

For all nine insoluble solid substrates, all the liquid handling for enzymatic hydrolysis reactions and bioconjugation chemistry were performed using a Biomek FX robot equipped with an AP96 multichannel pod (Beckman Coulter). For the enzymatic hydrolysis step, the Biomek FX transferred 180 μ L of 50 mM phosphate buffer, pH 6.0, into 96-well PCR plates containing ~2 mg solid substrate that was previously aliquoted using the Labman robot (plate A). Then 20 μ L of enzyme solution was transferred from the enzyme plate into plate A. After sealing the 96-well plate with PlateLoc peelable seal (2.5 s, 175°C), the plate was incubated at 60°C for 18 h in a shaker with shaker speed set to 200 rpm (HT INFORS). The recipe for a typical NIMS bioconjugation solution was a mixture of the following: (1) 1 mL of probe solution (150 mM in 1:1 (v/v) H₂O:MeOH); (2) 6 mL of 100 mM glycine buffer, pH 1.3; 3) 0.5 mL of 5 mM

¹³C glucose aqueous solution; (4) 0.5 mL of 5 mM ¹³C xylose aqueous solution; (5) 2 mL of acetonitrile; (6) 1 mL of methanol. Plate B is a 96-well PCR plate prepared by the Biomek robot containing with 22.2 μ L of NIMS tagging solution in each well. After plate A (enzymatic reaction) was cooled to room temperature, a 4 μ L aliquot from plate A was transferred into plate B by the Biomek robot. Plate B was sealed with PlateLoc peelable heat seal (2.5 s, 175°C) and left at room temperature for 16 h. Samples (12 μ L) from plate B was transferred into the assay plate (Greiner bio-one, 384 well μ clear-plate, coc black, Lo base, 10 pcs/bag, Lot No. E11060DN; Cat. No. 788876, Made in Germany).

Procedures for Handling Soluble Substrates with Biomek FX Robot

For the three soluble substrates, 1,4- β -D-cellobiose (G4), 1,4- β -D-xylotetraose (X4), and 1,4- β -D-mannotetraose (M4), 10 mM aqueous solutions were prepared as stock solutions. The Biomek robot was used to transfer 40 μ L of 50 mM phosphate, pH 6.0, into each well of a 96-well PCR plate (plate A). Then 5 μ L of soluble substrates [G4 (5 mM), X4 (5 mM), or M4 (5 mM)] was transferred to plate A. After that, Biomek transferred 5 μ L of enzymes from the enzyme plate to plate A. After sealing the 96-well plate with PlateLoc peelable heat seal (2.5 s, 175°C), the whole plate was incubated at 60°C for 18 h with 200 rpm in a shaker (HT INFORS). After this step, the Biomek robot was used to perform all subsequent liquid-handling steps as indicated above for insoluble substrates.

Acoustic Printing of Sample Arrays

Samples from the 384-well Greiner plate (1 μ L) were acoustically transferred by ATS Acoustic Liquid Dispenser (EDC Biosystems) onto a 2 \times 2 inch NIMS chip. Individual reaction spots on the NIMS chip were ionized by a laser and products were detected by a time-of-flight mass spectrometer (TOF/TOF 5800 MALDI systems, AB Sciex).

Nanostructure-Initiator Mass Spectrometry Imaging

The 4800 imaging acquisition software was used in these experiments. Chips were loaded using a modified standard MALDI plate. Instrument was set with laser intensity at 2,550 and 15 shots per sub-spectrum. The detector voltage multiplier was set as 0.77. For imaging, the step size was set up at 50 μ m for both x and y direction.

Mass Spectrometry Imaging Data Processing (Open MSI)

The imaging file was uploaded to openmsi.nersc.gov by Globus Connect Personal. The converted file was analyzed by a draggable points notebook written in Python. Signal intensities were identified for the ions of the tagging products. Enzyme activities were determined by measuring the concentration of glycan products using either [U]-¹³C glucose or [U]-¹³C xylose as an internal standard.

TABLE 2 | Standardized substrates list.

Soluble substrates	Insoluble solid substrates
(1) Cellotetraose	(4) Acid swollen cellulose (PASC)
	(5) Avicel PH-101 cellulose
	(6) Arabinoxylan
(2) Xylotetraose	(7) Carob galactomannan
	(8) Beechwood xylan
	(9) IL-switchgrass
(3) Mannotetraose	(10) AFEX-switchgrass
	(11) DA-switchgrass
	(12) UT-switchgrass

(1) Soluble substrates 1–3 can be dissolved in water and bypass the solid dispersion step by Labman. (2) Insoluble solid substrates 4–12 are being dispersed into a 96-well PCR plate by Labman, then Biomek was used for liquid handling.

RESULTS AND DISCUSSION

Substrate Panel

Table 2 lists some of the properties of the substrates included in the substrate panel. More detailed rationale for selection of individual substrates is described in the following experimental sections. The following general principles were used to assemble the substrate panel: (1) substrates should be readily available; (2) some substrates should have known structures (e.g., cellotetraose, xylotetraose, mannotetraose) so that enzyme specificity can be rapidly determined; (3) some substrates should be plant biomass substrates such as SG; and (4) examples of different pretreatments of the same biomass should be included. Based on evolving needs, other substrates can be added to the panel using similar principles.

After selection of these 12 substrates, we sourced large quantities of each substrate so as to permit detailed analytical characterizations and extended experimentation. Each of the substrates was acquired and then divided such that our two institutions (GLBRC and JBEI) each had large aliquots of each substrate.

Enzymes Selected for Platform Testing

To test this assay platform, we used catalytic domains from three *C. thermocellum* enzymes. In earlier studies (Deng et al., 2014; Takasuka et al., 2014), we showed that CelAcc, CelRcc, and CelEcc were able to release a variety of oligosaccharide products from IL- and AFEX-pretreated SG. As these enzymes are normally included in the cellulosome, we removed the dockerin domains and instead fused them to CBM3a (Yaniv et al., 2012). This engineered addition

of CBM3a targets the catalytic domain to a polysaccharide surface. Our studies also showed that CelEcc was a multifunctional GH5 catalytic domain that was reacted with cellulose, xylan, and mannan (classified as CMX). This broad specificity of reaction provides opportunities to understand contributions of an identical active site to biomass hydrolysis. By contrast, CelRcc_CBM3a reacted only with cellulose (classified as C), while CelAcc_CBM3a reacted with cellulose and only weakly with xylan (classified as CX), and so provided more specific enzyme reactions.

Automated Enzyme Assay Platform

Efforts have been made to establish a simple automated workflow that enables high-throughput and that minimizes assay error. In a preparatory step (**Figure 1**), insoluble substrates (**Table 2**, 4 through 12) were aliquoted into 96-well thermal cycler plates for activity assays using a Labman solids-handling robot and stored at 4°C until use. Immediately before initiating the activity assays, enzymes were manually aliquoted into 96-well plates.

The individual steps of the workflow are as follows. In step 1, reactions were prepared using a Biomek FX liquid-handling robot to add enzyme to substrate. In step 2, reactions were incubated at 60°C for 18 h with shaking. In step 3, an aliquot from the reaction plate was combined with the oxime-NIMS tagging solution using the liquid-handling robot, after which the NIMS tag reacted with reducing sugar to form a stable oxime linkage during overnight incubation. In step 4, the solution containing NIMS-tagged glycans was transferred from tagging reactions onto a NIMS chip mass spectrometry surface by sequentially

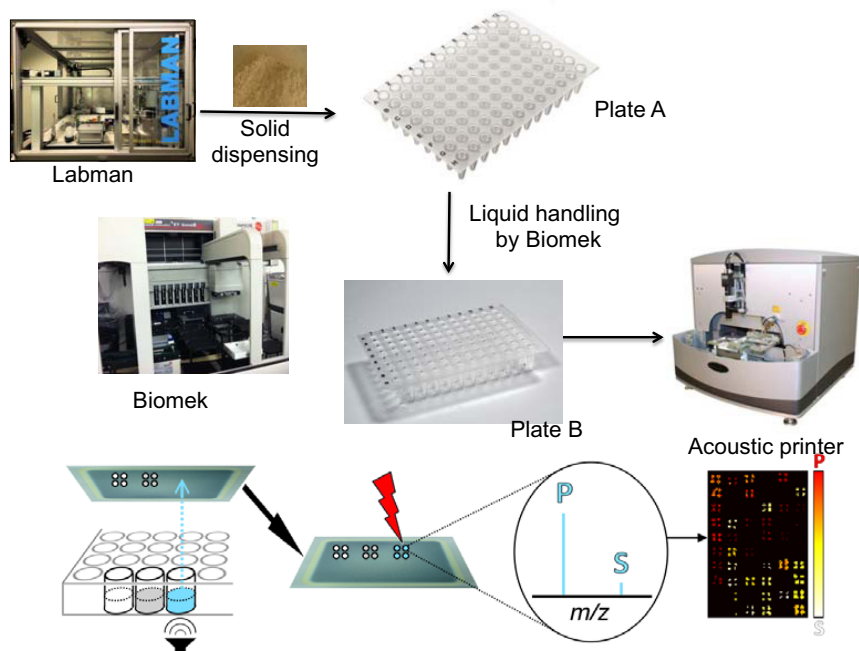


FIGURE 1 | Workflow of oxime-NIMS automation developed to study GHs. (1) Solid biomass was dispersed by Labman. (2) Liquid handling was performed by Biomek Automation Workstation, including setup of enzyme and oxime bioconjugation reactions. (3) Sample array on the NIMS chip was generated by an acoustic printer (ATS Acoustic Liquid Dispenser). (4) Mass spectrometry imaging (MSI) provided the readout of experimental assay results.

using (1) the liquid-handling robot to reformat aliquots from the reactions in 96-well plates into 384-well plates followed by (2) using an ATS Acoustic Liquid Dispenser to print 1-nL droplets onto the chip via non-contact dispensing. In step 5, an AB SCIEX TOF/TOF 5800 mass spectrometer was used to image the NIMS chip to collect data on the tagged glycans. In step 6, the data were processed using OpenMSI (a free web-based visualization, analysis and management package for mass spectrometry imaging (MSI) data) to quantify the tagged glycans in order to compute enzyme activities.

The different characteristics of the substrates introduced challenges for automated sample handling. Thus, while soluble samples can be dispensed using standard liquid handling (**Figure 1**), insoluble samples were more challenging and required application of solid dispensing using a Labman system. As shown in **Table 2**, only 3 out of the 12 substrates were soluble and could be processed using simple liquid handling, whereas the nine other substrates were insoluble and required Labman to carry out relatively slow protocol to generate batches of substrate-filled 96-wells PCR plates that were then stored at -20°C freezer until required.

Reactions with Soluble Substrates

Cellotetraose (G4), xylo-tetraose (X4), and mannotetraose (M4) are purified oligosaccharides that can be used to study cellulase, xylanase, and mannanase activities, respectively. These soluble substrates also permit simple liquid-handling approaches. However, it is important to recognize that some GH enzymes require longer oligosaccharide chains to show effective catalysis.

The reactions of G4, X4, and M4 were screened using a 96-well enzyme plate containing individual wells of CelAcc_CBM3a, CelRcc_CBM3a, and CelEcc_CBM3a, and mixtures of these three (**Table 1**). All three endoglucanases produced cellobiose as the major product from G4 under various concentrations (Tables S1–S3 in Supplementary Material). No products were observed at the lowest enzyme concentration tested (1 ng/ μL) for all three enzymes. When the enzyme concentration was increased (5 ng/ μL), CelEcc_CBM3a gave complete hydrolysis of the cellotetraose present into smaller oligosaccharides, while CelRcc_CBM3a hydrolyzed about 80% of G4 and CelAcc_CBM3a hydrolyzed about 13% of G4. For CelAcc_CBM3a, increasing the enzyme concentration above 0.01 $\mu\text{g}/\mu\text{L}$ was needed to obtain complete hydrolysis. Thus, the concentration screening can give a preliminary assessment of the affinity of an enzyme for the oligosaccharide. In the hydrolysis of cellotetraose, the following effective concentrations of enzyme were determined: CelAcc_CBM3a (10 ng/ μL), CelRcc_CBM3a (10 ng/ μL), and CelEcc_CBM3a (5 ng/ μL).

Xylo-tetraose (X4) is an oligosaccharide with β 1,4-linked xylose unit, and cellulases CelAcc_CBM3a and CelRcc_CBM3a were not reactive with xylo-tetraose. However, the multifunctional enzyme CelEcc_CBM3a was active with xylo-tetraose. Increasing the enzyme concentration increased oligomer and monomer products (cellotriose, cellobiose, and xylose, Table S4 in Supplementary Material). For CelEcc_CBM3a with xylo-tetraose, the minimum amount of enzyme needed for complete hydrolysis of X4 was 0.25 $\mu\text{g}/\mu\text{L}$. This is a 50-fold increase in the amount of enzyme needed to hydrolyze xylo-tetraose versus cellotetraose, suggesting higher affinity for the hexose oligosaccharide.

None of the enzymes studied reacted with mannotetraose (M4). Since CelAcc_CBM3a and CelRcc_CBM3a are cellulases, this was not surprising. However, since CelEcc_CBM3a has been shown to react with mannan and glucomannan, the lack of reaction with M4 was not expected. This may arise from the possibility that mannotetraose is not long enough to bind effectively in the catalytic channel of CelEcc_CBM3a.

Reactions with Avicel and PASC

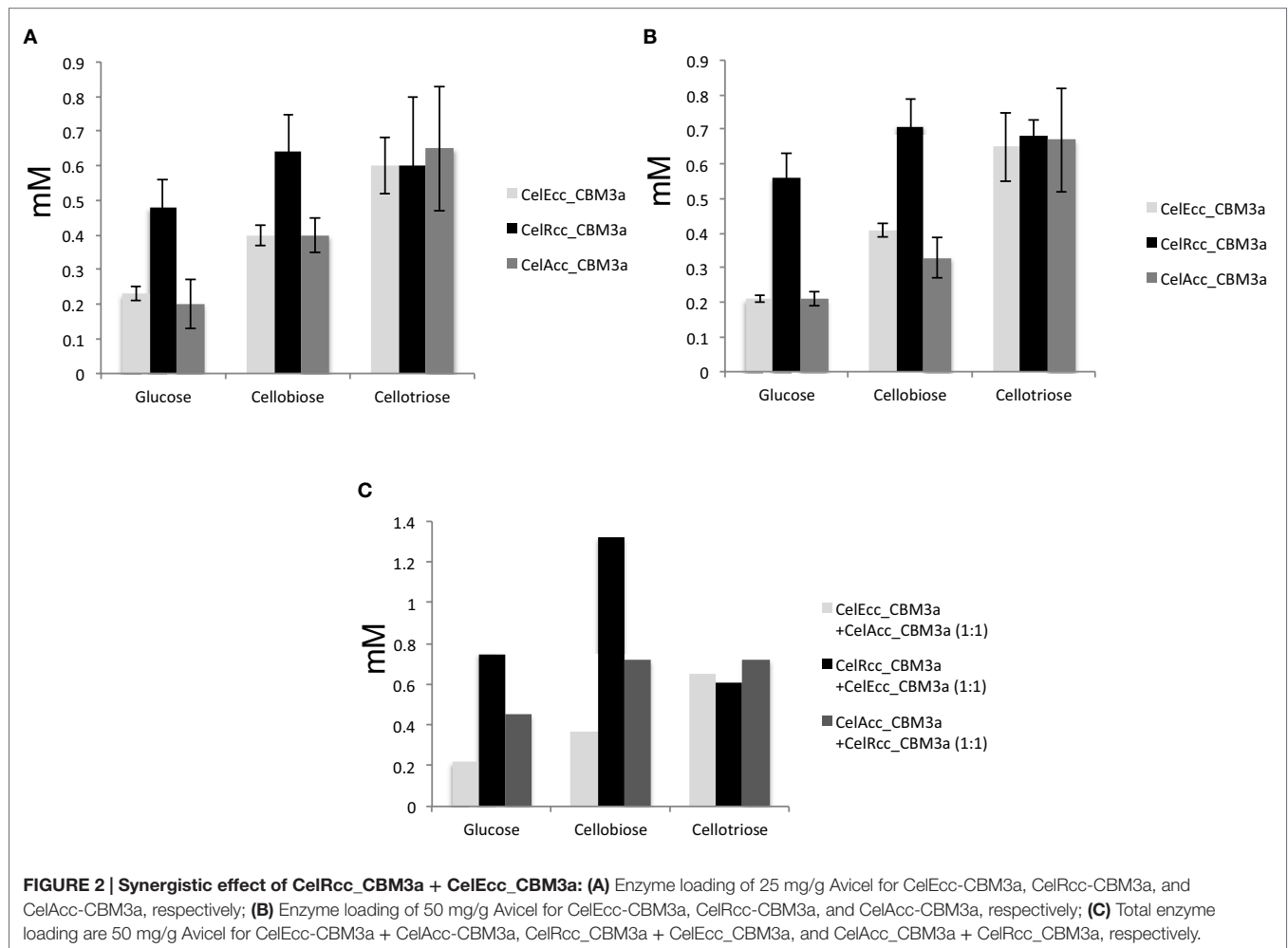
Phosphoric acid swollen cellulose (Zhang et al., 2006) has become a very popular substrate to study cellulase activities because it has a relatively easily hydrolyzed amorphous habit (Sharrock, 1988; Wood, 1988; Wood and Bhat, 1988). Solid-state cross-polarization magic angle spinning ^{13}C NMR was used to determine the crystallinity (Park et al., 2010) of both Avicel and PASC used in this work. The NMR results demonstrated that Avicel had crystallinity of 53% while PASC had crystallinity of less than 5% (Figure S1 in Supplementary Material). Compared with the microcrystalline cellulose (Avicel), amorphous PASC has the advantages of practical solubility and increased accessibility to the cellulases. Therefore, it is more easily hydrolyzed by cellulases. As expected, all three cellulase enzymes produced multiple times more soluble hexose products from PASC than Avicel (4 times for CelEcc_CBM3a, 2.5 times for CelRcc_CBM3a, and 3 times more for CelAcc_CBM3a) (Tables S5 and S6 in Supplementary Material). For the hydrolysis of Avicel, CelRcc_CBM3a performed better than CelEcc_CBM3a and CelAcc_CBM3a. Interestingly, the enzyme combination of CelEcc_CBM3a and CelRcc_CBM3a worked better than the combination of either CelAcc_CBM3a and CelRcc_CBM3a or CelEcc_CBM3a and CelAcc_CBM3a (**Figure 2**). Since the amount of glycan products produced by the combination of CelEcc_CBM3a and CelRcc_CBM3a was greater than the scaled contributions of the individual enzymes, this combination is demonstrated to have a synergistic effect in reaction.

Reactions with Beechwood Xylan and Arabinoxylan

Beechwood xylan has a high xylose content ($\sim 84\%$) with a majority of β -1,4 linkages. Arabinoxylan consists of a mixture of arabinose and xylose in an approximate 40:60 ratio with β -1,4- and β -1,6-linkages. Our screening results showed that CelRcc_CBM3a was unable to hydrolyze either beechwood xylan or arabinoxylan, even at high enzyme loading. By contrast, CelAcc_CBM3a showed weak activity with beechwood xylan, and at the highest enzyme loading tested (50 mg/g xylan), about 20% of the xylan was hydrolyzed (Table S8 in Supplementary Material). Furthermore, CelEcc_CBM3a reacted well with beechwood xylan, and at the highest enzyme concentration tested (50 mg/g xylan), about 40% of the xylan was hydrolyzed into soluble pentose products. Both CelAcc_CBM3a and CelEcc_CBM3a had only weak activities with arabinoxylan, even at the highest enzyme loading tested (50 mg/g xylan).

Reactions with Galactomannan

Galactomannans are polysaccharides consisting of a mannose backbone with galactose side groups. Mannans are important constituent of hemicellulose in some plant biomass. (Malherbe



et al., 2014) For example, softwoods contain 15–20% (w/w) mannans (Rodríguez-Gacio et al., 2012) and legume seeds can contain more than 30% of mannans (Buckeridge, 2010). For hydrolysis of these types of biomass into simple monosaccharides, it is important to find enzymes that can efficiently degrade mannans. Consequently, galactomannan was included in the substrate panel to test for mannanase activities. The screening results show that cellulases (CelAcc_CBM3a and CelRcc_CBM3a) could not deconstruct galactomannan at all. However, CelEcc_CBM3a (Table S9 in Supplementary Material) was able to hydrolyze galactomannan to hexose products (Fox et al., 2014).

It is interesting to compare the results of CelEcc_CBM3a reaction with galactomannan and its lack of reaction with mannotetraose. Besides the length of the oligosaccharide suggested above, it is also plausible that CelEcc_CBM3a may prefer reaction with the galactose-substituted mannans.

Reactions with Switchgrass

The structural features of SG, including surface area, crystallinity, the contents of cellulose, hemicellulose, and lignin vary considerably depending on pretreatment (Xu and Huang, 2014).

For example, cellulose is changed from cellulose I (untreated) to cellulose II after ionic liquid pretreatment (Cui et al., 2014). Other factors, like degree of delignification, hemicellulose solubilization (especially in diluted acid pretreatment), changes in porosity, and others affect enzyme accessibility so that different glycan products are produced for different pretreated SG. The compositional analysis of these four pretreated samples demonstrates these significant differences caused by pretreatment (Table S10 in Supplementary Material). These compositional differences imply the need for customized enzyme cocktails.

Four types of pretreated SG were included in the substrate panel to permit comparative studies of the consequences of pretreatment on enzymatic saccharification. These are UT-SG as control, IL-SG (Li et al., 2010), AFEX-SG (Bals et al., 2010), and DA-SG (Pu et al., 2013). This selection covers the three predominant pretreatment methods under investigation by the US-DOE funded Bioenergy Research Centers (Singh et al., 2015).

Screening results show that neither the three individual enzymes (CelAcc_CBM3a, CelRcc_CBM3a, and CelEcc_CBM3a) nor their combinations could deconstruct UT-SG. This is consistent with the substantial value of pretreatment before enzymatic saccharification.

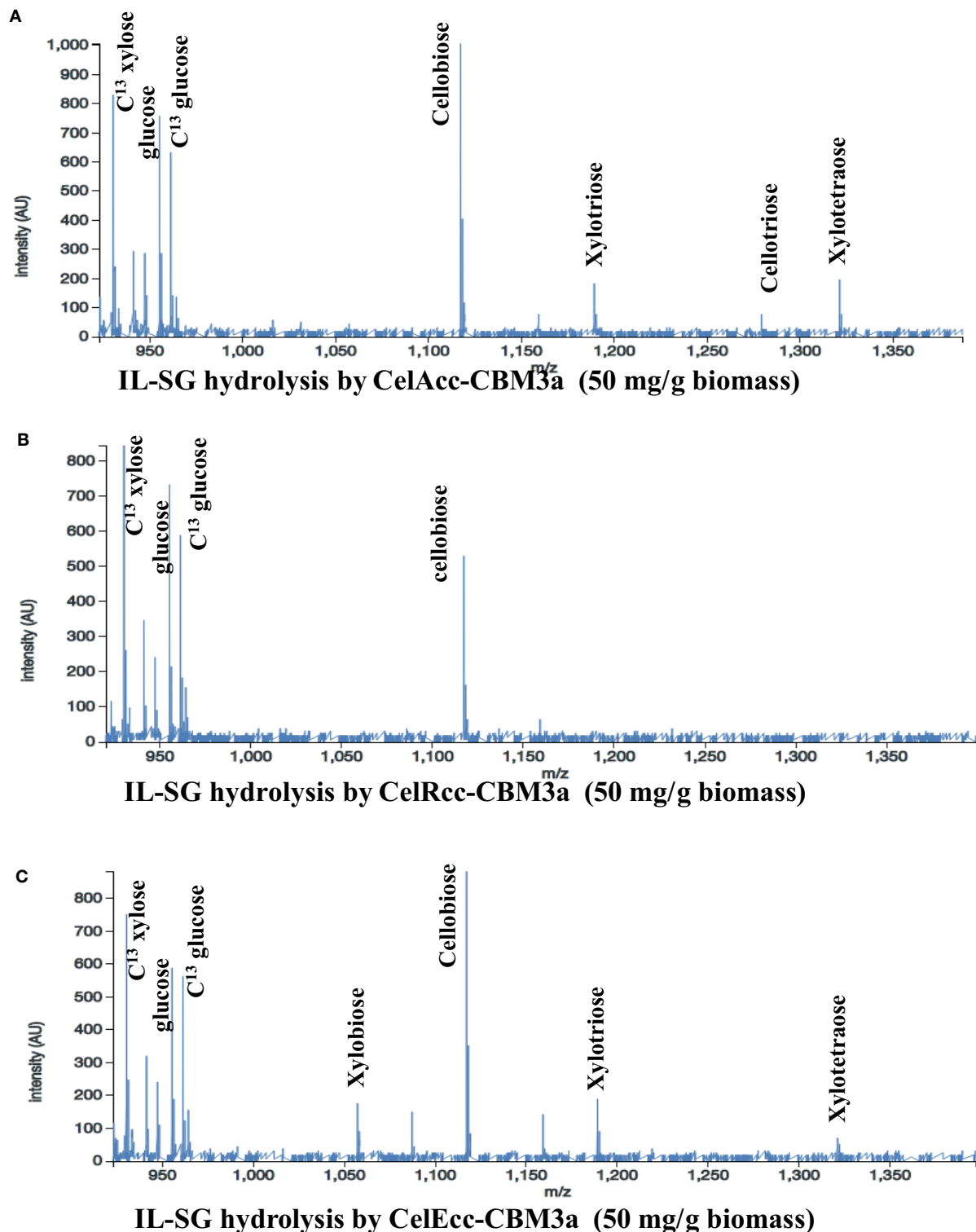


FIGURE 3 | Mass spectra from OpenMSI data processing: Reactions of IL-SG with CelAcc_CBM3a (A), CelRcc_CBM3a (B), and CelEcc_CBM3a (C).

For DA-SG, all assays across the breadth of enzyme concentrations gave similar, low yields for both hexose and pentose products. The relatively low reactivity of these individual

enzymes with crystalline cellulose is consistent with the low yield of hexose product. Presumably, the low amount of xylan remaining in DA-SG (4.58%) is not easily accessible to further enzyme

hydrolysis. For example, although CelEcc_CBM3a reacts well with pure xylan and the hemicellulose fractions in AFEX-SG and IL-SG (see below), it did not react with the remaining xylan in DA-SG. Whether this is a result of depletion of reactive substructures or other aspects of the diluted acid pretreatment is not clear.

For reactions with AFEX-SG, both CelRcc_CBM3a and CelAcc_CBM3a produced little soluble glycan and no soluble pentose. By contrast, CelEcc_CBM3a performed much better, and yielded both hexose and pentose products. CelEcc_CBM3a was especially reactive with the hemicellulose portion of AFEX-SG, yielding about 2.5× more pentose products than in its reaction with IL-SG (Table S11 in Supplementary Material).

For IL-SG, all three enzymes (Figure 3) produce significant amount of hexose products (Table S12 in Supplementary Material), which can be attributed to the ability of the IL pretreatment to reduce the crystallinity of cellulose. CelAcc_CBM3a produced the most hexose products among these three enzymes, and under the highest enzyme loading of 50 mg/g biomass a conversion of 24% of the glycan was observed. A comparison of the reaction of CelEcc_CBM3a with either IL-SG or AFEX-SG under the same experimental conditions (enzyme loading of 50 mg/g biomass) showed that CelEcc_CBM3a produced 8× more of hexose products with IL-SG than AFEX-SG. This result from the automated platform matches the conclusion that cellulose in IL-SG is much easier to be accessed and digested by CelEcc_CBM3a obtained in earlier oxime-NIMS studies (Deng et al., 2014).

CONCLUSION

In this work, we described a panel of 12 substrates and automated platform for characterization of GH enzymes using the oxime-NIMS approach. Standardization is an important step toward more in-depth comparison of GH enzymes activities, both from natural environments and from engineered systems. These studies are consistent with our earlier assignment that CelEcc_CBM3a

is a multifunctional enzyme that has cellulase, mannanase, and hemicellulase activities. By contrast, CelAcc_CBM3a has cellulase and only weak hemicellulase while CelRcc_CBM3a only has cellulase activity. This platform automates the handling of both solid biomass and soluble substrates, the introduction of enzymes as individuals or combinations, and the recovery of products for high sensitivity and high-resolution mass spectral analysis. Simplex optimization of the ratios of natural or engineered enzyme combinations produced by systems biology approaches such as gene synthesis and robotic cell-free translation, as well as optimization of the pretreatment conditions can be readily undertaken using this platform.

AUTHOR CONTRIBUTIONS

KD, BF, and TN designed experiments. KD, JMG, JG, HT, VR-O, XC, NS, RH, TT, LB, HG-H, and SD carried out experimental work, KD, BB, BF, and TN analyzed results, and KD, JMG, BF, and TN prepared the manuscript. DL, KS, BS, and PA supervised the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The DOE Joint BioEnergy Institute and DOE Great Lakes Bioenergy Research Center are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 and through contract DE-FC02-07ER64494, respectively. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00153>

REFERENCES

- Aceti, D. J., Bingman, C. A., Wrobel, R. L., Frederick, R. O., Makino, S., Nichols, K. W., et al. (2015). Expression platforms for producing eukaryotic proteins: a comparison of E. coli cell-based and wheat germ cell-free synthesis, affinity and solubility tags, and cloning strategies. *J. Struct. Funct. Genomics* 16, 67–80. doi:10.1007/s10969-015-9198-1
- Apweiler, R., Martin, M. J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Antunes, R., et al. (2011). Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* 39, D214–D219. doi:10.1093/nar/gkq1020
- Bals, B., Rogers, C., Jin, M. J., Balan, V., and Dale, B. (2010). Evaluation of ammonia fibre expansion (AFEX) pretreatment for enzymatic hydrolysis of switchgrass harvested in different seasons and locations. *Biotechnol. Biofuels* 3, 1. doi:10.1186/1754-6834-3-1
- Banerjee, G., Car, S., Scott-Craig, J. S., Borrusch, M. S., and Walton, J. D. (2010). Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. *Biotechnol. Biofuels* 12, 22. doi:10.1186/1754-6834-3-22
- Buckeridge, M. S. (2010). Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. *Plant Physiol.* 154, 1017–1023. doi:10.1104/pp.110.158642
- Carroll, A., and Somerville, C. (2009). Cellulosic biofuels. *Annu. Rev. Plant Biol.* 60, 165–182. doi:10.1146/annurev.arplant.043008.092125
- Cui, T., Li, J., Yan, Z., Yu, M., and Li, S. (2014). The correlation between the enzymatic saccharification and the multidimensional structure of cellulose changed by different pretreatments. *Biotechnol. Biofuels* 7, 134. doi:10.1186/s13068-014-0134-6
- Decker, S. R., Brunecky, R., Tucker, M. P., Himmel, M. E., and Selig, M. J. (2009). High-throughput screening techniques for biomass conversion. *Bioenerg. Res.* 2, 179–192. doi:10.1007/s12155-009-9051-0
- Deng, K., George, K. W., Reindl, W., Keasling, J. D., Adams, P. D., Lee, T. S., et al. (2012). Encoding substrates with mass tags to resolve stereospecific reactions using Nimzyme. *Rapid Commun. Mass Spectrom.* 26, 611–615. doi:10.1002/rcm.6134
- Deng, K., Takasuka, T. E., Heins, R., Cheng, X., Bergeman, L. F., Shi, J., et al. (2014). Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem. Biol.* 9, 1470–1479. doi:10.1021/cb5000289
- Fox, B. G., Takasuka, T., and Bianchetti, C. M. (2014). *Multifunctional Cellulase and Hemicellulase*. US Patent Application US 20140079683 A1 (Washington, DC: U.S. Patent and Trademark Office).
- Greving, M., Cheng, X. L., Reindl, W., Bowen, B., Deng, K., Louie, K., et al. (2012). Acoustic deposition with NIMS as a high-throughput enzyme activity assay. *Anal. Bioanal. Chem.* 403, 707–711. doi:10.1007/s00216-012-5908-8

- Heins, R., Cheng, X., Nath, S., Deng, K., Bowen, B. P., Chivian, D. C., et al. (2014). Hylogenomically guided identification of industrially relevant GH1 β -glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry. *ACS Chem. Biol.* 9, 2082–2091. doi:10.1021/cb500244v
- Keshwani, D. R., and Cheng, J. J. (2009). Switchgrass for bioethanol and other value-added applications: a review. *Bioresour. Technol.* 100, 1515–1523. doi:10.1016/j.biortech.2008.09.035
- Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., and Blanch, H. W. (2012). The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.* 109, 1083–1087. doi:10.1002/bit.24370
- Li, C. L., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V., Auer, M., et al. (2010). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. *Bioresour. Technol.* 101, 4900–4906. doi:10.1016/j.biortech.2009.10.066
- Malherbe, A. R., Rose, S. H., Viojen-Bloom, M., and van Zyl, W. H. (2014). Expression and evaluation of enzymes required for the hydrolysis of galactomannan. *J. Ind. Microbiol. Biotechnol.* 41, 1201–1209.
- Northen, T. R., Lee, J. C., Hoang, L., Raymond, J., Hwang, D. R., Yannone, S. M., et al. (2008). A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3678–3683. doi:10.1073/pnas.0712332105
- Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A., and Johnson, D. K. (2010). Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulose performance. *Biotechnol. Biofuels* 3, 10. doi:10.1186/1754-6834-3-10
- Pu, Y., Hu, F., Huang, F., Davison, B. H., and Ragauskas, A. J. (2013). Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. *Biotechnol. Biofuels* 6, 15.
- Reindl, W., Deng, K., Gladden, J. M., Cheng, G., Wong, A., Singer, S. W., et al. (2011). Colloid-based multiplexed screening for plant biomass-degrading glycoside hydrolase activities in microbial communities. *Energy Environ. Sci.* 4, 2884–2893. doi:10.1039/c1ee01112j
- Rodríguez-Gacio, M. C., Iglesias-Fernández, R., Carbonero, P., and Matilla, A. J. (2012). Softening-up mannan-rich cell walls. *J. Exp. Bot.* 63, 3976–3988. doi:10.1093/jxb/ers096
- Sharrock, K. R. (1988). Cellulase assay methods: a review. *J. Biochem. Biophys. Methods* 17, 81–105. doi:10.1016/0165-022X(88)90040-1
- Singh, S., Cheng, G., Sathitsuksanoh, N., Wu, D., Varanasi, P., George, A., et al. (2015). Comparison of different biomass pretreatment techniques and their impact on chemistry and structure. *Front. Energy Res.* 2:62. doi:10.3389/fenrg.2014.00062
- Takasuka, T. E., Walker, J. A., Bergeman, L. F., Vander Meulen, K. A., Makino, S.-I., Elsen, N. L., et al. (2014). Cell-free translation of biofuels enzymes. *Methods Mol. Biol.* 1118, 71–95. doi:10.1007/978-1-62703-782-2_5
- van Tilbeurgh, H., Claeysens, M., and de Bruyne, C. K. (1982). The use of 4-methylumbelliferyl and other chromophoric glycosides in the study of cellulolytic enzymes. *FEBS Lett.* 149, 152–156. doi:10.1016/0014-5793(82)81092-2
- Walton, J., Banerjee, G., and Car, S. (2011). GENPLAT: an automated platform for biomass enzyme discovery and cocktail optimization. *J. Vis. Exp.* 56, 3314. doi:10.3791/3314
- Wood, T. M. (1988). Preparation of crystalline, amorphous and dyed cellulase substrate. *Meth. Enzymol.* 166, 19–45. doi:10.1016/0076-6879(88)60103-0
- Wood, T. M., and Bhat, K. M. (1988). Methods for measuring cellulase activities. *Meth. Enzymol.* 160, 87–117. doi:10.1016/0076-6879(88)60109-1
- Xu, Z., and Huang, F. (2014). Pretreatment methods for bioethanol production. *Appl. Biochem. Biotechnol.* 174, 43–62. doi:10.1007/s12010-014-1015-y
- Yaniv, O., Frolov, F., Levy-Assraf, M., Lamed, R., and Bayer, E. A. (2012). Interactions between family 3 carbohydrate binding modules (CBMs) and cellosomal linker peptides. *Meth. Enzymol.* 510, 247–259. doi:10.1016/B978-0-12-415931-0.00013-6
- Zhang, Y.-H., Cui, J., Lynd, L. R., and Kuang, L. R. (2006). A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* 7, 644–648. doi:10.1021/bm050799c

Conflict of Interest Statement: Kai Deng and Trent R. Northen are co-inventors on a patent application that covers the oxime-NIMS assay. Taichi E. Takasuka and Brian G. Fox are co-inventors on a patent application that covers use of multi-functional enzymes. The remaining authors have no conflict of interest to declare.

Copyright © 2015 Deng, Guenther, Gao, Bowen, Tran, Reyes-Ortiz, Cheng, Sathitsuksanoh, Heins, Takasuka, Bergeman, Geertz-Hansen, Deutsch, Loqué, Sale, Simmons, Adams, Singh, Fox and Northen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases

Kai Deng^{1,2}, Taichi E. Takasuka^{3†}, Christopher M. Bianchetti^{3,4}, Lai F. Bergeman³, Paul D. Adams^{1,5,6}, Trent R. Northen^{1,5} and Brian G. Fox^{3,7*}

¹ US Department of Energy Joint BioEnergy Institute, Emeryville, CA, USA, ² Sandia National Laboratories, Livermore, CA, USA, ³ US Department of Energy Great Lakes Bioenergy Research Center, Madison, WI, USA, ⁴ Department of Chemistry, University of Wisconsin-Oshkosh, Oshkosh, WI, USA, ⁵ Lawrence Berkeley National Laboratory, Berkeley, CA, USA, ⁶ Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA, ⁷ Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

OPEN ACCESS

Edited by:

Robert Henry,
The University of Queensland,
Australia

Reviewed by:

Lixin Cheng,
Aarhus University, Denmark
Chiranjeevi Thulluri,
Jawaharlal Nehru Technological
University Hyderabad, India

*Correspondence:

Brian G. Fox
bgfox@biochem.wisc.edu

†Present address:

Taichi E. Takasuka,
Research Faculty of Agriculture,
Hokkaido University, Sapporo, Japan

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 28 July 2015

Accepted: 02 October 2015

Published: 27 October 2015

Citation:

Deng K, Takasuka TE, Bianchetti CM,
Bergeman LF, Adams PD,
Northen TR and Fox BG (2015) Use
of nanostructure-initiator mass
spectrometry to deduce selectivity of
reaction in glycoside hydrolases.
Front. Bioeng. Biotechnol. 3:165.
doi: 10.3389/fbioe.2015.00165

Chemically synthesized nanostructure-initiator mass spectrometry (NIMS) probes derivatized with tetrasaccharides were used to study the reactivity of representative *Clostridium thermocellum* β -glucosidase, endoglucanases, and cellobiohydrolase. Diagnostic patterns for reactions of these different classes of enzymes were observed. Results show sequential removal of glucose by the β -glucosidase and a progressive increase in specificity of reaction from endoglucanases to cellobiohydrolase. Time-dependent reactions of these polysaccharide-selective enzymes were modeled by numerical integration, which provides a quantitative basis to make functional distinctions among a continuum of naturally evolved catalytic properties. Consequently, our method, which combines automated protein translation with high-sensitivity and time-dependent detection of multiple products, provides a new approach to annotate glycoside hydrolase phylogenetic trees with functional measurements.

Keywords: cellulase, assay, kinetics, Nimzyme, mass spectrometry, protein engineering, biofuels

INTRODUCTION

The enzymatic hydrolysis of plant cell wall material is a formidable task because of the complexity of the plant cell wall (Himmel et al., 2007). In most currently deployed cellulosic ethanol plants, enzyme cocktails containing multiple classes of polysaccharide-degrading enzymes are used to hydrolyze plant biomass into fermentable sugars. Understanding the function, synergy, and stability of enzymes is thus of paramount importance in biofuels production.

Polysaccharide-degrading enzymes are classified into families in the carbohydrate active enzyme (CAZy) database (Henrissat and Davies, 1997; Cantarel et al., 2009; Levasseur et al., 2013), including glycoside hydrolases (GHs), pectic lyases (PLs), carbohydrate esterases (CEs), and others. Only a small fraction of the enzymes included in CAZy have a function assigned by biochemical analyses. One root of this limitation arises from difficulties in succeeding with heterologous expression of enzymes after selection from phylogenetic trees (Watson et al., 2007; Fox et al., 2008; Markley

Abbreviations: AFEX-SG, ammonia fiber expansion pretreated switchgrass; CBM, carbohydrate-binding module; CelE, broad specificity GH family 5 (GH5) domain from *C. thermocellum* Cthe_0797; GH, glycoside hydrolase; IL-SG, ionic liquid pretreated switchgrass; NIMS, nanostructure-initiator mass spectrometry.

et al., 2009; Nair et al., 2009; Pieper et al., 2013). As an option to address this limitation, we (Takasuka et al., 2014; Bianchetti et al., 2015) and others (Beebe et al., 2011, 2014; Madono et al., 2011; Hirano et al., 2013, 2015; Makino et al., 2014) have found that wheat germ cell-free protein translation can be used as an effective expression platform to make functional assignments of enzyme function.

Another limitation arises from experimental complications of carrying out high-throughput multisubstrate assays to screen for enzyme function (Gerlt et al., 2011). A breadth of assay methods have been developed for GHs, including use of soluble and insoluble chromogenic and/or fluorogenic substrates, HPLC, and others (Sharrock, 1988; Decker et al., 2003; Chundawat et al., 2008; Bansal et al., 2009; Dowe, 2009; Dashtban et al., 2010; Selig et al., 2011; Eklof et al., 2012; Horn et al., 2012; Kosik et al., 2012; McCleary et al., 2012; Pena et al., 2012; Whitehead et al., 2012; Wischmann et al., 2012). Each of these approaches has intrinsic advantages, but can suffer in sensitivity, complexity of analysis, throughput time, and volumes of reagents and enzyme needed. In comparison, nanostructure-initiator mass spectrometry (NIMS) offers high sensitivity, simplicity of detection of products derived from biomass hydrolysis, microliters or smaller volumes for reaction, and options for automation (Northen et al., 2008; Deng et al., 2012; de Rond et al., 2013; Heins et al., 2014). Recently, we used oxime-NIMS and numerical integration methods to provide time-dependent, quantitative characterization of reducing sugars released by individual enzymes in reactions with pretreated biomass (Deng et al., 2014).

Here, we report a new use of NIMS to provide quantitative analysis of time-dependent reactions of cellulases. The enzymes selected for this study were from *Clostridium thermocellum*, a Gram-positive anaerobe with high cellulolytic capacity (Ding et al., 2008; Fontes and Gilbert, 2010; Smith and Bayer, 2013). The *C. thermocellum* genome encodes ~130 CAZyme domains and ~90 carbohydrate-binding module (CBM) domains (Feinberg et al., 2011). The majority of CAZyme domains also possess dockerin domains, which serve to recruit these enzymes into the cellulosome via dockerin-cohesin interactions (Ding et al., 2008; Smith and Bayer, 2013). The specific gene regulatory and protein secretory patterns of this model consolidated bioprocessing organism have also been well described (Brown et al., 2007; Gold and Martin, 2007; Roberts et al., 2010; Feinberg et al., 2011; Raman et al., 2011; Riederer et al., 2011), and many of the enzymes have been characterized. Given this state of knowledge, individual enzymes from *C. thermocellum* have proven useful for the development and testing of new approaches for assignment of GH function.

In this work, we have used chemically synthesized tetrasaccharide-NIMS probes to study the reactivity of some cellulases from *C. thermocellum*. Patterns of reactivity identified by using the tetrasaccharide-NIMS probes provide a diagnostic approach to assess reaction specificity and also provide comparative apparent rate information. Our results show diagnostic patterns for reactions of a β -glucosidase, relaxed but varied specificity of several endoglucanases and high specificity of a cellobiohydrolase with the model substrate. Time-dependent reactions of these polysaccharide-selective enzymes were modeled by numerical

integration, which provides a quantitative basis to make functional distinctions among a continuum of naturally evolved reactive properties. Consequently, this method, which combines high-sensitivity detection of multiple products with quantitative numerical analysis of their time-dependent formation, provides a new approach to enhance the annotation of GH phylogenetic trees with functional measurements.

MATERIALS AND METHODS

Enzyme Preparation

Methods for cloning, cell-free translation, and purification of the enzymes studied have been reported elsewhere (Takasuka et al., 2014). Briefly, enzymes were cloned by PCR amplification of catalytic domains as indicated by the first and last codons indicated in **Table 1**. Cloned genes were transferred into an optimized wheat germ cell-free translation plasmid pEU-HSCB (Beebe et al., 2011; Takasuka et al., 2014), which is also available from the NIH Protein Structure Initiative Materials Repository (<http://psimr.asu.edu/>). Enzymes were prepared by cell-free translation using either bilayer or dialysis methods (Beebe et al., 2011, 2014; Makino et al., 2014), and active enzymes were identified (Takasuka et al., 2014). The enzymes listed in **Table 1** were also cloned by PCR into the *Escherichia coli* expression vector pEC_CBM3a to create enzyme_CBM3a fusion proteins, CelAcc_CBM3a. The vector pEC_CBM3a is a derivative of pEU_HSBC_CBM3a (Takasuka et al., 2014) that yields fusion proteins having an N-terminal enzyme catalytic domain fused by an ~40 aa linker sequence to the CBM3a domain from Cthe_3077. A stop codon was added to the PCR primer used to amplify the 3' end of the BglA gene so that no fusion to CBM3a was produced from pEU_HSBC_CBM3a. As needed, protein coding sequences were transferred between pEU and pEC vectors by use of FlexiVector cloning (Blommel et al., 2009). Methods for PCR amplification, capture and sequence verification of protein coding sequences, and transformation into *E. coli* 10G competent cells (Lucigen, Middleton, WI, USA) for DNA manipulations and *E. coli* B834 for protein expression were as previously reported (Takasuka et al., 2014). Additional details of the properties and methods for the use of pEU and pVP are described elsewhere (Aceti et al., 2015).

Synthesis of Cellotetraose-NIMS Substrate

The cellotetraose-NIMS substrate (**Figure 1A**) is an amphiphilic molecule that has a sugar head group coupled to a perfluorinated (F17) tag. The detailed synthetic procedure has been reported previously (Deng et al., 2012).

Enzyme Reactions

An enzyme reaction consisted of 10 μ L of 50 mM phosphate, pH 6.0, supplemented with 1 μ L of 1 mM cellotetraose-F17 dissolved in water. An aliquot of each enzyme preparation (containing 1–10 ng of enzyme) was added to initiate the reaction and the resulting mixture was incubated at 37°C. At times of 5, 10, 20, 40, 80, and 120 min, 0.2 μ L of the reaction mixture was withdrawn for analysis.

TABLE 1 | *Clostridium thermocellum* enzymes studied in this work.

Gene locus	Name	CAZy family	First codon ^a	Last codon ^b	Function ^c	CMX class ^d	Reference
Cthe_0212	BglA	GH1	1	448	Exo-β-glucosidase		Grabnitz et al. (1991)
Cthe_0269	CelA	GH8	34	368	Endo-β-1,4-glucanase	CX	1CEM; Alzari et al. (1996)
Cthe_0040	CelL	GH9	28	887	Endo-β-1,4-glucanase	C	2XFG
Cthe_0797	CelE	GH5	36	388	Cellulase, xylanase, mannanase	CMX	Deng et al. (2014, Takasuka)
Cthe_0578	CelR	GH9	27	640	Endo-β-1,4-glucanase; cellotetraohydrolase	C	Zverlov et al. (2005)
Cthe_0405	CelL	GH5	31	430	–	CX	Deng et al. (2014)
Cthe_0412	CelK	GH9	28	809	Cellobiohydrolase	C	Kataeva et al. (1999)
Cthe_3077 ^e	CipA	CBM3a	323	523	Cellulose-binding module		Kataeva et al. (1999)

^aFirst codon of the indicated gene locus that was included in the PCR primer design (Takasuka et al., 2014).

^bLast codon of the indicated gene locus that was included in the PCR primer design.

^cFunction assigned from annotation as defined in CAZy (Cantarel et al., 2009), from experimental evidence cited in the table, or a combination of both.

^dRepresentation of the breadth of substrate specificity for each enzyme (Deng et al., 2014). The CMX classification indicates that CelE can hydrolyze cellulose, xylan, or mannan; CX indicates that CelA and CelL can hydrolyze cellulose and xylan, while CelL, CelR, and CelK can only hydrolyze cellulose. This classification derives from reactions with pure polysaccharides and pretreated biomass (Deng et al., 2014; Takasuka et al., 2014).

^eCBM3a was subcloned from the scaffoldin CipA gene.

Mass Spectrometry

In each case, 0.2 μL per reaction sample was spotted onto the NIMS surface and removed after an incubation of ~30 s. A grid drawn manually on the NIMS chip using a diamond-tip scribe helped with spotting and identification of sample spots in the spectrometer. Chips were loaded using a modified standard MALDI plate. NIMS was performed on a 4800 MALDI TOF/TOF mass analyzer from AB Sciex (Foster City, CA, USA). In each case, signal intensities were identified for the ions of the cellotetraose substrate and, when present, each product shown in **Figure 1**. For each assay, ~1000 laser shots were collected. Enzyme activities were determined by measuring the intensity ratios of each product over the intensity total of ions of for the cellotetraose-, cellotriose-, cellobiose-, glucose-, and aglycone-NIMS (**Figure 2**).

Kinetic Analyses

The time dependence of hydrolysis of the tetrasaccharide-NIMS was analyzed by non-linear global optimization of differential equations accounting for the appearance and decay of products (Deng et al., 2014) using Mathematica routine NDSolve and the Nelder-Mead simplex method for constrained minimization (Nelder and Mead, 1965). The differential equations corresponding to the kinetic scheme of **Figure 3** are as follows:

$$y[1] = \text{cellotetraose-NIMS} \quad (1)$$

$$y[2] = \text{cellotriose-NIMS} \quad (2)$$

$$y[3] = \text{cellobiose-NIMS} \quad (3)$$

$$y[4] = \text{glucose-NIMS} \quad (4)$$

$$y[5] = \text{aglycone-NIMS} \quad (5)$$

$$dy[1]/dt = -(k_1 + k_9 + k_{11} + k_{13}) y[1] \quad (6)$$

$$dy[2]/dt = (k_1) y[1] - (k_3 + k_{15} + k_{17}) y[2] \quad (7)$$

$$dy[3]/dt = (k_9) y[1] + (k_3) y[2] - (k_5 + k_{19}) y[3] \quad (8)$$

$$dy[4]/dt = (k_{11}) y[1] + (k_{15}) y[2] + (k_5) y[3] - (k_7) y[4] \quad (9)$$

$$dy[5]/dt = (k_{13}) y[1] + (k_{17}) y[2] + (k_{19}) y[3] + (k_7) y[4] \quad (10)$$

Initial guesses for apparent rate constants were made by visual inspection of the match between the results of single NDSolve calculations and the experimental data. This process was continued in an iterative way until a set of initial apparent rates that adequately matched the experimental data was obtained. Successive rounds of least squares parameter optimization with adjustment of parameter constraints were carried out until the sum of the squares difference between calculated and experimental values reached a minimum and no parameter was artificially constrained.

RESULTS AND DISCUSSION

Enzymes Chosen for Study

Clostridium thermocellum enzymes were chosen for this study based on previous transcriptomic and proteomic results (Gold and Martin, 2007; Raman et al., 2011; Riederer et al., 2011) and other biochemical and structural results (**Table 1**). Genes encoding these enzymes were expressed using wheat germ cell-free protein synthesis and the translated proteins were assayed using fluorogenic substrates (Takasuka et al., 2014); among the synthesized enzymes, 13 reacted with MUG or MUC, 11 reacted with MUX or MUX2, and 5 reacted with other diagnostic fluorogenic substrates. Reactions of these enzymes with ionic liquid pretreated switchgrass (IL-SG) have been published (Deng et al., 2014). Enzymes from cell-free translation reactions that showed promising characteristics were produced by expression in *E. coli* and purified for use in the studies described here.

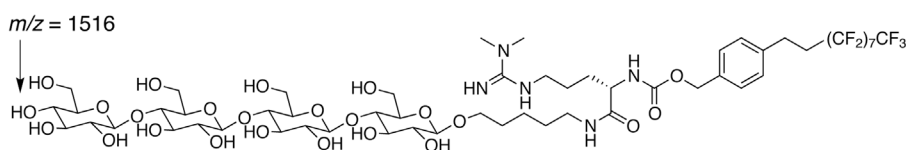
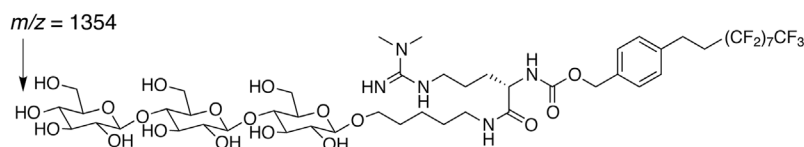
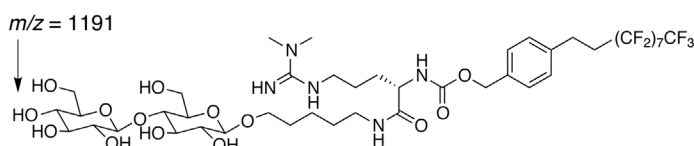
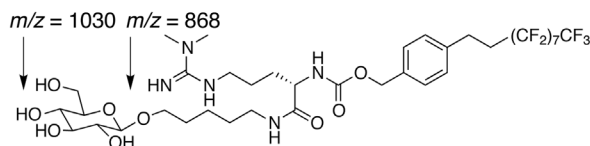
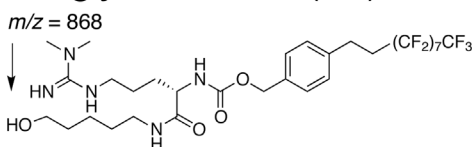
A cellotetraose-NIMS (G4)**B cellotriose-NIMS (G3)****C cellobiose-NIMS (G2)****D glucose-NIMS (G1)****E aglycone-NIMS (G0)**

FIGURE 1 | Structure of cellotetraose-NIMS and m/z values for products obtained from hydrolysis at the indicated anomeric position. (A) cellotetraose-NIMS; (B) cellotriose-NIMS; (C) cellobiose-NIMS; (D) glucose-NIMS; (E) aglycone-NIMS.

Cellotetraose-NIMS Substrate

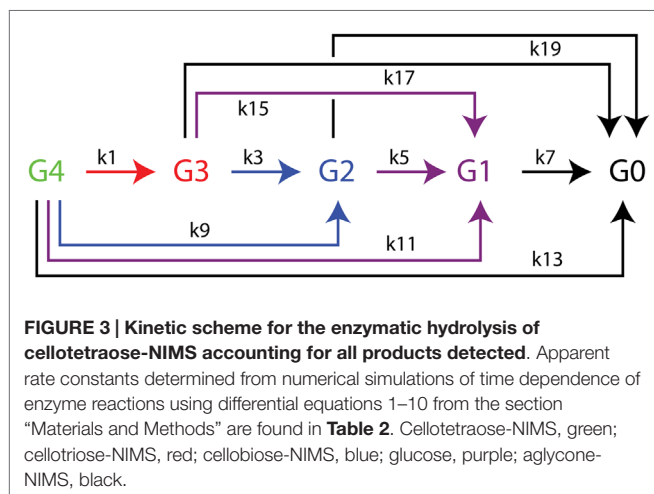
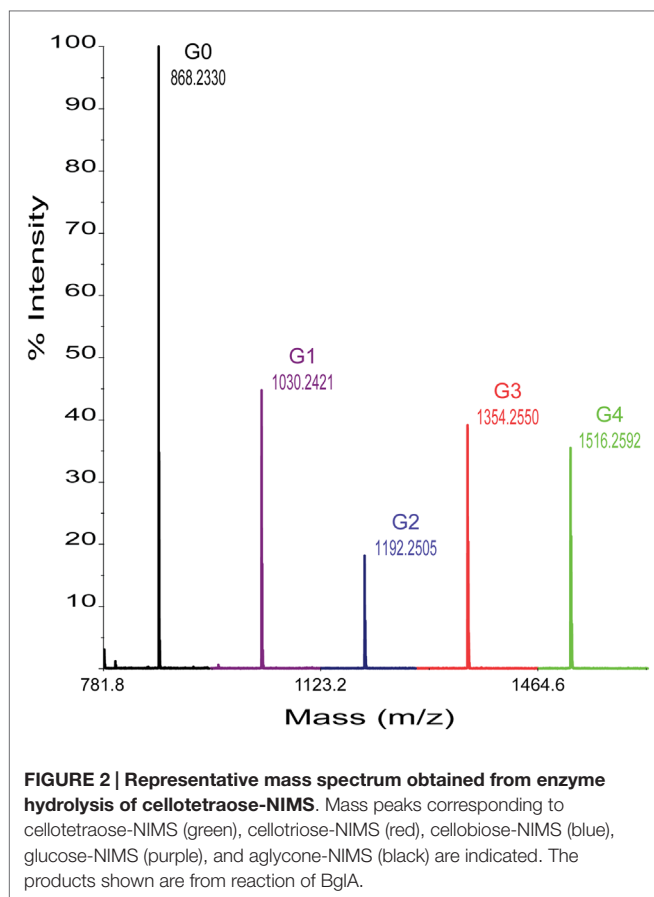
Figure 1 shows the structure of cellotetraose-NIMS and the products that can be formed by various GH reactions. In the synthesized probe, the tetra-saccharide is linked to the NIMS probe by a potentially hydrolyzable anomeric linkage. Synthesis of the NIMS probe and the tetra-saccharide derivatives are summarized in Materials and Methods (Deng et al., 2012; de Rond et al., 2013). The guanidium group on the NIMS probe provides improved ionization properties in the mass spectrometry experiment, while the perfluorinated portion of the NIMS probe provides hydrophobic anchoring of the molecule into the NIMS surface. Enzyme-catalyzed hydrolysis of the anomeric linkages give rise to a cascade of potential products retained on the NIMS surface. Reactions of GHs can progressively remove single glucose units or carry out other reactions that remove cellobiose, cellotriose, or cellotetraose.

Kinetic Scheme

Figure 2 shows a representative mass spectrum obtained after partial reaction with BglA (Cthe_0212), a β -glucosidase. At the selected time point in the reaction (120 min), the cellotetraose-NIMS probe (G4, green) has been partially converted into a mixture of cellotriose (G3, red), cellobiose (G2, blue), glucose (G1, purple), and aglycone (G0, black) derivatives of the NIMS probe. **Figure 3** shows a kinetic scheme that accounts for the potential products shown in **Figure 1**. The scheme accounts for release of one or more glucose units from the cellotetraose-NIMS probe (G4) and its successive products. Time course profiles provide the fundamental data used in this work for numerical analysis of enzyme hydrolysis reactions.

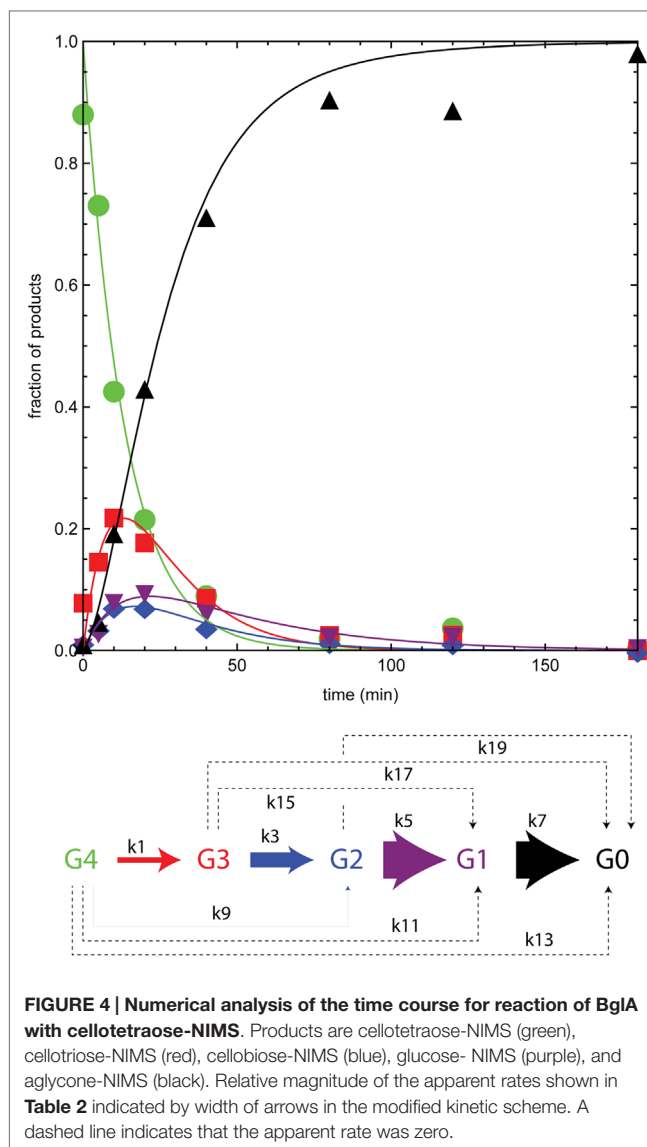
β -Glucosidase BglA Reaction

The nucleotide sequence of BglA (Grabnitz et al., 1991) was published before the genome sequence and annotated to be



a β -glucosidase from the GH1 family (Cantarel et al., 2009). The Cthe_02012 gene does not encode a signal peptide, so the entire gene was cloned for the studies described here. Beyond our characterization of the reaction of BglA with IL-SG (Deng et al., 2014), no other functional studies have been reported for this enzyme.

Figure 4 shows the time course for reaction of BglA with cellotetraose-NIMS. The plotted proportions of the different



products come from time series of mass spectra like those shown in **Figure 2**. The solid colored lines are results of simulations of the concentration of individual products based on the kinetic scheme of **Figure 3** and the differential equations shown in the section “Materials and Methods.” The apparent rate constants provided by the numerical simulation are given in **Table 2**, and a pictorial representation of the relative magnitudes of the apparent rate constants is also given in **Figure 4**. In the time course of the BglA reaction, cellotetraose-NIMS (green circles) was converted to a succession of intermediates by hydrolysis of a single glucose from the position most distal to the NIMS probe. This pattern of reactivity is as expected for the reaction of an exo- β -glucosidase with an oligosaccharide. Thus, cellotriose-NIMS (red squares) accumulated was subsequently converted to cellobiose-NIMS (purple down triangles), to glucose-NIMS (blue diamonds), and ultimately to aglycone-NIMS (black up triangles).

TABLE 2 | Apparent rate constants (min⁻¹) from numerical integration of time course reactions with cellotetraose-NIMS.

Rate	BglA	CeIL	CeLR	CeLE	CeLA	CeLI	CeLK
k1	0.05320	0.000	0.001	0.001	0.000001	0.0002	0.0000
k3	0.14805	0.0002	0.0000	0.0000	0.0000	0.0002	0.0000
k5	0.36803	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k7	0.36803	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k9	0.00047	0.0490	0.0170	0.0092	0.0004	0.0020	0.0439
k11	0.00006	0.0038	0.0453	0.0390	0.0116	0.0479	0.0000
k13	0.00006	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
k15	0.00006	0.0000	0.0000	0.0000	0.0010	0.0000	0.0004
k17	0.00006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k19	0.00006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

Rates for individual enzymes are color coded with smallest values in blue and larger values in red.

There are several features of the BglA reaction and simulation that warrant attention. The apparent rates k1, k3, k5, and k7, which correspond to successive removal of single glucose groups, dominate the numerical solution (Table 2; Figure 4). Under the reaction conditions used, BglA was able to completely convert cellotetraose-NIMS to aglycone-NIMS. It is also noteworthy that shortening the oligosaccharide chain led to an enhancement in the rate of hydrolysis, with reactions k5 (converting cellobiose-NIMS to glucose-NIMS) and k7 (converting glucose-NIMS to aglycone-NIMS) being fastest. Other apparent rates corresponding to side reactions for removal of cellobiose or larger oligosaccharides (e.g., k9 for removal of cellobiose from cellotetraose-NIMS) were less than 1/100th of the value observed for k1, the smallest of the central reactions. These simulation results are consistent with the assigned function of BglA as a β -glucosidase. Indeed, prior oxime-NIMS studies of the reaction of BglA with IL-SG revealed that glucose was the only product released from the biomass substrate (Deng et al., 2014). In the following paragraphs, these diagnostic behaviors of a beta-glucosidase are contrasted with two other classes of GHs, including five phylogenetically diverse endoglucanases and one cellobiohydrolase.

Endoglucanase and Cellobiohydrolase Reactions

Figure 5 shows time courses for reactions of endoglucanases CelA, CelI, CelE, CelR, CelL, and cellobiohydrolase CelK with cellotetraose-NIMS. The reactions of the individual enzymes were carried out and evaluated as described above for Figure 4. The appearance of the reaction time courses and the relative rates observed are markedly different than observed for BglA. Unlike the β -glucosidase reaction, no intermediates were observed to form and decay, and the central reactions corresponding to release of glucose units were negligible. This seemingly corresponds with the requirement of endoglucanases for a longer oligosaccharide chain to occupy the active site as a determinant of productive binding and catalysis. In effect, the endoglucanases and cellobiohydrolase primarily reacted only once with the cellotetraose-NIMS probe, leading to a markedly simpler cascade of products than observed for the beta-glucosidase. None of the enzymes characterized in Figure 5 was able to carry out reactions that yielded the aglycone-NIMS product

(black up triangles), suggesting unproductive binding or blocking steric interactions of the NIMS product with adjacent features of the active site. In contrast, the β -glucosidase BglA (Figure 4) was able to successively remove all glucose groups from cellotetraose-NIMS to yield aglycone-NIMS.

Endoglucanase CelA Reactions

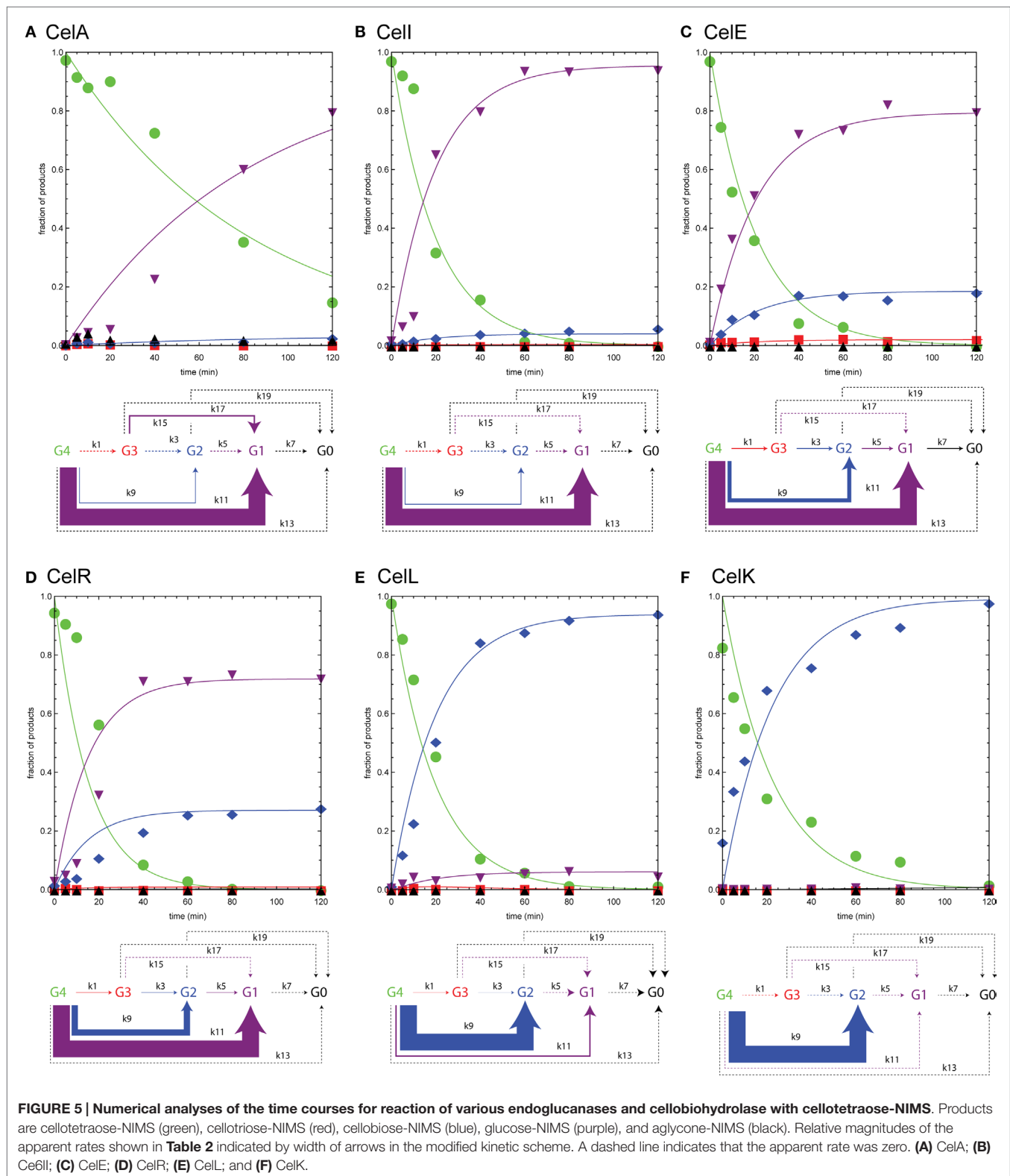
CeLA (Cthe_0269) is a GH8 endoglucanase. It is one of the most abundantly transcribed and secreted proteins in *C. thermocellum* during growth on cellulosic substrates (Brown et al., 2007; Gold and Martin, 2007; Raman et al., 2011; Riederer et al., 2011). Analysis of the crystal structure of the enzyme suggested that the substrate binding channel was optimally configured to bind a cellopentaose molecule (Alzari et al., 1996).

The functional characterizations of Figure 5 demonstrate a progression in reaction selectivity among the enzymes studied. This is a unique power arising from the combination of time-dependent NIMS with numerical analysis. For CelA (Figure 5A), k11 governed removal of cellotriose from cellotetraose-NIMS, leading to the predominant accumulation of glucose-NIMS (88%, purple down triangles). The alternative removal of cellotriose via the two step pathway of k1 (removal of glucose) and k15 (removal of cellobiose) contributed ~9% to the overall product yields, while reaction via k9 (removal of cellobiose) added only ~3% of total products as cellobiose-NIMS (blue diamonds). It is worth noting that CelA gave the slowest hydrolysis of cellotetraose-NIMS of all enzymes tested, which is reflected in the values of apparent rates reported in Table 2 and also in the shape of the plots in Figure 5. This may also reflect a partial rate diminution caused by a mismatch between cellotetraose-NIMS and a preferred cellopentaose occupying the active site channel.

In our earlier reactions of CelA with IL-SG (Deng et al., 2014), a mixture of glucose, cellobiose, triose, and tetraose was observed. Other than cellotetraose, whose release from cellotetraose-NIMS was probably prevented by improper binding of the NIMS moiety in the active site channel, the suite of products given by CelA reaction with cellotetraose-NIMS was comparable to that observed from reactions with the pretreated biomass (Deng et al., 2014).

Endoglucanase CelI, CelE, and CelR Reactions

For the reactions of CelI (Figure 5B), CelE (Figure 5C), and CelR (Figure 5D), the dominant pattern of preferred removal of cellotriose units to yield glucose-NIMS (purple down triangles) was retained. However, functional differences of these three enzymes were identified as the removal of cellobiose leading to cellobiose-NIMS (blue diamonds) assumed an increasing contribution to the total product distribution. For example, the observed change corresponds to an approximately eightfold increase in k9 between CelI and CelR. In the middle of these boundary enzymes, CelE was unique among the endoglucanases tested as it was also able to release a glucose unit from cellotetraose-NIMS in ~2% yield. In reactions with IL-SG and ammonia fiber expansion pretreated switchgrass (AFEX-SG) (Deng et al., 2014), these three enzymes released a mixture of glucose, cellobiose, and cellotriose, with the distribution of products in the biomass reaction shifted toward cellobiose and glucose. However, this shift is, in part, due



to the ability of these enzymes to cleave solubilized cellotriose into cellobiose and glucose. Subsequent hydrolysis of released oligosaccharides could not be detected when cellotetraose-NIMS was the substrate.

CelI (Cthe_0040) has a structure consisting of GH9 and two CBM3 domains (Hazlewood et al., 1993). It catalyzes the hydrolysis of 1,4- β -glucosidic linkages in cellulose and other glucans. The structure suggests the position of a tunnel that can permit the

release of either cellotriose or cellobiose from cellotetraose-NIMS (PDB 2XFG, no associated publication).

CelE (Cthe_0797) is a multidomain enzyme consisting of GH5, dockerin, and GSDL-lipase domains. Our work has shown that the GH5 domain has broad specificity for reaction with cellulose, xylan, mannan, xyloglucan, and other polysaccharides (Deng et al., 2014; Takasuka et al., 2014). The active site channel of this enzyme is open and tolerates the placement of each of these different linear and branched polysaccharides in a way that a glycosidic bond can be placed in the appropriate position for hydrolysis (Bianchetti et al., 2015). The release of cellotriose, cellobiose, and glucose from cellotetraose-NIMS is compatible with this broad specificity active site. Nevertheless, the active site is not sufficiently tolerant to remove cellotetraose, leading to the formation of aglycone-NIMS.

Previous studies have reported that CelR (Cthe_0578) is a β -glucanase with preference for release of cellotetraose in reactions with amorphous cellulose (Zverlov et al., 2005). Subsequently, CelR was able to convert the longer solubilized oligosaccharide to shorter oligosaccharides. The present studies provide support for this conclusion, as k11 for release of cellotriose was the predominant reaction with cellotetraose-NIMS. Our studies of CelR in reactions with IL-SG and AFEX-SG gave glucose and cellobiose as the dominant hydrolysis products (Deng et al., 2014), suggesting a kinetically rapid conversion of longer oligosaccharides to shorter during the duration of the reaction. Removal of cellotetraose was not observed from cellotetraose-NIMS, which as proposed above likely represents ineffective binding of the NIMS probe in the active site adjacent to the active site.

Endoglucanase CelL and Cellobiohydrolase CelK Reactions

We tested the cellotetraose-NIMS reactions with an additional endoglucanase, CelL (Cthe_0405, **Figure 5E**), and a reducing end cellobiohydrolase, CelK (Cthe_0212, **Figure 5F**). These enzymes show a shift in reaction specificity so that removal of cellobiose to produce cellobiose-NIMS (blue diamonds) became the dominant pattern of reaction. Notably, CelL had an approximately threefold enhanced ability to remove cellobiose relative to CelR because of a higher k9 value and also an ~ 10 -fold decrease in the ability to remove cellotriose associated with a lower k11 value (**Table 2**). CelL reacted with IL-SG also showed preference for release of cellobiose (Deng et al., 2014). Furthermore, although CelK also had an approximately threefold enhanced ability to remove

cellobiose relative to CelR because of a higher k9 value, it showed no ability to produce either cellotriose or glucose (e.g., k1 and k11 = 0; **Table 2**).

The high specificity for release of cellobiose by a cellobiohydrolase is a characteristic reactivity (Amano et al., 1996; Barr et al., 1996; Divne et al., 1998), including CelK (Kataeva et al., 1999) and also CelK reacted with IL-SG (Deng et al., 2014). Thus, cellotetraose-NIMS clearly reports on this catalytic function of CelK. There are no previously published reactivity studies or crystal structures of CelL, beyond our studies of reaction with IL-SG, where CelL showed strong preference for release of cellobiose and xylobiose from the pretreated biomass (Deng et al., 2014).

CONCLUSION

This work establishes the utility of a chemically synthesized mass spectral probe for characterization of GHs. We have shown remarkable correspondence between the products obtained from enzyme reactions with the synthetic cellotetraose-NIMS probe and IL- and AFEX-pretreated switchgrass (Deng et al., 2014). Because of the emerging success of robotic cell-free translation to provide active enzyme samples from synthesized genes (Takasuka et al., 2014; Bianchetti et al., 2015), the substantial advantages of automation and miniaturization afforded by the Nimzyme platform (Deng et al., 2012, 2014; de Rond et al., 2013; Heins et al., 2014), and the predictive power inherent in numerical analysis of enzyme reaction time courses (Cleland, 1975; Orsi and Tipton, 1979; Duggleby, 1995; Marangoni, 2003), our combination offers a powerful new approach for functional annotation of bioenergy phylogenetic space.

AUTHOR CONTRIBUTIONS

KD, TT, CB, LB, PA, TN, and BF designed experiments, carried out experimental work, analyzed results, and prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

The DOE Great Lakes Bioenergy Research Center and DOE Joint BioEnergy Institute are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-FC02-07ER64494 and through contract DE-AC02-05CH11231, respectively.

REFERENCES

- Aceti, D. J., Bingman, C. A., Wrobel, R. L., Frederick, R. O., Makino, S., Nichols, K. W., et al. (2015). Expression platforms for producing eukaryotic proteins: a comparison of *E. coli* cell-based and wheat germ cell-free synthesis, affinity and solubility tags, and cloning strategies. *J. Struct. Funct. Genomics* 16, 67–80. doi:10.1007/s10969-015-9198-1
- Alzari, P. M., Souchon, H., and Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 4, 265–275. doi:10.1016/S0969-2126(96)00031-7
- Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E., and Kanda, T. (1996). Fine substrate specificities of four exo-type cellulases produced by *Aspergillus niger*, *Trichoderma reesei*, and *Irpex lacteus* on (1 \rightarrow 3), (1 \rightarrow 4)- β -D-glucans and xyloglucan. *J. Biochem.* 120, 1123–1129. doi:10.1093/oxfordjournals.jbchem.a021531
- Bansal, P., Hall, M., Realff, M. J., Lee, J. H., and Bommarius, A. S. (2009). Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnol. Adv.* 27, 833–848. doi:10.1016/j.biotechadv.2009.06.005
- Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996). Identification of two functionally different classes of exocellulases. *Biochemistry* 35, 586–592. doi:10.1021/bi9520388
- Beebe, E. T., Makino, S., Markley, J. L., and Fox, B. G. (2014). Automated cell-free protein production methods for structural studies. *Methods Mol. Biol.* 1140, 117–135. doi:10.1007/978-1-4939-0354-2_9
- Beebe, E. T., Makino, S., Nozawa, A., Matsubara, Y., Frederick, R. O., Primm, J. G., et al. (2011). Robotic large-scale application of wheat cell-free translation to

- structural studies including membrane proteins. *N Biotechnol.* 28, 239–249. doi:10.1016/j.nbt.2010.07.003
- Bianchetti, C. M., Takasuka, T. E., Deutsch, S., Udell, H. S., Yik, E. J., Bergeman, L. F., et al. (2015). Active site and laminarin binding in glycoside hydrolase family 55. *J. Biol. Chem.* 290, 11819–11832. doi:10.1074/jbc.M114.623579
- Blommel, P. G., Martin, P. A., Seder, K. D., Wrobel, R. L., and Fox, B. G. (2009). Flexi vector cloning. *Methods Mol. Biol.* 498, 55–73. doi:10.1007/978-1-59745-196-3_4
- Brown, S. D., Raman, B., Mckeown, C. K., Kale, S. P., He, Z., and Mielenz, J. R. (2007). Construction and evaluation of a *Clostridium thermocellum* ATCC 27405 whole-genome oligonucleotide microarray. *Appl. Biochem. Biotechnol.* 13, 663–674. doi:10.1007/s12010-007-9087-6
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi:10.1093/nar/gkn663
- Chundawat, S. P., Balan, V., and Dale, B. E. (2008). High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. *Biotechnol. Bioeng.* 99, 1281–1294. doi:10.1002/bit.21805
- Cleland, W. W. (1975). Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry* 14, 3220–3224. doi:10.1021/bi00685a029
- Dashtban, M., Maki, M., Leung, K. T., Mao, C., and Qin, W. (2010). Cellulase activities in biomass conversion: measurement methods and comparison. *Crit. Rev. Biotechnol.* 30, 302–309. doi:10.3109/07388551.2010.490938
- de Rond, T., Peralta-Yahya, P., Cheng, X., Northen, T. R., and Keasling, J. D. (2013). Versatile synthesis of probes for high-throughput enzyme activity screening. *Anal. Bioanal. Chem.* 405, 4969–4973. doi:10.1007/s00216-013-6888-z
- Decker, S. R., Adney, W. S., Jennings, E., Vinzant, T. B., and Himmel, M. E. (2003). Automated filter paper assay for determination of cellulase activity. *Appl. Biochem. Biotechnol.* 105–108, 689–703. doi:10.1385/ABAB:107:1-3:689
- Deng, K., George, K. W., Reindl, W., Keasling, J. D., Adams, P. D., Lee, T. S., et al. (2012). Encoding substrates with mass tags to resolve stereospecific reactions using Nimzyme. *Rapid Commun. Mass Spectrom.* 26, 611–615. doi:10.1002/rcm.6134
- Deng, K., Takasuka, T. E., Heins, R., Cheng, X., Bergeman, L. F., Shi, J., et al. (2014). Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem. Biol.* 9, 1470–1479. doi:10.1021/cb5000289
- Ding, S. Y., Xu, Q., Crowley, M., Zeng, Y., Nimlos, M., Lamed, R., et al. (2008). A biophysical perspective on the cellulosome: new opportunities for biomass conversion. *Curr. Opin. Biotechnol.* 19, 218–227. doi:10.1016/j.copbio.2008.04.008
- Divne, C., Stahlberg, J., Teeri, T. T., and Jones, T. A. (1998). High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 275, 309–325. doi:10.1006/jmbi.1997.1437
- Dowe, N. (2009). Assessing cellulase performance on pretreated lignocellulosic biomass using saccharification and fermentation-based protocols. *Methods Mol. Biol.* 581, 233–245. doi:10.1007/978-1-60761-214-8_15
- Duggleby, R. G. (1995). Analysis of enzyme progress curves by nonlinear regression. *Meth. Enzymol.* 249, 61–90. doi:10.1016/0076-6879(95)49031-0
- Eklof, J. M., Ruda, M. C., and Brumer, H. (2012). Distinguishing xyloglucanase activity in endo-beta(1 - >4)glucanases. *Meth. Enzymol.* 510, 97–120. doi:10.1016/B978-0-12-415931-0.00006-9
- Feinberg, L., Foden, J., Barrett, T., Davenport, K. W., Bruce, D., Detter, C., et al. (2011). Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. *J. Bacteriol.* 193, 2906–2907. doi:10.1128/JB.00322-11
- Fontes, C. M., and Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79, 655–681. doi:10.1146/annurev-biochem-091208-085603
- Fox, B. G., Goulding, C., Malkowski, M. G., Stewart, L., and Deacon, A. (2008). Structural genomics: from genes to structures with valuable materials and many questions in between. *Nat. Methods* 5, 129–132. doi:10.1038/nmeth0208-129
- Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., et al. (2011). The enzyme function initiative. *Biochemistry* 50, 9950–9962. doi:10.1021/bi201312u
- Gold, N. D., and Martin, V. J. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* 189, 6787–6795. doi:10.1128/JB.00882-07
- Grabnitz, F., Seiss, M., Rucknagel, K. P., and Staudenbauer, W. L. (1991). Structure of the beta-glucosidase gene bglA of *Clostridium thermocellum*. Sequence analysis reveals a superfamily of cellulases and beta-glycosidases including human lactase/phlorizin hydrolase. *Eur. J. Biochem.* 200, 301–309. doi:10.1111/j.1432-1033.1991.tb16186.x
- Hazlewood, G. P., Davidson, K., Laurie, J. I., Huskisson, N. S., and Gilbert, H. J. (1993). Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. *J. Gen. Microbiol.* 139, 307–316. doi:10.1099/00221287-139-2-307
- Heins, R. A., Cheng, X., Nath, S., Deng, K., Bowen, B. P., Chivian, D. C., et al. (2014). Phylogenomically guided identification of industrially relevant GH1 beta-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry. *ACS Chem. Biol.* 9, 2082–2091. doi:10.1021/cb500244v
- Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi:10.1016/S0959-440X(97)80072-3
- Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi:10.1126/science.1137016
- Hirano, K., Nihei, S., Hasegawa, H., Haruki, M., and Hirano, N. (2015). Stoichiometric assembly of the cellulosome generates maximum synergy for the degradation of crystalline cellulose, as revealed by in vitro reconstitution of the *Clostridium thermocellum* cellulosome. *Appl. Environ. Microbiol.* 81, 4756–4766. doi:10.1128/AEM.00772-15
- Hirano, N., Hasegawa, H., Nihei, S., and Haruki, M. (2013). Cell-free protein synthesis and substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest multi-enzyme subunit of the *Clostridium thermocellum* cellulosome. *FEMS Microbiol. Lett.* 344, 25–30. doi:10.1111/1574-6968.12149
- Horn, S. J., Sorlie, M., Varum, K. M., Valjamae, P., and Eijsink, V. G. (2012). Measuring processivity. *Meth. Enzymol.* 510, 69–95. doi:10.1016/B978-0-12-415931-0.00005-7
- Kataeva, I., Li, X. L., Chen, H., Choi, S. K., and Ljungdahl, L. G. (1999). Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum*: evidence for gene duplication and recombination. *J. Bacteriol.* 181, 5288–5295.
- Kosik, O., Bromley, J. R., Busse-Wicher, M., Zhang, Z., and Dupree, P. (2012). Studies of enzymatic cleavage of cellulose using polysaccharide analysis by carbohydrate gel electrophoresis (PACE). *Meth. Enzymol.* 510, 51–67. doi:10.1016/B978-0-12-415931-0.00004-5
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6, 41. doi:10.1186/1754-6834-6-41
- Madono, M., Sawasaki, T., Morishita, R., and Endo, Y. (2011). Wheat germ cell-free protein production system for post-genomic research. *N Biotechnol.* 28, 211–217. doi:10.1016/j.nbt.2010.08.009
- Makino, S., Beebe, E. T., Markley, J. L., and Fox, B. G. (2014). Cell-free protein synthesis for functional and structural studies. *Methods Mol. Biol.* 1091, 161–178. doi:10.1007/978-1-62703-691-7_11
- Marangoni, A. G. (2003). *Enzyme Kinetics*. Hoboken, NJ: John Wiley & Sons, Inc.
- Markley, J. L., Aceti, D. J., Bingman, C. A., Fox, B. G., Frederick, R. O., Makino, S., et al. (2009). The center for eukaryotic structural genomics. *J. Struct. Funct. Genomics* 10, 165–179. doi:10.1007/s10969-008-9057-4
- McCleary, B. V., Mckie, V., and Draga, A. (2012). Measurement of endo-1,4-beta-glucanase. *Meth. Enzymol.* 510, 1–17. doi:10.1016/B978-0-12-415931-0.00001-X
- Nair, R., Liu, J., Soong, T. T., Acton, T. B., Everett, J. K., Kouranov, A., et al. (2009). Structural genomics is the largest contributor of novel structural leverage. *J. Struct. Funct. Genomics* 10, 181–191. doi:10.1007/s10969-008-9055-6
- Nelder, J. A., and Mead, R. (1965). A simplex-method for function minimization. *Comput. J.* 7, 308–313. doi:10.1093/comjnl/7.4.308
- Northen, T. R., Lee, J. C., Hoang, L., Raymond, J., Hwang, D. R., Yannone, S. M., et al. (2008). A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3678–3683. doi:10.1073/pnas.0712321105
- Orsi, B. A., and Tipton, K. F. (1979). Kinetic analysis of progress curves. *Meth. Enzymol.* 63, 159–183. doi:10.1016/0076-6879(79)63010-0
- Pena, M. J., Tuomivaara, S. T., Urbanowicz, B. R., O'Neill, M. A., and York, W. S. (2012). Methods for structural characterization of the products of cellulose- and xyloglucan-hydrolyzing enzymes. *Meth. Enzymol.* 510, 121–139. doi:10.1016/B978-0-12-415931-0.00007-0

- Pieper, U., Schlessinger, A., Kloppmann, E., Chang, G. A., Chou, J. J., Dumont, M. E., et al. (2013). Coordinating the impact of structural genomics on the human alpha-helical transmembrane proteome. *Nat. Struct. Mol. Biol.* 20, 135–138. doi:10.1038/nsmb.2508
- Raman, B., Mckeown, C. K., Rodriguez, M. Jr., Brown, S. D., and Mielenz, J. R. (2011). Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol.* 11:134. doi:10.1186/1471-2180-11-134
- Riederer, A., Takasuka, T. E., Makino, S., Stevenson, D. M., Bukhman, Y. V., Elsen, N. L., et al. (2011). Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl. Environ. Microbiol.* 77, 1243–1253. doi:10.1128/AEM.02008-10
- Roberts, S. B., Gowen, C. M., Brooks, J. P., and Fong, S. S. (2010). Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst. Biol.* 4:31. doi:10.1186/1752-0509-4-31
- Selig, M. J., Tucker, M. P., Law, C., Doeppke, C., Himmel, M. E., and Decker, S. R. (2011). High throughput determination of glucan and xylan fractions in lignocelluloses. *Biotechnol. Lett.* 33, 961–967. doi:10.1007/s10529-011-0526-7
- Sharrock, K. R. (1988). Cellulase assay methods: a review. *J. Biochem. Biophys. Methods* 17, 81–105. doi:10.1016/0165-022X(88)90040-1
- Smith, S. P., and Bayer, E. A. (2013). Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* 23, 686–694. doi:10.1016/j.sbi.2013.09.002
- Takasuka, T. E., Walker, J. A., Bergeman, L. F., Vander Meulen, K. A., Makino, S., Elsen, N. L., et al. (2014). Cell-free translation of biofuel enzymes. *Methods Mol. Biol.* 1118, 71–95. doi:10.1007/978-1-62703-782-2_5
- Watson, J. D., Sanderson, S., Ezersky, A., Savchenko, A., Edwards, A., Orengo, C., et al. (2007). Towards fully automated structure-based function prediction in structural genomics: a case study. *J. Mol. Biol.* 367, 1511–1522. doi:10.1016/j.jmb.2007.01.063
- Whitehead, C., Gomez, L. D., and Mcqueen-Mason, S. J. (2012). The analysis of saccharification in biomass using an automated high-throughput method. *Meth. Enzymol.* 510, 37–50. doi:10.1016/B978-0-12-415931-0.00003-3
- Wischmann, B., Toft, M., Malten, M., and Mcfarland, K. C. (2012). Biomass conversion determined via fluorescent cellulose decay assay. *Meth. Enzymol.* 510, 19–36. doi:10.1016/B978-0-12-415931-0.00002-1
- Zverlov, V. V., Schantz, N., and Schwarz, W. H. (2005). A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-beta-1,4-glucanase producing cellotetraose. *FEMS Microbiol. Lett.* 249, 353–358. doi:10.1016/j.femsle.2005.06.037

Conflict of Interest Statement: Kai Deng and Trent R. Northen are coinventors on a patent application that covers the oxime-NIMS assay. Taichi E. Takasuka, Christopher M. Bianchetti, and Brian G. Fox are coinventors on a patent application that covers use of multifunctional enzymes. Lai F. Bergeman and Paul D. Adams have no conflict of interest to declare.

Copyright © 2015 Deng, Takasuka, Bianchetti, Bergeman, Adams, Northen and Fox. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Immunological approaches to biomass characterization and utilization

Sivakumar Pattathil^{1,2*}, Utku Avci^{1,2}, Tiantian Zhang¹, Claudia L. Cardenas^{1†} and Michael G. Hahn^{1,2}

¹ Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA, ² Oak Ridge National Laboratory, BioEnergy Science Center (BESC), Oak Ridge, TN, USA

OPEN ACCESS

Edited by:

Jason Lupoi,
University of Queensland, USA

Reviewed by:

Xu Fang,
Shandong University, China
Arumugam Muthu,
Council of Scientific and Industrial
Research, India

*Correspondence:

Sivakumar Pattathil
siva@ccrc.uga.edu

†Present address:

Claudia L. Cardenas,
Central Piedmont Community
College, Charlotte, NC, USA

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the
journal *Frontiers in Bioengineering
and Biotechnology*

Received: 20 August 2015

Accepted: 12 October 2015

Published: 28 October 2015

Citation:

Pattathil S, Avci U, Zhang T,
Cardenas CL and Hahn MG (2015)
*Immunological approaches to
biomass characterization
and utilization.*
Front. Bioeng. Biotechnol. 3:173.
doi: 10.3389/fbioe.2015.00173

Plant biomass is the major renewable feedstock resource for sustainable generation of alternative transportation fuels to replace fossil carbon-derived fuels. Lignocellulosic cell walls are the principal component of plant biomass. Hence, a detailed understanding of plant cell wall structure and biosynthesis is an important aspect of bioenergy research. Cell walls are dynamic in their composition and structure, varying considerably among different organs, cells, and developmental stages of plants. Hence, tools are needed that are highly efficient and broadly applicable at various levels of plant biomass-based bioenergy research. The use of plant cell wall glycan-directed probes has seen increasing use over the past decade as an excellent approach for the detailed characterization of cell walls. Large collections of such probes directed against most major cell wall glycans are currently available worldwide. The largest and most diverse set of such probes consists of cell wall glycan-directed monoclonal antibodies (McAbs). These McAbs can be used as immunological probes to comprehensively monitor the overall presence, extractability, and distribution patterns among cell types of most major cell wall glycan epitopes using two mutually complementary immunological approaches, glycome profiling (an *in vitro* platform) and immunolocalization (an *in situ* platform). Significant progress has been made recently in the overall understanding of plant biomass structure, composition, and modifications with the application of these immunological approaches. This review focuses on such advances made in plant biomass analyses across diverse areas of bioenergy research.

Keywords: glycome profiling, immunolocalization, cell walls, biomass, antibodies

INTRODUCTION

Complexity and Dynamics of Plant Cell Walls Constituting Biomass

Plant biomass, the prime feedstock for lignocellulosic biofuel production, constitutes the principal sustainable resource for renewable bioenergy. Identifying the optimal plant biomass types that are most suitable for biofuel production and optimizing their downstream processing and utilization are at the forefront of modern-day lignocellulosic feedstock research. The focus of much of this research is the examination of diverse classes of plants for their potential as cost-effective and sustainable raw

materials for biofuel production. For example, biomass materials originating from classes of plants ranging from herbaceous dicots (e.g., alfalfa), woody dicots (e.g., poplar), perennial monocots (e.g., *Agave* spp.), herbaceous monocots (e.g., grasses such as *Miscanthus*, sugarcane, and switchgrass), and woody gymnosperms (e.g., pines) are regarded as potentially promising resources for biofuel production (Galbe and Zacchi, 2007; Gomez et al., 2008; Somerville et al., 2010).

Cell walls constitute the major part of plant biomass, and physicochemical features of these cell walls vary among biomass materials from diverse plant classes (Pauly and Keegstra, 2008; Popper, 2008; Fangel et al., 2012). For example, cell walls from grass biomass have distinct structural and compositional features [with a higher abundance of glucuronarabinoxylans and the presence of mixed-linkage glucans (Vogel, 2008)] that are quite different from those of highly lignified woody biomass (Studer et al., 2011) or herbaceous dicot biomass (Burton et al., 2010; Liepman et al., 2010). Even within a plant, the structure and composition of cell walls can vary significantly depending on the cell types, organs, age, developmental stage, and growth environment (Freshour et al., 1996; Knox, 2008). These cell wall variations are the result of differences in the relative proportions and structural dynamics that occur among the major cell wall polymers, which include (but are not limited to) cellulose, hemicelluloses, pectic polysaccharides, and lignin (Pauly and Keegstra, 2008). Several structural models for plant cell walls have been proposed and published (McNeil et al., 1984; McCann and Roberts, 1991; Carpita and Gibeau, 1993; Carpita, 1996; Cosgrove, 1997; Somerville et al., 2004; Loqué et al., 2015); all of these models focus on the primary wall. To our knowledge, no model has been proposed for secondary plant cell walls, which constitute the bulk of the biomass used for bioenergy production. In vascular plants, non-glycan components such as lignin (especially in secondary cell wall-containing tissues such as sclerenchyma and xylem cells) are important for optimal growth and development of plants by playing important roles in maintaining cell wall integrity to optimally facilitate water transportation, rendering mechanical support and defense against pathogens (Weng and Chapple, 2010; Voxeur et al., 2015). A high abundance of lignin in cell walls is regarded as disadvantageous for biomass utilization for biofuel production as it contributes significantly to recalcitrance. Transgenic plants that are genetically modified for reduced lignin biosynthesis have been shown to exhibit reduced recalcitrance properties (Chen and Dixon, 2007; Pattathil et al., 2012b). The abundance of diverse potential plant biomass feedstocks that are available to be studied and the aforementioned variations among the cell walls constituting them pose a major challenge in lignocellulosic bioenergy research.

Research on the structure, function, and biosynthesis of plant cell walls has received new impetus with advances in genome sequencing that have made available, for the first time, whole genomes from diverse plant families. Thus, complete genomes have been sequenced for plants from diverse phylogenetic classes including both herbaceous [e.g., *Arabidopsis* (The Arabidopsis Genome Initiative, 2000); *Medicago* (Young et al., 2011)] and woody dicots [e.g., *Populus* (Tuskan et al., 2006)] and

monocotyledonous grasses [e.g., maize (Schnable et al., 2009), rice (Goff et al., 2002; Yu et al., 2002), and brachypodium (The International Brachypodium Initiative, 2010)]. The availability of these genome sequences has, in turn, dramatically expanded experimental access to genes and gene families involved in plant primary and secondary cell wall biosynthesis and modification. Functional characterization of cell wall-related genes and the proteins that they encode, combined with expanded research on cell wall deconstruction, have dramatically enhanced our understanding of wall features important for biomass utilization.

Genetic Approaches to Studies of Cell Walls with Impacts on Lignocellulosic Bioenergy Research

Cell walls are known for their innate resistance to degradation and specifically to the breakdown of their complex polysaccharides into simpler fermentable sugars that can be utilized for microbial production of biofuels. This property of plant cell walls is referred to as “recalcitrance” (Himmel et al., 2007; Fu et al., 2011). Cell wall recalcitrance has been identified as the most well-documented challenge that limits biomass conversion into sustainable and cost-effective biofuel production (Himmel et al., 2007; Pauly and Keegstra, 2008; Scheller et al., 2010). Hence, identifying cell wall components that affect recalcitrance has been an important target of lignocellulosic bioenergy research (Ferraz et al., 2014). A number of plant cell wall polymers, including lignin, hemicelluloses, and pectic polysaccharides, have been shown to contribute to cell wall recalcitrance (Mohnen et al., 2008; Fu et al., 2011; Studer et al., 2011; Pattathil et al., 2012b).

Most of the studies directed toward overcoming recalcitrance focus on genetically modifying plants by specifically targeting genes involved in the biosynthesis or modification of wall polymers (Chen and Dixon, 2007; Mohnen et al., 2008; Fu et al., 2011; Studer et al., 2011; Pattathil et al., 2012b) with the objective of generating a viable, sustainable biomass crop that synthesizes cell walls with reduced recalcitrance. Identification of target genes for reducing recalcitrance has relied largely on model plant systems, particularly *Arabidopsis*, and then to transfer that information to biofuel crops. This has been particularly successful for genes and pathways that participate directly or indirectly in secondary cell wall biosynthesis and development. Secondary walls constitute the bulk of most biofuel feedstocks and thus become a main target for genetic modification (Chundawat et al., 2011; Yang et al., 2013). Secondary wall synthetic genes that have been investigated in this way include, for example, several genes that are involved in cellulose [such as various *CesA* genes (Joshi et al., 2004, 2011; Taylor et al., 2004; Brown et al., 2005; Ye et al., 2006)] and xylan biosynthesis [*IRX8* (Brown et al., 2005; Ye et al., 2006; Peña et al., 2007; Oikawa et al., 2010; Liang et al., 2013), *IRX9* (Brown et al., 2005; Lee et al., 2007, 2011a; Peña et al., 2007; Oikawa et al., 2010; Liang et al., 2013), *IRX9L* (Oikawa et al., 2010; Wu et al., 2010), *IRX14* (Oikawa et al., 2010; Wu et al., 2010; Lee et al., 2011a), *IRX14L* (Wu et al., 2010; Lee et al., 2011a), *IRX15* (Brown et al., 2011), and *IRX15L* (Brown et al., 2011)] in dicots. In addition, a number of transcription factors including plant-specific NAC-domain

transcription factors [*SND1*, *NST1*, *VND6*, and *VND7* in *Arabidopsis* (Kubo et al., 2005; Zhong et al., 2006, 2007b)], WRKY transcription factors [in *Medicago* and *Arabidopsis* (Wang et al., 2010; Wang and Dixon, 2012)], and MYB transcription factors [*MYB83* (McCarthy et al., 2009) and *MYB46* (Zhong et al., 2007a) in *Arabidopsis*] with potential involvement in secondary wall biosynthesis and development have been functionally characterized. Examples of the successful transfer of insights gained in model dicots to studies of orthologous genes in monocots include investigations of rice *IRX* orthologs involved in xylan biosynthesis and secondary wall formation (Oikawa et al., 2010) and experiments on transcription factors controlling secondary wall formation in several grasses (Handakumbura and Hazen, 2012; Shen et al., 2013; Valdivia et al., 2013). These molecular genetic approaches toward understanding and manipulating cell wall-related genes for biofuel feedstock improvement would be assisted by improved methods for rapidly identifying and characterizing the effects of genetic changes on cell wall components.

Need for Efficient Tools for Plant Cell Wall/Biomass Analyses

The structural complexity of plant cell walls, regardless of their origin, is challenging to analyze, particularly in a high-throughput manner. To date, most of the plant cell wall analytical platforms have been based on the preparation of cell wall materials and/or extracts that are selectively enriched for particular wall polysaccharides, followed by colorimetric assays (Selvendran and O'Neill, 1987), chemical derivatizations coupled with gas chromatography (Albersheim et al., 1967; Sweet et al., 1974, 1975a,b), mass spectroscopy (Lerouxel et al., 2002), and nuclear magnetic resonance spectroscopy (NMR) (Peña et al., 2008) to gain compositional and structural information about those polysaccharides. Some of these methods have been adapted for biomass analytics [see, for review, Sluiter et al. (2010)]. Overall, these tools have allowed extensive progress in delineating basic structural features of diverse classes of plant cell wall polysaccharides. However, these experimental approaches for plant cell wall/biomass analysis are time-consuming, require specialized and, in some cases, expensive equipment, are low in throughput, and usually provide information only about a single polysaccharide of specific interest. However, given the number of wall components that have already been shown to influence cell wall recalcitrance, and the complex and heterogeneous nature of cell wall components in diverse plants, it is desirable to have additional tools, particularly those with higher throughput and the capability to monitor a broad spectrum of wall polymers. Over the past 10 years, immunological approaches for plant cell wall and biomass analyses have emerged as tools that are broadly applicable to multiple aspects of interests to the biofuel research community, including characterization of genetically altered plant feedstocks, investigations of the effects of diverse biomass pretreatment processes, and the effects of enzymatic or microbial deconstruction of cell walls. In the following sections, we review applications of two immunological tools for studies on plant biomass that employ a comprehensive collection of plant cell wall glycan-directed probes.

PROBES FOR BIOMASS ANALYSES

Currently, well-characterized cell wall-directed probes range from small molecules (Wallace and Anderson, 2012) to larger proteinaceous probes such as carbohydrate-binding modules (CBMs) and monoclonal or polyclonal antibodies (Knox, 2008; Pattathil et al., 2010; Lee et al., 2011b). In this review, we will focus on the latter cell wall-directed probes.

Glycan-Directed Probes

Monoclonal Antibodies

Plant cell wall glycan-directed monoclonal antibodies (McAbs) are among the most commonly used probes for plant cell wall analyses. McAbs, commonly available as hybridoma culture supernatants, are monospecific probes that recognize specific glycan sub-structures (epitopes) present in plant polysaccharides (Knox, 2008; Pattathil et al., 2010). McAbs have several advantages that make them particularly suited for use as glycan-directed probes. First, since each antibody is the product of a single clonal cell line, each McAb is by definition monospecific with regard to the epitope that is recognized. This is important for studies of glycans, whose structures are frequently repetitive and whose substructures can be found in multiple macromolecular contexts (e.g., arabinogalactan epitopes present on glycoproteins and on rhamnogalacturonan I). The monospecific nature of McAbs also means that, in theory, the binding specificity of the antibody can be determined unambiguously, although this is still difficult for glycan-directed antibodies given the complexity of plant cell wall glycan structures. McAbs also typically bind to their epitopes with high affinity ($K_d \sim 10^{-6}$ M), which makes them very sensitive reagents for detecting and quantitating molecules to which they bind. Finally, another significant advantage with McAbs is that their supply is not limited, as cell lines producing them can be cryopreserved indefinitely (some hybridoma lines whose plant glycan-directed antibodies are frequently used today were generated more than 20 years ago) and can be regrown at any time to produce additional McAb, which retains the binding selectivity and affinity of the original McAb, as needed in any quantities required. Currently, a worldwide collection of over 200 McAbs (Pattathil et al., 2010, 2012a) exists (**Figure 1**) that encompasses antibodies recognizing diverse structural features of most major non-cellulosic cell wall glycans, including arabinogalactans, xyloglucans, xylans, mannans, homogalacturonans, and rhamnogalacturonan I. So far, McAbs that bind reliably and specifically to rhamnogalacturonan II have not been reported. The available plant glycan-directed McAbs can be obtained from several stock centers (see **Table 1**) or from the individual research laboratories that generated them. A listing of the McAbs currently available is not practical here. The reader is referred to a plant cell wall McAb database, *WallMabDB*,¹ where detailed descriptions of most of the currently available plant glycan-directed McAbs, including immunogen, antibody isotype, and epitope structure (to the extent known), can be obtained.

¹<http://www.wallmabdb.net>.

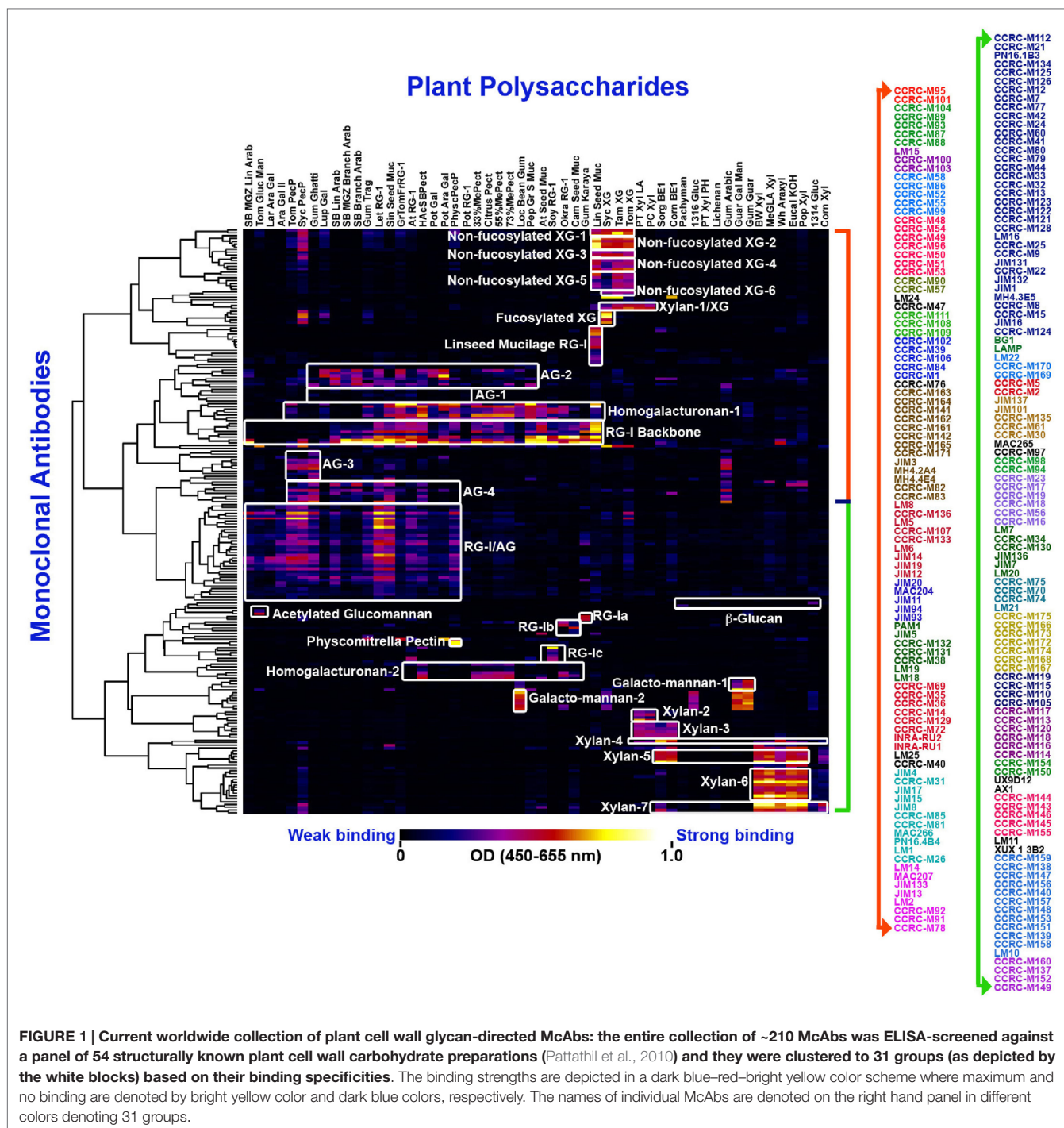


FIGURE 1 | Current worldwide collection of plant cell wall glycan-directed McAbs: the entire collection of ~210 McAbs was ELISA-screened against a panel of 54 structurally known plant cell wall carbohydrate preparations (Pattathil et al., 2010) and they were clustered to 31 groups (as depicted by the white blocks) based on their binding specificities. The binding strengths are depicted in a dark blue–red–bright yellow color scheme where maximum and no binding are denoted by bright yellow color and dark blue colors, respectively. The names of individual McAbs are denoted on the right hand panel in different colors denoting 31 groups.

Early studies in our laboratory screened 130 of the plant glycan-directed McAbs available at the time for their binding specificity to 54 structurally characterized polysaccharide preparations from diverse plants (Pattathil et al., 2010). Hierarchical clustering analyses of the resultant binding response data resolved the McAbs into 19 antibody clades based on their binding specificities to the 54 plant glycans tested (Pattathil et al., 2010). A more recent study that included almost all available plant

glycan-directed McAbs further resolved the antibody collection into about 31 clades of McAbs (Pattathil et al., 2012a). **Figure 1** shows the data from most recent screening studies employing ~210 plant glycan-directed McAbs. While these broad specificity screens provide considerable information about the binding specificities of the McAbs in the collection, they do not provide complete detailed epitope information for the antibodies. Such detailed epitope characterization studies require the availability

TABLE 1 | List of major CBMs currently used for plant cell wall analyses.

Group	Protein	Enzyme	Organism	Type	Reference
A. Cellulose-binding group	CBM1	Cellulase	<i>Trichoderma reesei</i>	Crystalline cellulose	Reinikainen et al. (1992)
	CBM2a	Xylanase 10A	<i>Cybister japonicus</i>	Crystalline cellulose	Bolam et al. (1998)
	CBM3a	Scaffoldin	<i>Clostridium thermocellum</i>	Crystalline cellulose	Tormo et al. (1996)
	CBM10	Xylanase 10A	<i>Cybister japonicus</i>	Crystalline cellulose	Gill et al. (1999)
	CBM4-1	Cellulase 9B	<i>Cellulomonas fimi</i>	Amorphous cellulose	Tomme et al. (1996)
	CBM17	Cellulase 5A	<i>Clostridium cellulovorans</i>	Amorphous cellulose	Boraston et al. (2000)
	CBM28	Cellulase 5A	<i>Bacillus</i> sp. no. 1139	Amorphous cellulose	Boraston et al. (2002)
	CBM9-2	Xylanase 10A	<i>Thermotoga maritima</i>	The ends of cellulose chain	Boraston et al. (2001)
B. Xylan-binding group	CBM2b-1-2	Xylanase 11A	<i>Cellulomonas fimi</i>	Both decorated and unsubstituted xylan	Bolam et al. (2001)
	CBM4-2	Xylanase 10A	<i>Rhodothermus marinus</i>	Both decorated and unsubstituted xylan	Abou Hachem et al. (2000)
	CBM6	Xylanase 11A	<i>Clostridium thermocellum</i>	Both decorated and unsubstituted xylan	Czjzek et al. (2001)
	CBM15	Xylanase 10C	<i>Cybister japonicus</i>	Both decorated and unsubstituted xylan	Szabó et al. (2001)
	CBM22-2	Xylanase 10B	<i>Clostridium thermocellum</i>	both decorated and unsubstituted xylan	Charnock et al. (2000)
	CBM35	Arabino-furano-sidase 62A	<i>Cybister japonicus</i>	Unsubstituted xylan	Bolam et al. (2004)
C. Mannan-binding group	CBM27 (TmMan5)	Mannanase 5C	<i>Thermotoga maritima</i>	Mannan	Filonova et al. (2007) and Zhang et al. (2014)
	CBM35 (Cjman5C)	Mannanase 5C	<i>Cybister japonicus</i>	Mannan	Filonova et al. (2007) and Zhang et al. (2014)
D. Xyloglucan-binding group	CBMXG34	Modified xylanase 10A	<i>Rhodothermus marinus</i>	Non-fucosylated xyloglucan	Gunnarsson et al. (2006)
	CBMXG34/1-X	Modified xylanase 10A	<i>Rhodothermus marinus</i>	Non-fucosylated xyloglucan	von Schantz et al. (2009)
	CBMXG34/2-VI	Modified xylanase 10A	<i>Rhodothermus marinus</i>	Non-fucosylated xyloglucan	von Schantz et al. (2009)
	CBMXG35	Modified xylanase 10A	<i>Rhodothermus marinus</i>	Non-fucosylated xyloglucan	Gunnarsson et al. (2006)
E. Pectic galactan-binding group	TmCBM61	GH53 endo- β -1,4-galac-tanase	<i>Thermotoga maritima</i>	β -1,4-galactan	Cid et al. (2010)

of purified, structurally characterized oligosaccharide fragments and/or purified and characterized glycosylhydrolases capable of selectively attacking epitope structures. To date, a relatively small number of plant glycan-directed McAbs have had their epitopes characterized in detail using these resources (Meikle et al., 1991, 1994; Puhlmann et al., 1994; Steffan et al., 1995; Willats et al., 2000a; Clausen et al., 2003, 2004; McCartney et al., 2005; Verhertbruggen et al., 2009; Marcus et al., 2010; Ralet et al., 2010; Pedersen et al., 2012; Schmidt et al., 2015). Recent advances in methods for immobilization of oligosaccharides on solid surfaces (Fukui et al., 2002; Wang et al., 2002; Willats et al., 2002; Blixt et al., 2004; Pedersen et al., 2012) is facilitating such epitope characterization studies, but the bottleneck remains the availability of comprehensive sets of purified, well-characterized plant glycan-related oligosaccharides.

Carbohydrate-Binding Modules

Carbohydrate-binding modules are another set of proteinaceous probes that have been used to study plant polysaccharide localization patterns *in vivo* (Knox, 2008). CBMs are amino acid sequences that are contiguous with the catalytic domain in a carbohydrate-active enzyme and are capable of binding to a

carbohydrate structural domain (McCartney et al., 2006; Knox, 2008). CBMs have been shown to enhance the efficiency of cell wall hydrolytic enzymes by facilitating sustained and close contact between their associated catalytic modules and targeted substrates (Boraston et al., 2004; Zhang et al., 2014). Although CBMs have been known to occur in several plant enzymes, most CBMs that are used as probes for cell wall glycans are microbial in origin (Boraston et al., 2004; Shoseyov et al., 2006). CBMs, in contrast to the antibody probes described above, are relatively easy to prepare, given that their gene/protein sequences are known (McCann and Knox, 2011). CBMs have been classified into 71 sequence-based families.² CBMs from approximately half of these families have been shown to bind to diverse plant cell wall polysaccharides, including cellulose (Blake et al., 2006), mannans (Filonova et al., 2007), xylans (McCartney et al., 2006), and most recently, the galactan side chains of rhamnogalacturonan I (Cid et al., 2010). Protein engineering of a xylan-binding CBM using random mutagenesis, phage-display technology, and affinity maturation has been employed to generate xyloglucan-specific

²<http://www.cazy.org/Carbohydrate-Binding-Modules.html>.

CBMs (Gunnarsson et al., 2006; von Schantz et al., 2009, 2012), showing that it is possible to generate CBMs with new and heretofore unseen specificities.

Carbohydrate-binding modules that have been used to detect cellulose, xylan, mannan, xyloglucan, and pectic galactans in plant cells and tissues, together with information about their origins, are listed in **Table 1**. Binding of various CBMs is usually assessed by an indirect triple-labeling immunofluorescence procedure (His-tagged CBM, anti-His mouse-Ig, and anti-mouse Ig fluorescein isothiocyanate) in plant tissue sections (Knox, 2008; Hervé et al., 2010), which is slightly more complicated than the double-labeling procedure used with McAbs (Avci et al., 2012). The binding specificities exhibited by the CBMs enlarge the suite of probes available for biomass analyses, given that at least some of them bind to carbohydrate structures, such as cellulose substructures, for which no McAbs probes have been developed to date. Additional advantages of the CBMs are the availability of their gene and protein sequences and the wealth of structural information, including in many instances X-ray crystal structures, about their binding sites. Potential disadvantages of CBMs are their typically lower affinity for their ligands and the lower selectivity of their binding sites compared with McAb probes. Nonetheless, CBMs are useful probes for analyzing biomass.

Immunological Probes Against Lignin

Lignins are phenylpropanoid polymers comprising 5–30% of biomass weight and have been considered as important sources of renewable aromatics (McKendry, 2002). Lignin composition and structure vary considerably depending on the plant species and on the cell type where lignins are deposited (Ruel et al., 1994; Donaldson, 2001). For example, in gymnosperms, lignins are mainly composed of guaiacyl units, whereas in angiosperms, lignins are formed by guaiacyl and syringyl units (Donaldson, 2001). In angiosperms, the guaiacyl-containing lignins are located mainly in secondary cell walls of vessels while syringyl-containing lignins are found on fibers (Ruel et al., 1994; Joseleau et al., 2004; Patten et al., 2010). Lignin composition and localization are also affected by pretreatment strategies aimed at removing lignin from biomass. For example, potassium permanganate labeling and electron microscopy studies revealed morphological alterations in *Zea mays* lignins subjected to different thermochemical pretreatments (Donohoe et al., 2008).

Lignin is most frequently visualized in plant tissue sections using selectively reactive histochemical stains such as phloroglucinol–HCl and Mäule reaction that can distinguish guaiacyl-enriched from syringyl-enriched cell wall regions (Patten et al., 2010). Although the various histochemical lignin stains provide general information about the localization of different lignin types, they cannot provide detailed information about specific lignin substructures; this would require more highly selective probes.

Given the structural complexity and variability of lignin, several laboratories have undertaken the development of immunological probes for lignins and/or lignin substructures. Much of the early work in this area focused on the production of polyclonal antisera. Thus, polyclonal antisera were raised against synthetic dehydrogenative polymers (DHPs)

prepared from the appropriate *p*-hydroxycinnamic alcohols [*p*-hydroxyphenylpropane (H), guaiacyl (G), or syringyl (S), or mixtures of these] (Ruel et al., 1994; Joseleau et al., 2004). These polyclonal sera showed specificity toward the DHPs used to generate them. Other laboratories have generated polyclonal sera against milled wood lignin (Kim and Koh, 1997) or model compounds based on lignin substructures (Kukkola et al., 2003, 2004). The main difficulty with these polyclonal sera is that they are in limited supply, and many of these antisera are no longer available. Thus, new immunizations must be carried out, with uncertain outcomes with regard to the ability to reproduce the specificity of the original antisera; a fundamental problem with polyclonal antisera. In an effort to overcome this limited supply issue, two lignin-related model compounds, dehydrodiconiferyl alcohol and pinosresinol, were used to generate McAbs against these two lignin dimers (Kiyoto et al., 2013); supplies of these antibodies should not be limited. The antibody directed to dehydrodiconiferyl alcohol (KM1) displayed specificity toward a dehydrodiconiferyl alcohol 8-5' model compound, whereas the antibody directed against pinosresinol (KM2) responded to two 8-8' model compounds, pinosresinol and syringaresinol. This recent development suggests that it will be possible, in principle, to generate specific McAbs against diverse lignin substructures. The number and diversity of lignin-directed McAbs will need to be increased in order to fully exploit these probes for greater insights into lignin structural diversity, localization patterns, and integration into the plant cell wall.

TWO MAJOR APPROACHES FOR McAb/CBM-BASED ANALYSES OF PLANT BIOMASS

The use of McAb/CBM probes to define the localization of plant cell wall components has a long history. These probes have been used in basic plant cell wall research to study the effects of mutations in wall-related genes on plant cell wall structure and composition, to study changes in plant cell walls during growth, development, and differentiation, and to study changes in plant cell walls that result from environmental and pathogenic influences. A comprehensive review of this literature is beyond the scope of this minireview and the reader is referred to several recent reviews to gain an overview of this literature (Knox, 1997, 2008; Willats et al., 2000b; Lee et al., 2011b; McCann and Knox, 2011). The use of McAb probes, in particular, is rapidly expanding due to the recent dramatic increase in the number and diversity of plant cell wall-directed antibodies (Pattathil et al., 2010) and the availability of more detailed information about the epitopes recognized by these McAbs (Pedersen et al., 2012; Schmidt et al., 2015).

We will concentrate here on an overview of recent studies that have taken advantage of the availability of the comprehensive collection of cell wall-directed McAb/CBM probes for studying plant biomass of interest as possible lignocellulosic feedstocks for biofuel production. These studies have focused on using these probes to understand the effects of genetic modification on biomass recalcitrance, to study the effects of different pretreatment

regimes on biomass digestibility, and to study how microbes being considered for consolidated bioprocessing deconstruct plant biomass. Two complementary experimental approaches have been principally employed in these studies, namely, glycome profiling (Moller et al., 2007, 2008; Pattathil et al., 2012a) and immunolocalization (Avci et al., 2012). The following sections provide an overview of the studies with bioenergy implications done to date using these approaches.

Studies Using Glycome Profiling

Glycome profiling involves the sequential extraction of insoluble cell wall/biomass samples with a series of reagents of increasing harshness and then screening the extracted cell wall materials with McAbs to determine which cell wall polymers are released in which extract. Thus, this experimental method provides two pieces of important information: (1) it provides detailed information about the composition of the biomass/cell walls; and (2) it provides information on how tightly the various wall components that can be detected are linked into the wall structure. The method is limited by the number of probes (McAbs, CBMs, etc.) used in the screen and the extent to which they are able to recognize the full breadth of wall components released by the extractive reagents. The substantial increase in number and diversity of cell wall probes over the past 10 years has dramatically improved the power and versatility of glycome profiling as a technique for rapid screening of cell wall/biomass samples.

The versatility of glycome profiling is also limited by the ability to immobilize the extracted wall components to a solid support. Diverse solid supports have been used, including nitrocellulose (Moller et al., 2007, 2008), glass slides (Pedersen et al., 2012), and multiwell plastic plates (Pattathil et al., 2012a). All of these suffer the limitation that most low-molecular-weight cell wall components that might be released in the wall extracts, especially low-molecular-weight glycans, do not bind to the solid supports without modification and therefore cannot be assayed by glycome profiling. The lower limit of the glycan size that will adhere has not been definitively determined, but is greater than 10 kDa (Pattathil et al., 2010).

The choice of extractive reagents that have been used for glycome profiling analyses has varied, as has their order. However, typically, the extractive reagents are used in order of increasing severity. Thus, relatively mild reagents, such as CDTA (Moller et al., 2007) or oxalate (Pattathil et al., 2012a), are used first, typically extracting primarily arabinogalactans and pectins. Harsher base extractions then follow, in which primarily hemicelluloses (e.g., xylans and xyloglucans) are extracted (Moller et al., 2007; Pattathil et al., 2012a). For samples that contain significant amounts of lignin, which is the case for most biomass samples of interest to the biofuel industry, an acidic chlorite extraction (Ahlgren and Goring, 1971; Selvendran et al., 1975) is used to degrade the lignin and release lignin-associated wall glycans; this chlorite extraction has most frequently been used after the first base extractions (Pattathil et al., 2012a) but has also been used as the first extraction step (de Souza et al., 2013). None of the extraction sequences used to date yield exclusively one kind of polymer in any given extract, an indication that each wall glycan exists as different subclasses that vary in their extent of cross-linking/

interactions within the wall. Ultimately, the choice of extraction reagents and their order depends on the individual investigator and the specific research questions under investigation.

Two approaches for glycome profiling of plant biomass/cell wall samples have been described. The first, termed comprehensive microarray polymer profiling (CoMPP), is a dot blot-based assay system utilizing nitrocellulose as the solid support (Moller et al., 2007, 2008) and typically employs ~20 glycan-directed probes for screening of three sequential extracts [CDTA (50 mM), 4M NaOH, and Cadoxen (33% v/v)] prepared from plant cell walls. The number of glycan-directed probes that could be used in CoMPP can readily be expanded. An alternative, ELISA-based approach, termed glycome profiling, uses 384-well microtiter plates as the solid support, and uses a broadly diverse toolkit of 155 plant glycan-directed McAbs (Pattathil et al., 2012a) to screen sets of sequentially prepared plant biomass/cell wall extracts [typically, oxalate (50 mM), carbonate (50 mM), 1M KOH, 4M KOH, acidified chlorite, and 4M KOH post-chlorite]. The use of a suite of 155 McAbs ensures a wide-ranging coverage of multiple structural features on most of the major non-cellulosic plant wall glycans (Zhu et al., 2010; Pattathil et al., 2012a). The ELISA-based approach used in glycome profiling lends itself to facile automation and quantitation of antibody binding, hence substantially increasing the throughput of the analyses.

Glycome profiling has seen broad application to diverse experimental approaches in lignocellulosic bioenergy research, including analyzing cell walls from native/genetically modified, variously pretreated, and microbially/enzymatically converted plant biomass (DeMartini et al., 2011; Duceppe et al., 2012; Lee et al., 2012; Tan et al., 2013; Biswal et al., 2015; de Souza et al., 2015; Pattathil et al., 2015; Trajano et al., 2015). Both CoMPP and glycome profiling have been used to undertake comparative glycomics of plant cell wall samples originating from diverse plant phylogenies (Popper et al., 2011; Sørensen et al., 2011; Duceppe et al., 2012; Kulkarni et al., 2012). Examples of such analyses applied to questions related to bioenergy research include a recent study assessing the genetic variability of cell wall degradability of a selected number of *Medicago* cultivars with superior saccharification properties (Duceppe et al., 2012) and an examination of five grass species that revealed commonalities and variations in the overall wall composition and extractability of epitopes among these grasses (Kulkarni et al., 2012). Glycome profiling has also been employed as an effective tool for analyzing cell walls from biomass crops that are genetically modified with the aim of reducing recalcitrance. Examples include examination of the effects on recalcitrance of mutations in lignin biosynthesis in alfalfa [*cad1* (cinnamyl alcohol dehydrogenase 1) (Zhao et al., 2013) and *hct* (hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase) (Pattathil et al., 2012b)] and overexpression of the secondary wall-related transcription factor, PvMYB4 in switchgrass (Shen et al., 2013).

Analyses using cell wall-directed probes have allowed the rapid identification and monitoring of structural and compositional alterations that occur in plant biomass under various regimes of pretreatments (Alonso-Simón et al., 2010; DeMartini et al., 2011; Li et al., 2014; Socha et al., 2014; Pattathil et al., 2015; Trajano et al., 2015). Studies on hydrothermally pretreated wheat straw using CoMPP showed that severe pretreatment regimes

induce significant alterations in wheat straw biomass, including reduction in various hemicellulose and mixed-linkage glucan epitopes (Alonso-Simón et al., 2010). In a more recent study, glycome profiling of poplar biomass subjected to low, medium, and severe hydrothermal pretreatment regimes demonstrated that a series of structural and compositional changes occur in poplar cell walls during this pretreatment, including the rapid disruption of lignin–polysaccharide interactions even under mild conditions, with a concomitant loss of pectins and arabinogalactans, followed by significant removal of hemicellulose (xylans and xyloglucans) (DeMartini et al., 2011). The major inference from this study was that lignin content *per se* does not affect recalcitrance; instead, it is the associations/cross-links between polymers, for example, between lignin and various polysaccharides, within cell walls that play a larger role (DeMartini et al., 2011). Glycome profiling has also been used to examine the effects of other types of pretreatment regimes such as Ammonia Fiber Expansion (AFEX™), alkaline hydrogen peroxide (AHP), and various types of ionic liquids (ILs) on the composition and extractability of wall glycan epitopes in biomass samples from diverse bioenergy crop plants (Li et al., 2014; Socha et al., 2014; Pattathil et al., 2015). These studies demonstrate that, unlike hydrothermal pretreatment, these three types of pretreatment, in general, cause loosening of specific classes of non-cellulosic glycans from plant cell walls, thereby contributing to the reduced recalcitrance exhibited by the pretreated biomasses. Conclusions from these studies contribute significantly to a deeper understanding of pretreatment mechanisms and ultimately will enable optimization of biomass pretreatment regimes and perhaps further downstream utilization processes for biomass from different plant feedstocks.

Glycome profiling has also been used to identify cell wall components that affect biomass recalcitrance. A recent study examined poplar and switchgrass biomass subjected to different pretreatments and correlated pretreatment-induced changes in the biomass with recalcitrance properties of the treated biomass samples (DeMartini et al., 2013). A set of samples with varying composition and structure was generated from native poplar and switchgrass biomass via defined chemical and enzymatic extraction. Subsequently, glycome profiling of the extracts was employed to delineate which wall components were removed and residual solid pretreated biomass samples were analyzed for their recalcitrance features. Major conclusions from this study are that pretreatment regimes affect distinct biomass samples differently and that the most important contributors to recalcitrance vary depending on the biomass. Thus, lignin content appears to play an important role in biomass recalcitrance particularly in woody biomass such as poplar (as they contain higher levels of lignin). However, subclasses of hemicellulose were key recalcitrance-causing factors in grasses such as switchgrass. These results may have important implications for the biofuel industry as they suggest that biomass-processing conditions may have to be tailored to the biomass being used as the feedstock for biofuel generation (DeMartini et al., 2013).

Another bioenergy-related area that has benefited from the use of plant cell wall glycan-directed probes is research into how microbes, particularly those being selected for biomass

deconstruction, degrade plant biomass during culture. Such knowledge will be useful for bioengineering microbes for better biomass conversion. An analysis of biological conversion of unpretreated wild-type sorghum and various *brown midrib* (*bmr*) lines by *Clostridium phytofermentans* examined variations in extractable polysaccharide epitopes of the cell-wall fractions in detail using glycome profiling (Lee et al., 2012). The conclusions were that the loosely integrated xylans and pectins are the primary polysaccharide targets of *C. phytofermentans* and that these are more accessible in the *bmr* mutants than in the wild-type plants (Lee et al., 2012). In another study, an anaerobic thermophilic bacterium, *Caldicellulosiruptor bescii*, was shown to solubilize both lignin and carbohydrates simultaneously in switchgrass biomass at high temperature (Kataeva et al., 2013). Further studies with *C. bescii* demonstrated that deletion of a cluster of genes encoding pectic-degrading enzymes in this organism compromised the ability of *C. bescii* to grow on diverse biomass samples (Chung et al., 2014). A comparative analysis of hemicellulose utilization potentials of *Clostridium clariflavum* and *Clostridium thermocellum* strains demonstrated that *C. clariflavum* strains were better able to grow on untreated switchgrass biomass and degraded easily extractable xylans more readily than do *C. thermocellum* strains (Izquierdo et al., 2014). In all of these studies, glycome profiling proved to be a very effective tool for understanding what was happening to the biomass during culture with the microbes. Studies of this kind provide information about the mode of action of microbial strains on plant biomass, thus identifying wall components that are resistant/recalcitrant to microbial actions.

Studies Using Immunolocalization

Immunolocalization techniques use fixed and embedded (generally in plastic resins) biomass samples (Knox, 1997; Lee et al., 2011b). Primary probes (polyclonal antibodies, McAbs, and CBMs) are applied on semithin sections followed by probing with a fluorescently tagged secondary antibody that allows visualization of glycan epitope localization/distribution under a fluorescent microscope (Avci et al., 2012; Lee and Knox, 2014). This approach for biomass analyses provides information regarding the distribution of cell wall glycans at the cellular and subcellular levels.

A handful of studies thus far have employed this technique in the context of bioenergy research for analyses of cell walls in wall biosynthetic mutants and in pretreated biomass. Examination of *Arabidopsis* and *Medicago* mutants in which a WRKY transcription factor was knocked out revealed secondary cell wall thickening in pith cells caused by ectopic deposition of lignin, xylan, and cellulose. In the *Arabidopsis* mutant, this ectopic secondary wall formation resulted in an approximately 50% increase in biomass density in stem tissue (Yu et al., 2014). The use of three xylan-directed McAbs and a cellulose-directed CBM were instrumental in proving the ectopic deposition of these cell wall glycans in pith cells. In another recent study, the use of two xylan-directed CBMs (CBM2b-1-2 and CBM35 recognizing different degrees of methyl esterification on xylan) on the *Arabidopsis gxmt-1* mutant demonstrated a reduction of 4-O-methyl esterification of xylans (up to 75% as detected by chemical analyses) with a concomitant

reduction in the recalcitrance of mutant walls (Urbanowicz et al., 2012). Additional studies also implicate the importance of secondary wall xylan for cell wall recalcitrance. Restoration of xylan synthesis in xylan-deficient mutants, as documented using xylan-directed McAbs, could, in some cases, yield plants with reduced xylan deposition compared with wild-type plants, but with normal growth habits and decreased recalcitrance (Petersen et al., 2012). Likewise, reduction of xylan in rice culm cell walls yielded plants with slightly lower stature, but with reduced recalcitrance (Chen et al., 2013).

Plant glycan-directed probes (McAbs and CBMs) can also be used to study the distribution patterns of glycan epitopes in plant biomass after diverse pretreatments used to reduce cell wall recalcitrance. One example of such a study is the demonstration that increasingly harsh hydrothermal pretreatments lead to an increased loss of various hemicellulosic, pectic, and cellulosic epitopes in cell walls of the pretreated tissues (DeMartini et al., 2011). The effects of other pretreatment methods (Alonso-Simón et al., 2010; DeMartini et al., 2013; Li et al., 2014; Socha et al., 2014; Pattathil et al., 2015; Trajano et al., 2015) on glycan epitope distribution patterns have not yet been carried out. Such information could be potentially useful to chemical engineers for the optimization of pretreatment conditions to enable optimal biomass conversion.

Immunolocalization studies have documented lignin distribution patterns in plant cell walls that may be relevant to bioenergy research. For instance, cell wall ultrastructure studies using three polyclonal antisera against DHPs allowed visualization of where these types of lignin-related polymers were located in cells of *Zea mays* L. (Joseleau and Ruel, 1997), *Arabidopsis thaliana*, *Nicotiana tabacum*, and *Populus tremula* (Ruel et al., 2002). These studies showed that H-DHPs were present in cell corners and middle lamella, whereas G-DHPs and G/S-DHPs were mainly present in secondary cell walls. The syringylpropane DHP epitope was visualized mainly in the S2 layer of secondary cell walls of *A. thaliana*, *N. tabacum*, and *P. tremula* (Joseleau et al., 2004). Recently, immunogold labeling analyses using KM1 and KM2 demonstrated the presence of 8-5' and 8-8' linked structures, respectively, on either developed xylem or phloem fibers of *Chamaecyparis obtusa* (Kiyoto et al., 2013). It will likely be informative to use these and other lignin-directed probes to monitor lignin distribution patterns in biomass that has been subjected to various pretreatment regimes and/or subjected to microbial degradation in the context of biomass conversion.

Concluding Remarks

The application of high affinity, highly selective molecular probes against plant cell wall polymers clearly has high potential to provide complementary and supplementary data to existing chemical and biochemical analyses for studies on plant biomass structure and conversion. The number and diversity of McAb and CBM probes directed against plant polymers is now sufficiently large that these probes can provide extensive information about cell wall composition and structure in native and pretreated or microbially digested biomass. We have reviewed two main approaches using these probes for biomass characterization and conversion studies. Both glycome profiling/CoMPP and immunolocalization

methods provide distinct but complementary information about the cell walls that constitute the bulk of plant biomass. Glycome profiling and CoMPP provide extensive information about the epitope composition and epitope extractability of polymers present in the biomass. Histochemical approaches using these probes provide valuable information about the spatial distribution of wall epitopes at all levels of organization, ranging from whole plants, to organs, to tissues, to cells, and even to individual cell walls and cell wall domains.

It is important to recognize several attributes of molecular probes directed against cell wall glycan epitopes, in particular, when interpreting the results of experiments. Both McAbs and CBMs are epitope-directed probes, that is, they specifically recognize particular structural motifs. Hence, glycan-directed McAbs and CBMs may not always be polymer-specific, in as much as glycan structures are frequently present in multiple molecular contexts within plant cell walls (e.g., arabinogalactan epitopes present on both polypeptide and polysaccharide backbones). Hence, positive binding of a McAb or CBM probe does not necessarily infer the presence of a particular cell wall glycan polymer. Likewise, the absence of binding of a given McAb or CBM does not unambiguously infer the absence of the glycan detected by this probe; the epitope may be absent or chemically modified (e.g., acetylated or methylated) such that the probe does not bind, but the polymer may still be present (Avci et al., 2012). Furthermore, plant glycans exist as families of polymers, whose epitope composition may not be uniform among all family members. Thus, a single McAb or CBM probe may not bind to all members of a polymer family, and it is therefore advisable to use multiple probes against diverse epitopes on a particular glycan to obtain a comprehensive picture of its abundance either in cell wall extracts or in histochemical localization studies. The size and diversity of the McAb/CBM collections now make such comprehensive studies possible.

Glycome profiling and CoMPP are dependent on the successful immobilization of cell wall-derived molecules to solid supports (e.g., plastic ELISA plates or nitrocellulose). Cell wall glycans with lower molecular masses (less than 20 kDa) have been found not to adhere reliably to the plates (Pattathil et al., 2010, 2012a). Hence, using glycome profiling as a tool to gather information regarding low-molecular-weight cell wall glycans is not advisable unless alternative strategies are employed to ensure adherence of these molecules to a solid support [e.g., covalent attachment directly to the solid support (Schmidt et al., 2015) or to a protein carrier that adheres to the solid support (Pedersen et al., 2012)]. Both glycome profiling and CoMPP also rely on chemical/enzymatic extractions of biomass/cell wall samples. Such extractions are rarely complete or quantitative and thus absolute quantitation of epitope composition in biomass/cell wall samples using these approaches is problematic. Thus, these approaches are best used as initial broad glycome characterization screens, particularly in comparative studies (e.g., mutant vs. wild-type and pretreated vs. untreated) where they provide valuable information regarding changes in the cell wall/biomass samples as a result of a particular experimental manipulation. In histochemical studies, the embedding medium used may influence the results of labeling experiments; in our laboratory, we

have found LR White to give the most consistent results with both McAb and CBM probes (Avci et al., 2012).

FUTURE PERSPECTIVES

The molecular probe toolkits (McAb and CBM) currently available provide an invaluable resource for plant biomass analyses of relevance to bioenergy research and biomass conversion process development. In spite of the number and diversity of the probes currently available, there is still a need for additional probes against structural features not encompassed by the binding specificities of the probes currently available. Thus, additional probes against lignin substructures, rhamnogalacturonan II, and cellulose would further enhance the utility of the probe toolkit. In addition, coverage by the current probe collection of the epitope diversity for some cell wall glycans (e.g., mannans, glucomannans, and galactomannans) is limited. Finally, there remains a need to obtain more detailed information regarding the binding

specificities of many of the molecular probes in the toolkit; about one third of the glycan-directed McAbs have had their epitope specificities characterized in detail. Efforts are underway in multiple laboratories to address these needs. Thus, we can look forward to an enhanced toolkit of probes against plant cell wall polymers in the future.

ACKNOWLEDGMENTS

Immunological studies on biomass characterization conducted in our laboratory are supported by the BioEnergy Science Center administered by Oak Ridge National Laboratory and funded by a grant (DE-AC05-00OR22725) from the Office of Biological and Environmental Research, Office of Science, United States, Department of Energy. The generation of the CCRC series of plant cell wall glycan-directed monoclonal antibodies used in this work was supported by the NSF Plant Genome Program (DBI-0421683 and IOS-0923992).

REFERENCES

- Abou Hachem, M., Karlsson, E. N., Bartonek-Roxa, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., et al. (2000). Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. *Biochem. J.* 345, 53–60. doi:10.1042/0264-6021:3450053
- Ahlgren, P. A., and Goring, D. A. I. (1971). Removal of wood components during chlorite delignification of black spruce. *Can. J. Chem.* 49, 1272–1275. doi:10.1139/v71-207
- Albersheim, P., Nevins, D. J., English, P. D., and Karr, A. (1967). A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography. *Carbohydr. Res.* 5, 340–345. doi:10.1016/S0008-6215(00)80510-8
- Alonso-Simón, A., Kristensen, J. B., Øbro, J., Felby, C., Willats, W. G. T., and Jørgensen, H. (2010). High-throughput microarray profiling of cell wall polymers during hydrothermal pre-treatment of wheat straw. *Biotechnol. Bioeng.* 105, 509–514. doi:10.1002/bit.22546
- Avci, U., Pattathil, S., and Hahn, M. G. (2012). Immunological approaches to plant cell wall and biomass characterization: immunolocalization of glycan epitopes. *Methods Mol. Biol.*, 908, 73–82. doi:10.1007/978-1-61779-956-3_7
- Biswal, A. K., Hao, Z., Pattathil, S., Yang, X., Winkeler, K., Collins, C., et al. (2015). Downregulation of *GAUT12* in *Populus deltoides* by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnol. Biofuels* 8, 41. doi:10.1186/s13068-015-0218-y
- Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., et al. (2006). Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J. Biol. Chem.* 281, 29321–29329. doi:10.1074/jbc.M605903200
- Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., et al. (2004). Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17033–17038. doi:10.1073/pnas.0407902101
- Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., et al. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* 331, 775–781. doi:10.1042/bj3310775
- Bolam, D. N., Xie, H. F., Pell, G., Hogg, D., Galbraith, G., Henrissat, B., et al. (2004). X4 modules represent a new family of carbohydrate-binding modules that display novel properties. *J. Biol. Chem.* 279, 22953–22963. doi:10.1074/jbc.M313317200
- Bolam, D. N., Xie, H. F., White, P., Simpson, P. J., Hancock, S. M., Williamson, M. P., et al. (2001). Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A. *Biochemistry* 40, 2468–2477. doi:10.1021/bi002564l
- Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781. doi:10.1042/BJ20040892
- Boraston, A. B., Chiu, P., Warren, R. A. J., and Kilburn, D. G. (2000). Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* cellulase 5A. *Biochemistry* 39, 11129–11136. doi:10.1021/bi0007728
- Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., et al. (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry* 40, 6240–6247. doi:10.1021/bi0101695
- Boraston, A. B., Ghaffari, M., Warren, R. A. J., and Kilburn, D. G. (2002). Identification and glucan-binding properties of a new carbohydrate-binding module family. *Biochem. J.* 361, 35–40. doi:10.1042/bj3610035
- Brown, D., Wightman, R., Zhang, Z. N., Gomez, L. D., Atanassov, I., Bukowski, J. P., et al. (2011). *Arabidopsis* genes *IRREGULAR XYLEM (IRX15)* and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *Plant J.* 66, 401–413. doi:10.1111/j.1365-3113X.2011.04501.x
- Brown, D. M., Zeef, L. A. H., Ellis, J., Goodacre, R., and Turner, S. R. (2005). Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17, 2281–2295. doi:10.1105/tpc.105.031542
- Burton, R. A., Gidley, M. J., and Fincher, G. B. (2010). Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat. Chem. Biol.* 6, 724–732. doi:10.1038/nchembio.439
- Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 445–476. doi:10.1146/annurev.arplant.47.1.445
- Carpita, N. C., and Gibeaut, D. M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 3, 1–30. doi:10.1111/j.1365-3113X.1993.tb00007.x
- Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M. A., Davies, G. J., et al. (2000). The X6 “thermostabilizing” domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* 39, 5013–5021. doi:10.1021/bi992821q
- Chen, F., and Dixon, R. A. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–761. doi:10.1038/nbt1316
- Chen, X., Vega-Sánchez, M. E., Verhertbruggen, Y., Chiniquy, D., Canlas, P. E., Fagerström, A., et al. (2013). Inactivation of *OsIRX10* leads to decreased xylan

- content in rice culm cell walls and improved biomass saccharification. *Mol. Plant*. 6, 570–573. doi:10.1093/mp/sss135
- Chundawat, S. P. S., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2, 121–145. doi:10.1146/annurev-chembioeng-061010-114205
- Chung, D., Pattathil, S., Biswal, A. K., Hahn, M. G., Mohnen, D., and Westpheling, J. (2014). Deletion of a gene cluster encoding pectin degrading enzymes in *Caldicellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. *Biotechnol. Biofuels* 7, 147. doi:10.1186/s13068-014-0147-1
- Cid, M., Pedersen, H. L., Kaneko, S., Coutinho, P. M., Henrissat, B., Willats, W. G. T., et al. (2010). Recognition of the helical structure of β -1,4-galactan by a new family of carbohydrate-binding modules. *J. Biol. Chem.* 285, 35999–36009. doi:10.1074/jbc.M110.166330
- Clausen, M. H., Ralet, M. C., Willats, W. G. T., McCartney, L., Marcus, S. E., Thibault, J.-F., et al. (2004). A monoclonal antibody to feruloylated-(1,4)- β -D-galactan. *Planta* 219, 1036–1041. doi:10.1007/s00425-004-1309-3
- Clausen, M. H., Willats, W. G. T., and Knox, J. P. (2003). Synthetic methyl hexagalacturonate hapten inhibitors of anti-homogalacturonan monoclonal antibodies LM7, JIM5 and JIM7. *Carbohydr. Res.* 338, 1797–1800. doi:10.1016/S0008-6215(03)00272-6
- Cosgrove, D. J. (1997). Assembly and enlargement of the primary cell wall in plants. *Annu. Rev. Cell Dev. Biol.* 13, 171–201. doi:10.1146/annurev.cellbio.13.1.171
- Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M. G. A., Ferreira, L. M. A., et al. (2001). The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. *J. Biol. Chem.* 276, 48580–48587. doi:10.1074/jbc.M109142200
- de Souza, A. P., Kamei, C. L. A., Torres, A. F., Pattathil, S., Hahn, M. G., Trindade, L. M., et al. (2015). How cell wall complexity influences saccharification efficiency in *Miscanthus sinensis*. *J. Exp. Bot.* 66, 4351–4365. doi:10.1093/jxb/erv183
- de Souza, A. P., Leite, D. C. C., Pattathil, S., Hahn, M. G., and Buckeridge, M. S. (2013). Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production. *Bioenerg. Res.* 6, 564–579. doi:10.1007/s12155-012-9268-1
- DeMartini, J. D., Pattathil, S., Avci, U., Szekalski, K., Mazumder, K., Hahn, M. G., et al. (2011). Application of monoclonal antibodies to investigate plant cell wall deconstruction for biofuels production. *Energy Environ. Sci.* 4, 4332–4339. doi:10.1039/c1ee02112e
- DeMartini, J. D., Pattathil, S., Miller, J. S., Li, H., Hahn, M. G., and Wyman, C. E. (2013). Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energy Environ. Sci.* 6, 898–909. doi:10.1039/c3ee23801f
- Donaldson, L. A. (2001). Lignification and lignin topochemistry – an ultrastructural view. *Phytochemistry* 57, 859–873. doi:10.1016/S0031-9422(01)00049-8
- Donohoe, B. S., Decker, S. R., Tucker, M. P., Himmel, M. E., and Vinzant, T. B. (2008). Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment. *Biotechnol. Bioeng.* 101, 913–925. doi:10.1002/bit.21959
- Duceppe, M.-O., Bertrand, A., Pattathil, S., Miller, J., Castonguay, Y., Hahn, M. G., et al. (2012). Assessment of genetic variability of cell wall degradability for the selection of alfalfa with improved saccharification efficiency. *Bioenerg. Res.* 5, 904–914. doi:10.1007/s12155-012-9204-4
- Fangel, J. U., Ulvskov, P., Knox, J. P., Mikkelsen, M. D., Harholt, J., Popper, Z., et al. (2012). Cell wall evolution and diversity. *Front. Plant Sci.* 3:152. doi:10.3389/fpls.2012.00152
- Ferraz, A., Costa, T. H. F., Siqueira, G., and Milagres, A. M. F. (2014). “Mapping of cell wall components in lignified biomass as a tool to understand recalcitrance,” in *Biofuels in Brazil*, eds Da Silva S. S. and Chandel A. K. (Heidelberg: Springer International), 173–202.
- Filonova, L., Kallas, A. M., Greffe, L., Johansson, G., Teeri, T. T., and Daniel, G. (2007). Analysis of the surfaces of wood tissues and pulp fibers using carbohydrate-binding modules specific for crystalline cellulose and mannan. *Biomacromolecules* 8, 91–97. doi:10.1021/bm060632z
- Freshour, G., Clay, R. P., Fuller, M. S., Albersheim, P., Darvill, A. G., and Hahn, M. G. (1996). Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of *Arabidopsis thaliana* roots. *Plant Physiol.* 110, 1413–1429.
- Fu, C., Mielenz, J. R., Xiao, X., Ge, Y., Hamilton, C. Y., Rodriguez, M., et al. (2011). Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3803–3808. doi:10.1073/pnas.1100310108
- Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., and Chai, W. (2002). Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nat. Biotechnol.* 20, 1011–1017. doi:10.1038/nbt735
- Galbe, M., and Zacchi, G. (2007). “Pretreatment of lignocellulosic materials for efficient bioethanol production,” in *Biofuels*, Vol. 108, ed. Olsson L. (Berlin: Springer Verlag), 41–65.
- Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., et al. (1999). The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochem. J.* 342, 473–480. doi:10.1042/0264-6021:3420473
- Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100. doi:10.1126/science.1068275
- Gomez, L. D., Steele-King, C. G., and McQueen-Mason, S. J. (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytol.* 178, 473–485. doi:10.1111/j.1469-8137.2008.02422.x
- Gunnarsson, L. C., Zhou, Q., Montanier, C., Karlsson, E. N., Brumer, H., and Ohlin, M. (2006). Engineered xyloglucan specificity in a carbohydrate-binding module. *Glycobiology* 16, 1171–1180. doi:10.1093/glycob/cwl038
- Handakumbura, P. P., and Hazen, S. P. (2012). Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with *Arabidopsis thaliana*. *Front. Plant Sci.* 3:74. doi:10.3389/fpls.2012.00074
- Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15293–15298. doi:10.1073/pnas.1005732107
- Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi:10.1126/science.1137016
- Izquierdo, J. A., Pattathil, S., Guseva, A., Hahn, M. G., and Lynd, L. R. (2014). Comparative analysis of the ability of *Clostridium clariflavum* strains and *Clostridium thermocellum* to utilize hemicellulose and untreated plant material. *Biotechnol. Biofuels* 7, 136. doi:10.1186/s13068-014-0136-4
- Joseleau, J.-P., Faix, O., Kuroda, K.-I., and Ruel, K. (2004). A polyclonal antibody directed against syringylpropane epitopes of native lignins. *C. R. Biol.* 327, 809–815. doi:10.1016/j.crv.2004.06.003
- Joseleau, J.-P., and Ruel, K. (1997). Study of lignification by noninvasive techniques in growing maize internodes – an investigation by Fourier transform infrared cross-polarization magic angle spinning ^{13}C -nuclear magnetic resonance spectroscopy and immunocytochemical transmission electron microscopy. *Plant Physiol.* 114, 1123–1133.
- Joshi, C. P., Bhandari, S., Ranjan, P., Kalluri, U. C., Liang, X., Fujino, T., et al. (2004). Genomics of cellulose biosynthesis in poplars. *New Phytol.* 164, 53–61. doi:10.1111/j.1469-8137.2004.01155.x
- Joshi, C. P., Thammannagowda, S., Fujino, T., Gou, J. Q., Avci, U., Haigler, C. H., et al. (2011). Perturbation of wood cellulose synthesis causes pleiotropic effects in transgenic aspen. *Mol. Plant*. 4, 331–345. doi:10.1093/mp/ssq081
- Kataeva, I., Foston, M. B., Yang, S.-J., Pattathil, S., Biswal, A. K., Poole, I. F. L., et al. (2013). Carbohydrate and lignin are simultaneously solubilized from untreated switchgrass by microbial action at high temperature. *Energy Environ. Sci.* 6, 2186–2195. doi:10.1039/c3ee40932e
- Kim, Y. S., and Koh, H. B. (1997). Immuno electron microscopic study on the origin of milled wood lignin. *Holzforchung* 51, 411–413.
- Kiyoto, S., Yoshinaga, A., Tanaka, N., Wada, M., Kamitakahara, H., and Takabe, K. (2013). Immunolocalization of 8-5' and 8-8' linked structures of lignin in cell walls of *Chamaecyparis obtusa* using monoclonal antibodies. *Planta* 237, 705–715. doi:10.1007/s00425-012-1784-x
- Knox, J. P. (1997). The use of antibodies to study the architecture and developmental regulation of plant cell walls. *Int. Rev. Cytol.* 171, 79–120. doi:10.1016/S0074-7696(08)62586-3
- Knox, J. P. (2008). Revealing the structural and functional diversity of plant cell walls. *Curr. Opin. Plant Biol.* 11, 308–313. doi:10.1016/j.pbi.2008.03.001
- Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* 19, 1855–1860. doi:10.1101/gad.1331305

- Kukkola, E. M., Koutaniemi, S., Gustafsson, M., Karhunen, P., Ruel, K., Lundell, T. K., et al. (2003). Localization of dibenzodioxocin substructures in lignifying Norway spruce xylem by transmission electron microscopy-immunogold labeling. *Planta* 217, 229–237. doi:10.1007/s00425-003-0983-x
- Kukkola, E. M., Koutaniemi, S., Pollanen, E., Gustafsson, M., Karhunen, P., Lundell, T. K., et al. (2004). The dibenzodioxocin lignin substructure is abundant in the inner part of the secondary wall in Norway spruce and silver birch xylem. *Planta* 218, 497–500. doi:10.1007/s00425-003-1107-3
- Kulkarni, A. R., Pattathil, S., Hahn, M. G., York, W. S., and O'Neill, M. A. (2012). Comparison of arabinoxylan structure in bioenergy and model grasses. *Ind. Biotechnol.* 8, 222–229. doi:10.1089/ind.2012.0014
- Lee, C., O'Neill, M. A., Tsumuraya, Y., Darvill, A. G., and Ye, Z.-H. (2007). The *irregular xylem9* mutant is deficient in xylan xylosyltransferase activity. *Plant Cell Physiol.* 48, 1624–1634. doi:10.1093/pcp/pcm135
- Lee, C., Teng, Q., Zhong, R., and Ye, Z.-H. (2011a). Molecular dissection of xylan biosynthesis during wood formation in poplar. *Mol. Plant.* 4, 730–747. doi:10.1093/mp/ssr035
- Lee, K. J. D., Marcus, S. E., and Knox, J. P. (2011b). Cell wall biology: perspectives from cell wall imaging. *Mol. Plant.* 4, 212–219. doi:10.1093/mp/ssq075
- Lee, K. J., and Knox, J. P. (2014). Resin embedding, sectioning, and immunocytochemical analyses of plant cell walls in hard tissues. *Methods Mol. Biol.* 1080, 41–52. doi:10.1007/978-1-62703-643-6_3
- Lee, S. J., Warnick, T. A., Pattathil, S., Alvelo-Maurosa, J. G., Serapiglia, M. J., McCormick, H., et al. (2012). Biological conversion assay using *Clostridium phytofermentans* to estimate plant feedstock quality. *Biotechnol. Biofuels* 5, 5. doi:10.1186/1754-6834-5-5
- Lerouxel, O., Choo, T. S., Séveno, M., Usadel, B., Faye, L., Lerouge, P., et al. (2002). Rapid structural phenotyping of plant cell wall mutants by enzymatic oligosaccharide fingerprinting. *Plant Physiol.* 130, 1754–1763. doi:10.1104/pp.011965
- Li, M., Pattathil, S., Hahn, M. G., and Hodge, D. B. (2014). Identification of features associated with plant cell wall recalcitrance to pretreatment by alkaline hydrogen peroxide in diverse bioenergy feedstocks using glycome profiling. *RSC Adv.* 4, 17282–17292. doi:10.1039/C4RA00824C
- Liang, Y., Pattathil, S., Xu, W.-L., Basu, D., Venetos, A., Faik, A., et al. (2013). Biochemical and physiological characterization of *fut4* and *fut6* mutants defective in arabinogalactan-protein fucosylation in *Arabidopsis*. *J. Exp. Bot.* 64, 5537–5551. doi:10.1093/jxb/ert321
- Liepmann, A. H., Wightman, R., Geshi, N., Turner, S. R., and Scheller, H. V. (2010). *Arabidopsis* – a powerful model system for plant cell wall research. *Plant J.* 61, 1107–1121. doi:10.1111/j.1365-313X.2010.04161.x
- Loqué, D., Scheller, H. V., and Pauly, M. (2015). Engineering of plant cell walls for enhanced biofuel production. *Curr. Opin. Plant Biol.* 25, 151–161. doi:10.1016/j.pbi.2015.05.018
- Marcus, S. E., Blake, A. W., Benians, T. A. S., Lee, K. J. D., Poyser, C., Donaldson, L., et al. (2010). Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J.* 64, 191–203. doi:10.1111/j.1365-313X.2010.04319.x
- McCann, M. C., and Knox, J. P. (2011). “Plant cell wall biology: polysaccharides in architectural and developmental contexts,” in *Plant Polysaccharides: Biosynthesis and Bioengineering*, Vol. 41, ed. Ulvskov P. (Chichester: Wiley-Blackwell), 343–366.
- McCann, M. C., and Roberts, K. (1991). “Architecture of the primary cell wall,” in *Cytoskeletal Basis of Plant Growth and Form*, ed. Lloyd C. W. (London: Academic Press Ltd.), 109–129.
- McCarthy, R. L., Zhong, R., and Ye, Z.-H. (2009). MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell Physiol.* 50, 1950–1964. doi:10.1093/pcp/pcm139
- McCartney, L., Blake, A. W., Flint, J., Bolam, D. N., Boraston, A. B., Gilbert, H. J., et al. (2006). Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4765–4770. doi:10.1073/pnas.0508887103
- McCartney, L., Marcus, S. E., and Knox, J. P. (2005). Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *J. Histochem. Cytochem.* 53, 543–546. doi:10.1369/jhc.4B6578.2005
- McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresour. Technol.* 83, 37–46. doi:10.1016/S0960-8524(01)00118-3
- McNeil, M., Darvill, A., Fry, S. C., and Albersheim, P. (1984). Structure and function of the primary cell walls of plants. *Annu. Rev. Biochem.* 53, 625–663. doi:10.1146/annurev.bi.53.070184.003205
- Meikle, P. J., Bonig, I., Hoogenraad, N. J., Clarke, A. E., and Stone, B. A. (1991). The location of (1,3)- β -glucans in the walls of pollen tubes of *Nicotiana glauca* using a (1,3)- β -glucan-specific monoclonal antibody. *Planta* 185, 1–8. doi:10.1007/BF00194507
- Meikle, P. J., Hoogenraad, N. J., Bonig, I., Clarke, A. E., and Stone, B. A. (1994). A (1,3;1,4)- β -glucan-specific monoclonal antibody and its use in the quantitation and immunocytochemical location of (1,3;1,4)- β -glucans. *Plant J.* 5, 1–9. doi:10.1046/j.1365-313X.1994.5010001.x
- Mohnen, D., Bar-Peled, M., and Somerville, C. (2008). “Cell wall polysaccharide synthesis,” in *Biomass Recalcitrance – Deconstructing the Plant Cell Wall for Bioenergy*, ed. Himmel D. E. (Oxford: Blackwell Publishing Ltd.), 94–187.
- Moller, I., Marcus, S. E., Haeger, A., Verhertbruggen, Y., Verhoeve, R., Schols, H., et al. (2008). High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles. *Glycoconj. J.* 25, 37–48. doi:10.1007/s10719-007-9059-7
- Moller, I., Sorensen, I., Bernal, A. J., Blaukopf, C., Lee, K., Obro, J., et al. (2007). High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant J.* 50, 1118–1128. doi:10.1111/j.1365-313X.2007.03114.x
- Oikawa, A., Joshi, H. J., Rennie, E. A., Ebert, B., Manisseri, C., Heazlewood, J. L., et al. (2010). An integrative approach to the identification of *Arabidopsis* and rice genes involved in xylan and secondary wall development. *PLoS ONE* 5:e15481. doi:10.1371/journal.pone.0015481
- Pattathil, S., Avci, U., Baldwin, D., Swennes, A. G., McGill, J. A., Popper, Z., et al. (2010). A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol.* 153, 514–525. doi:10.1104/pp.109.151985
- Pattathil, S., Avci, U., Miller, J. S., and Hahn, M. G. (2012a). Immunological approaches to plant cell wall and biomass characterization: glycome profiling. *Methods Mol. Biol.* 908, 61–72. doi:10.1007/978-1-61779-956-3_6
- Pattathil, S., Saffold, T., Gallego-Giraldo, L., O'Neill, M., York, W. S., Dixon, R. A., et al. (2012b). Changes in cell wall carbohydrate extractability are correlated with reduced recalcitrance of HCT downregulated alfalfa biomass. *Ind. Biotechnol.* 8, 217–221. doi:10.1089/ind.2012.0013
- Pattathil, S., Hahn, M. G., Dale, B. E., and Chundawat, S. P. S. (2015). Insights into plant cell wall structure, architecture, and integrity using glycome profiling of native and AFEX™-pre-treated biomass. *J. Exp. Bot.* 66, 4279–4294. doi:10.1093/jxb/erv107
- Patten, A. M., Jourdes, M., Cardenas, C. L., Laskar, D. D., Nakazawa, Y., Chung, B.-Y., et al. (2010). Probing native lignin macromolecular configuration in *Arabidopsis thaliana* in specific cell wall types: further insights into limited substrate degeneracy and assembly of the lignins of *ref8*, *fah* 1-2 and C4H:F5H lines. *Mol. Biosyst.* 6, 499–515. doi:10.1039/b819206e
- Pauly, M., and Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54, 559–568. doi:10.1111/j.1365-313X.2008.03463.x
- Pedersen, H. L., Fangel, J. U., McCleary, B., Ruzanski, C., Rydahl, M. G., Ralet, M.-C., et al. (2012). Versatile high resolution oligosaccharide microarrays for plant glycobiology and cell wall research. *J. Biol. Chem.* 287, 39429–39438. doi:10.1074/jbc.M112.396598
- Peña, M. J., Darvill, A. G., Eberhard, S., York, W. S., and O'Neill, M. A. (2008). Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* 18, 891–904. doi:10.1093/glycob/cwn078
- Peña, M. J., Zhong, R., Zhou, G.-K., Richardson, E. A., O'Neill, M. A., Darvill, A. G., et al. (2007). *Arabidopsis irregular xylem8* and *irregular xylem9*: implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19, 549–563. doi:10.1105/tpc.106.049320
- Petersen, P. D., Lau, J., Ebert, B., Yang, F., Verhertbruggen, Y., Kim, J. S., et al. (2012). Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnol. Biofuels* 5, 84. doi:10.1186/1754-6834-5-84
- Popper, Z. A. (2008). Evolution and diversity of green plant cell walls. *Curr. Opin. Plant Biol.* 11, 286–292. doi:10.1016/j.pbi.2008.02.012
- Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G. T., Tuohy, M. G., et al. (2011). Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* 62, 567–588. doi:10.1146/annurev-arplant-042110-103809

- Puhlmann, J., Bucheli, E., Swain, M. J., Dunning, N., Albersheim, P., Darvill, A. G., et al. (1994). Generation of monoclonal antibodies against plant cell wall polysaccharides. I. Characterization of a monoclonal antibody to a terminal α -(1,2)-linked fucosyl-containing epitope. *Plant Physiol.* 104, 699–710. doi:10.1104/pp.104.2.699
- Ralet, M.-C., Tranquet, O., Poulain, D., Moïse, A., and Guillon, F. (2010). Monoclonal antibodies to rhamnogalacturonan I backbone. *Planta* 231, 1373–1383. doi:10.1007/s00425-010-1116-y
- Reinikainen, T., Ruohonen, L., Nevenen, T., Laaksonen, L., Kraulis, P., Jones, T. A., et al. (1992). Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase-I. *Proteins* 14, 475–482. doi:10.1002/prot.340140408
- Ruel, K., Faix, O., and Joseleau, J.-P. (1994). New immunogold probes for studying the distribution of the different lignin types during plant cell wall biogenesis. *J. Trace Microprobe Tech.* 12, 247–265.
- Ruel, K., Montiel, M.-D., Goujon, T., Jouanin, L., Burlat, V., and Joseleau, J.-P. (2002). Interrelation between lignin deposition and polysaccharide matrices during the assembly of plant cell walls. *Plant Biol.* 4, 2–8. doi:10.1055/s-2002-20429
- Scheller, H. V., Singh, S., Blanch, H., and Keasling, J. D. (2010). The Joint BioEnergy Institute (JBEI): developing new biofuels by overcoming biomass recalcitrance. *Bioenerg. Res.* 3, 105–107. doi:10.1007/s12155-010-9086-2
- Schmidt, D., Schuhmacher, F., Geissner, A., Seeberger, P. H., and Pfengle, F. (2015). Automated synthesis of arabinoxylan-oligosaccharides enables characterization of antibodies that recognize plant cell wall glycans. *Chemistry* 21, 5709–5713. doi:10.1002/chem.201500065
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi:10.1126/science.1178534
- Selvendran, R. R., Davies, A. M. C., and Tidder, E. (1975). Cell wall glycoproteins and polysaccharides of mature runner beans. *Phytochemistry* 14, 2169–2174. doi:10.1016/S0031-9422(00)91094-X
- Selvendran, R. R., and O'Neill, M. A. (1987). "Isolation and analysis of cell walls from plant material," in *Methods of Biochemical Analysis*, 32 Edn, ed. Glick D. (New York, NY: John Wiley & Sons, Inc), 25–153.
- Shen, H., Poovaiah, C. R., Ziebell, A., Tschaplinski, T. J., Pattathil, S., Gjersing, E., et al. (2013). Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol. Biofuels* 6, 71. doi:10.1186/1754-6834-6-71
- Shoseyov, O., Shani, Z., and Levy, I. (2006). Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283–295. doi:10.1128/MMBR.00028-05
- Sluiter, J. B., Ruiz, R. O., Scarlata, C. J., Sluiter, A. D., and Templeton, D. W. (2010). Compositional analysis of lignocellulosic feedstocks. I. Review and description of methods. *J. Agric. Food Chem.* 58, 9043–9053. doi:10.1021/jf1008023
- Socha, A. M., Parthasarathi, R., Shi, J., Pattathil, S., Whyte, D., Bergeron, M., et al. (2014). Efficient biomass pretreatment using ionic liquids derived from lignin and hemicellulose. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3587–E3595. doi:10.1073/pnas.1405685111
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., et al. (2004). Toward a systems approach to understanding plant cell walls. *Science* 306, 2206–2211. doi:10.1126/science.1102765
- Somerville, C., Youngs, C., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi:10.1126/science.1189268
- Sørensen, I., Pettolino, F. A., Bacic, A., Ralph, J., Lu, F., O'Neill, M. A., et al. (2011). The charophyte green algae provide insights into the early origins of plant cell walls. *Plant J.* 68, 201–211. doi:10.1111/j.1365-3113.2011.04686.x
- Steffan, W., Kováč, P., Albersheim, P., Darvill, A. G., and Hahn, M. G. (1995). Characterization of a monoclonal antibody that recognizes an arabinosylated (1,6)- β -D-galactan epitope in plant complex carbohydrates. *Carbohydr. Res.* 275, 295–307. doi:10.1016/0008-6215(95)00174-R
- Studer, M. H., DeMartini, J. D., Davis, M. F., Sykes, R. W., Davison, B., Keller, M., et al. (2011). Lignin content in natural *Populus* variants affects sugar release. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6300–6305. doi:10.1073/pnas.1009252108
- Sweet, D. P., Albersheim, P., and Shapiro, R. H. (1975a). Partially ethylated alditol acetates as derivatives for elucidation of the glycosyl linkage-composition of polysaccharides. *Carbohydr. Res.* 40, 199–216. doi:10.1016/S0008-6215(00)82603-8
- Sweet, D. P., Shapiro, R. H., and Albersheim, P. (1975b). Quantitative analysis by various G.L.C. response-factor theories for partially methylated and partially ethylated alditol acetates. *Carbohydr. Res.* 40, 217–225. doi:10.1016/S0008-6215(00)82604-X
- Sweet, D. P., Shapiro, R. H., and Albersheim, P. (1974). The mass spectral fragmentation of partially ethylated alditol acetates, a derivative used in determining the glycosyl linkage composition of polysaccharides. *Biomed. Mass Spectrom.* 1, 263–268. doi:10.1002/bms.1200010410
- Szabó, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., et al. (2001). Structure of a family 15 carbohydrate-binding module in complex with xylopentaose – evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* 276, 49061–49065. doi:10.1074/jbc.M109558200
- Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., et al. (2013). An *Arabidopsis* cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* 25, 270–287. doi:10.1105/tpc.112.107334
- Taylor, N. G., Gardiner, J. C., Whiteman, R., and Turner, S. R. (2004). Cellulose synthesis in the *Arabidopsis* secondary cell wall. *Cellulose* 11, 329–338. doi:10.1023/B:CELL.0000046405.11326.a8
- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi:10.1038/35048692
- The International Brachypodium Initiative. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi:10.1038/nature08747
- Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 1. Binding specificity and calorimetric analysis. *Biochemistry* 35, 13885–13894. doi:10.1021/bi961185i
- Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., et al. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* 15, 5739–5751.
- Trajano, H. L., Pattathil, S., Tomkins, B. A., Tschaplinski, T. J., Hahn, M. G., Van Berkel, G. J., et al. (2015). Xylan hydrolysis in *Populus trichocarpa* x *P. deltoides* and model substrates during hydrothermal pretreatment. *Bioresour. Technol.* 179, 202–210. doi:10.1016/j.biortech.2014.11.090
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi:10.1126/science.1128691
- Urbanowicz, B. R., Pena, M. J., Ratnaparkhe, S., Avci, U., Backe, J., Steet, H. F., et al. (2012). 4-O-methylation of glucuronic acid in *Arabidopsis* glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14253–14258. doi:10.1073/pnas.1208097109
- Valdivia, E. R., Herrera, M. T., Gianzo, C., Fidalgo, J., Revilla, G., Zarra, I., et al. (2013). Regulation of secondary wall synthesis and cell death by NAC transcription factors in the monocot *Brachypodium distachyon*. *J. Exp. Bot.* 64, 1333–1343. doi:10.1093/jxb/ers394
- Verhertbruggen, Y., Marcus, S. E., Haeger, A., Ordaz-Ortiz, J. J., and Knox, J. P. (2009). An extended set of monoclonal antibodies to pectic homogalacturonan. *Carbohydr. Res.* 344, 1858–1862. doi:10.1016/j.carres.2008.11.010
- Vogel, J. (2008). Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* 11, 301–307. doi:10.1016/j.pbi.2008.03.002
- von Schantz, L., Gullföt, F., Scheer, S., Filonova, L., Gunnarsson, L. C., Flint, J. E., et al. (2009). Affinity maturation generates greatly improved xyloglucan-specific carbohydrate binding modules. *BMC Biotechnol.* 9:92. doi:10.1186/1472-6750-9-92
- von Schantz, L., Håkansson, M., Logan, D. T., Walse, B., Österlin, J., Nordberg-Karlsson, E., et al. (2012). Structural basis for carbohydrate-binding specificity – a comparative assessment of two engineered carbohydrate-binding modules. *Glycobiology* 22, 948–961. doi:10.1093/glycob/cws063
- Voxeur, A., Wang, Y., and Sibout, R. (2015). Lignification: different mechanisms for a versatile polymer. *Curr. Opin. Plant Biol.* 23, 83–90. doi:10.1016/j.pbi.2014.11.006
- Wallace, I., and Anderson, C. T. (2012). Small molecule probes for plant cell wall polysaccharide imaging. *Front. Plant Sci.* 3:89. doi:10.3389/fpls.2012.00089
- Wang, D. N., Liu, S., Trummer, B. J., Deng, C., and Wang, A. L. (2002). Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat. Biotechnol.* 20, 275–281. doi:10.1038/nbt0302-275

- Wang, H., Avci, U., Nakashima, J., Hahn, M. G., Chen, F., and Dixon, R. A. (2010). Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22338–22343. doi:10.1073/pnas.1016436107
- Wang, H.-Z., and Dixon, R. A. (2012). On-off switches for secondary cell wall biosynthesis. *Mol. Plant* 5, 297–303. doi:10.1093/mp/ssr098
- Weng, J.-K., and Chapple, C. (2010). The origin and evolution of lignin biosynthesis. *New Phytol.* 187, 273–285. doi:10.1111/j.1469-8137.2010.03327.x
- Willats, W. G. T., Limberg, G., Buchholt, H. C., Van Alebeek, G.-J., Benen, J., Christensen, T. M. I. E., et al. (2000a). Analysis of pectic epitopes recognised by hybridoma and phage display monoclonal antibodies using defined oligosaccharides, polysaccharides, and enzymatic degradation. *Carbohydr. Res.* 327, 309–320. doi:10.1016/S0008-6215(00)00039-2
- Willats, W. G. T., Steele-King, C. G., McCartney, L., Orfila, C., Marcus, S. E., and Knox, J. P. (2000b). Making and using antibody probes to study plant cell walls. *Plant Physiol. Biochem.* 38, 27–36. doi:10.1016/S0981-9428(00)00170-4
- Willats, W. G. T., Rasmussen, S. E., Kristensen, T., Mikkelsen, J. D., and Knox, J. P. (2002). Sugar-coated microarrays: a novel slide surface for the high-throughput analysis of glycans. *Proteomics* 2, 1666–1671. doi:10.1002/1615-9861(200212)2:12<1666::AID-PROT1666>3.0.CO;2-E
- Wu, A.-M., Hörnblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., et al. (2010). Analysis of the *Arabidopsis* *IRX9/IRX9-L* and *IRX14/IRX14-L* pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiol.* 153, 542–554. doi:10.1104/pp.110.154971
- Yang, F., Mitra, P., Zhang, L., Prak, L., Verhertbruggen, Y., Kim, J.-S., et al. (2013). Engineering secondary cell wall deposition in plants. *Plant Biotechnol. J.* 11, 325–335. doi:10.1111/pbi.12016
- Ye, Z.-H., York, W. S., and Darvill, A. G. (2006). Important new players in secondary wall synthesis. *Trends Plant Sci.* 11, 162–164. doi:10.1016/j.tplants.2006.02.001
- Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi:10.1038/nature10625
- Yu, J., Hu, S. N., Wang, J., Wong, G. K. S., Li, S. G., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92. doi:10.1126/science.1068037
- Yu, L., Shi, D., Li, J., Kong, Y., Yu, Y., Chai, G., et al. (2014). CELLULOSE SYNTHASE-LIKE A2, a glucomannan synthase, is involved in maintaining adherent mucilage structure in *Arabidopsis* seed. *Plant Physiol.* 164, 1842–1856. doi:10.1104/pp.114.236596
- Zhang, X., Rogowski, A., Zhao, L., Hahn, M. G., Avci, U., Knox, J. P., et al. (2014). Understanding how the complex molecular architecture of mannan-degrading hydrolases contributes to plant cell wall degradation. *J. Biol. Chem.* 289, 2002–2012. doi:10.1074/jbc.M113.527770
- Zhao, Q., Tobimatsu, Y., Zhou, R., Pattathil, S., Gallego-Giraldo, L., Fu, C., et al. (2013). Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13660–13665. doi:10.1073/pnas.1312234110
- Zhong, R., Demura, T., and Ye, Z.-H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* 18, 3158–3170. doi:10.1105/tpc.106.047399
- Zhong, R., Richardson, E. A., and Ye, Z.-H. (2007a). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *Plant Cell* 19, 2776–2792. doi:10.1105/tpc.107.053678
- Zhong, R., Richardson, E. A., and Ye, Z.-H. (2007b). Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* 225, 1603–1611. doi:10.1007/s00425-007-0498-y
- Zhu, X., Pattathil, S., Mazumder, K., Brehm, A., Hahn, M. G., Dinesh-Kumar, S. P., et al. (2010). Virus-induced gene silencing offers a functional genomics platform for studying plant cell wall formation. *Mol. Plant* 3, 818–833. doi:10.1093/mp/ssq023

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Pattathil, Avci, Zhang, Cardenas and Hahn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Relationships between biomass composition and liquid products formed via pyrolysis

Fan Lin^{1†}, Christopher L. Waters^{2†}, Richard G. Mallinson², Lance L. Lobban² and Laura E. Bartley^{1*}

¹ Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA, ² School of Chemical, Biological, and Materials Engineering, University of Oklahoma, Norman, OK, USA

OPEN ACCESS

Edited by:

Jason Lupoi,
Joint BioEnergy Institute, USA;
University of Queensland, Australia

Reviewed by:

Suyin Gan,
The University of Nottingham
Malaysia Campus, Malaysia
Xu Fang,
Shandong University, China

*Correspondence:

Laura E. Bartley
lbartley@ou.edu

[†]Fan Lin and Christopher L. Waters
have contributed equally to this work.

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Energy Research

Received: 22 July 2015

Accepted: 29 September 2015

Published: 21 October 2015

Citation:

Lin F, Waters CL, Mallinson RG,
Lobban LL and Bartley LE (2015)
Relationships between biomass
composition and liquid products
formed via pyrolysis.
Front. Energy Res. 3:45.
doi: 10.3389/fenrg.2015.00045

Thermal conversion of biomass is a rapid, low-cost way to produce a dense liquid product, known as bio-oil, that can be refined to transportation fuels. However, utilization of bio-oil is challenging due to its chemical complexity, acidity, and instability – all results of the intricate nature of biomass. A clear understanding of how biomass properties impact yield and composition of thermal products will provide guidance to optimize both biomass and conditions for thermal conversion. To aid elucidation of these associations, we first describe biomass polymers, including phenolics, polysaccharides, acetyl groups, and inorganic ions, and the chemical interactions among them. We then discuss evidence for three roles (i.e., models) for biomass components in the formation of liquid pyrolysis products: (1) as direct sources, (2) as catalysts, and (3) as indirect factors whereby chemical interactions among components and/or cell wall structural features impact thermal conversion products. We highlight associations that might be utilized to optimize biomass content prior to pyrolysis, though a more detailed characterization is required to understand indirect effects. In combination with high-throughput biomass characterization techniques, this knowledge will enable identification of biomass particularly suited for biofuel production and can also guide genetic engineering of bioenergy crops to improve biomass features.

Keywords: thermochemical conversion, plant biomass, bio-oil, lignin, polysaccharides, cell wall, fast pyrolysis, minerals

Biomass can be a renewable and sustainable source of transportation fuels not associated with fossil CO₂ release. Numerous studies highlight the advantages of displacing petroleum fuels with industrial production of liquid fuels from thermochemical conversion of biomass (Bridgwater et al., 1999; Perlack et al., 2005; Mohan et al., 2006; NSF, 2008). Thermochemical conversion entails heating of biomass in an anoxic environment; condensation of organic liquid products, known as bio-oil; and subsequent treatment of the products with catalysts to create liquid fuels, i.e., refined bio-oil, similar to petroleum-derived gasoline or diesel. This is in contrast to biochemical conversion, which utilizes enzymes to release sugars followed by microbial production of ethanol or other fuel molecules (Somerville, 2007; Youngs and Somerville, 2012). Relative to biochemical approaches, thermal conversion has the potential to make use of all carbon (C)-containing biomass components, would allow society to retain existing infrastructure associated with liquid hydrocarbon fuels, and, due to the rapidity of the process, may reduce production costs by permitting scalability and distribution of production (Huber et al., 2006; Mettler et al., 2012). For both

thermochemical and biochemical biofuels, lowering processing costs and improving fuel yields per hectare are major engineering challenges that hinder economic viability. Thermochemical fuel production also faces challenges related to maintaining a high C-yield while obtaining a fungible fuel. We posit that this latter challenge might be addressed by understanding the relationships between biomass composition and bio-oil components and using this information to alter biomass through genetic, chemical, or thermal means.

THERMAL CONVERSION CHALLENGES

Two types of pyrolysis have been developed: fast pyrolysis and slow pyrolysis. Slow pyrolysis is usually performed over several hours and has a high solid yield, and as such has little relevance for liquid fuels production. Fast pyrolysis, however, is typically performed quickly, in seconds, at temperatures between 400 and 600°C and decomposes most of the solid biomass into a volatile mixture of various organic molecules, water, and CO/CO₂. Pyrolysis oil or bio-oil constitutes the condensable portion of this vapor. Non-condensable components (primarily CO₂ and CO) and a mineral-rich solid (char) are other product classes that will not be addressed here, except in that they detract from the overall C-yield of raw and refined bio-oil. Bio-oil comprises water (15–30%) plus compounds from several chemical families including the following (Table 1): organic acids, light (C1–C3) oxygenates, furan and furan derivatives, phenolic species with various methyl and methoxy substituents, pyrones, and sugar derivatives like levoglucosan (Faix et al., 1991a,b). Bio-oil's chemically complex nature prohibits its direct use in combustion applications or petroleum refining. The reasons for this include low heating value; ignition difficulty; high chemical reactivity, which results in oligomerization and polymerization over time and upon heating, prohibiting distillative separation (Oasmaa and Czernik, 1999; Demirbas, 2011; Patwardhan et al., 2011a); immiscibility with petroleum; and high corrosivity (Oasmaa and Czernik, 1999). Many of these features are associated with the

high oxygen content of biomass and the resulting bio-oil, relative to fossil fuels.

In order to obtain desirable fuel properties and allow integration with the existing transportation fuels infrastructure (gasoline and diesel engines), the bio-oil must be chemically converted to reduce the undesirable characteristics mentioned above. Catalytic upgrading is typically used to refine bio-oil, improving its stability and making it an acceptable liquid fuel. The simplest method is hydrotreating or hydrodeoxygenation, which removes oxygen via catalytic hydrogenation (Furimsky, 2000), decreasing both the chemical reactivity and corrosivity. However, this process converts any C1–C5 oxygenates, representing as much as half of the carbon in bio-oil, to C1–C5 hydrocarbons that are too volatile for liquid fuels (Resasco, 2011). Another straightforward approach is to “crack” the pyrolysis vapors using acidic zeolite catalysts into light olefins and aromatic hydrocarbons (primarily benzene, toluene, and *o/m/p*-xylene) (Bridgwater, 1994; Carlson et al., 2008, 2009). This approach is appealing because of the lack of an external H₂ requirement and the simplicity of the product streams. Furthermore, since zeolite cracking is widely used in traditional petroleum refining/valorization (Wan et al., 2015), other advantages are the product compatibility with existing refinery infrastructure and the maturity of the process (Wan et al., 2015). However, zeolite cracking is crippled by poor usable carbon yield due to the high amounts of coke, CO, and CO₂ formed during the catalytic process (Carlson et al., 2008) and the concomitant rapid catalyst deactivation. Additionally, further catalytic oligomerization and reforming for olefins and aromatics, respectively, are needed to make these products suitable for addition to refinery fuel product streams, increasing the process costs and further reducing overall carbon yield.

More advanced strategies propose to use reactions such as ketonization, condensation, alkylation, and others to retain a higher fraction of the biomass carbon into liquid fuel-range molecules (Zhu et al., 2011; Zapata et al., 2012; Pham et al., 2013; Gonzalez-Borja and Resasco, 2015). However, catalytic upgrading of any one family of compounds (e.g., light oxygenates) typically requires a catalyst and reaction conditions different than those required for another family of compounds (e.g., substituted phenolics). Moreover, catalysts used for upgrading one family of compounds may be ill-suited for other families, either facilitating undesirable reactions (breaking C–C bonds unnecessarily or increasing H:C ratios above the 2:1 optimum) or undergoing rapid deactivation due to reactions with other non-targeted bio-oil oxygenates. These upgrading challenges suggest the desirability of thermal conversion producing more selective product streams, i.e., each stream comprising fewer families of chemical compounds. Developing such thermal conversion processes would be aided by clearer knowledge of the relationship between biomass composition and thermal conversion products.

BIOMASS COMPOSITION AND CHEMICAL STRUCTURES

Recent reviews have addressed the general relationships between biomass composition and thermal products, such as

TABLE 1 | The percentage ranges and categories of major bio-oil components.

Category	Major components	Wet weight (%)
Light oxygenates	Glycolaldehyde, acetol	3–26
Organic acids	Acetic acid, formic acid, propanoic acid	2–27
Aldehydes	Acetaldehyde, formaldehyde, ethanedial	3–18
Sugars	1,6-Anhydroglucose (levoglucosan)	5–14
Phenols	Phenol, catechol (di-OH benzene), methyl phenol, dimethyl phenol	3–13
Guaiacols	Isoeugenol, eugenol, 4-methylguaiacol	3–15
Furans	Furfural, hydroxymethyl furfural, furfural	2–11
Syringols	2,6-Dimethoxy phenol, syringaldehyde, propyl syringol	2–9
Ketones	Acetone	4–6
Alcohols	Methanol, ethylene glycol, ethanol	2–6
Esters	Methyl formate, butyrolactone, methylfuranone	<1–3

Source: Huber et al. (2006).

increasing the content of phenolics relative to carbohydrates to reduce the oxygen content of bio-oil (Tanger et al., 2013). Here, we provide a more detailed description of the chemical structure and interactions among major cell wall components to aid in understanding more subtle relationships between biomass and bio-oil content. Biomass consists of cell walls that establish the structure of the plant and, to a lesser extent, non-structural components (Table 2). Cell walls determine the shape of leaves and stems and the cells that compose them and consist of cellulose, hemicellulose, lignin, as well as structural proteins and wall-associated mineral components (O'Neill and York, 2009; Vogel et al., 2011; Tanger et al., 2013). Non-structural components include sugars, proteins, and additional minerals (O'Neill and York, 2009; Vogel et al., 2011; Tanger et al., 2013). For example, in switchgrass, an important potential bioenergy crop, dry biomass consists of ~70% cell walls, 9% intrinsic water, 8% minerals, 6% proteins, and 5% non-structural sugars (Vogel et al., 2011). The relative fractions of different components, chemical linkages within and between polymers, and cellular patterning vary among plant species, organs, developmental stages, and growth conditions (Adler et al., 2006; El-Nashaar et al., 2009; Singh et al., 2012; Zhao et al., 2012). Here, we review the components of secondary cell walls, which are formed as plant growth ceases, as they constitute the majority of plant biomass (Pauly and Keegstra, 2008), and then discuss evidence for interactions among components. Table 2 lists the different major and minor components of biomass and the broad ranges of their representation within biomass for biofuel conversion.

TABLE 2 | The variation of biomass components among vascular plants including grasses, softwoods, and hardwoods.

Biomass component	Dry weight (%) ^a
Cellulose	15–49 ^b
Hemicellulose	12–50 ^{b,c,d}
Xylan	5–50 ^c
Mixed-linkage glucan	0–5 ^{c,e}
Xyloglucan	Minor ^c
Mannan (and galactoglucomannan)	0–30 ^{c,f}
Soluble (mainly sucrose)	9–67 ^{b,g}
Pectin	<0.1 ^h
Lignin	6–28 ^b
Ferulic acid and <i>p</i> -coumaric acid	<1.5 ^h
Protein	4–5 ^b
Ash (mainly silicate)	0.4–14.4 ^b
Intrinsic moisture	11–34 ⁱ

^aPercent mass composition of secondary cell walls.

^bPauly and Keegstra (2008).

^cScheller and Ulvskov (2010).

^dAs the highest percentage of xylan in Scheller and Ulvskov (2010) is higher than the highest percentage of hemicellulose in Pauly and Keegstra (2008), the highest percentage of hemicellulose is set to the highest percentage of xylan.

^eMLG is only abundant in grasses. The maximum percentage of MLG we are aware of is that of the mature rice stem after flowering (Vega-Sanchez et al., 2012).

^fGalactoglucomannan is only abundant in gymnosperm woods. Dicots and grasses possess <8% of mannan and galactoglucomannan (Scheller and Ulvskov, 2010).

^gThe high abundance of solubles is only for sorghum biomass. Other plants usually have less than 15% soluble content (Pauly and Keegstra, 2008).

^hVogel (2008).

ⁱMcKendry (2002).

Figure 1(1–3) shows the chemical structures and atom numbering of the most abundant cell wall monomeric species.

Cellulose and hemicellulose represent 15–49% and 12–50% of biomass by dry weight, respectively (Pauly and Keegstra, 2008; Vogel, 2008; Zhao et al., 2012). Cellulose is an unbranched homopolymer of >500 β -(1,4)-linked glucose units. In plant cell walls, cellulose is primarily in the form of crystalline microfibrils consisting of approximately 36 hydrogen-bonded cellulose chains, but also has amorphous regions (Somerville, 2006; Newman et al., 2013).

Hemicelluloses are typically branched polysaccharides substituted with various sugars and acyl groups. As discussed further in the Section “Evidence Relating Biomass Content and Bio-oil Composition,” the different sugar composition and linkages of hemicelluloses influence thermal products (Shafizadeh et al., 1972; Mante et al., 2014). The structure and composition of hemicellulosic polysaccharides differ depending on plant species classification, i.e., taxonomy. Major taxonomic divisions with relevance to bioenergy production are grasses, such as switchgrass and wheat; woody dicots, i.e., hardwoods, such as poplar; and woody gymnosperms, i.e., softwoods, such as pine. The most abundant grass hemicelluloses are mixed-linkage glucan (MLG) and glucuronoarabinoxylan (GAX) (Scheller and Ulvskov, 2010; Vega-Sanchez et al., 2013); the hemicelluloses of hardwood are primarily composed of glucuronoxylans (GX) but also contain a small amount of galactomannans (GM) (Pauly and Keegstra, 2008); and softwood hemicelluloses are largely galactoglucomannan (GGM) and GAXs (Scheller and Ulvskov, 2010). MLG is an unbranched glucose polymer similar to cellulose but containing both β -(1-3)- and β -(1-4)-linkages (Vega-Sanchez et al., 2012). MLG is nearly unique to the order Poales, which includes the grasses, but has also been found in horsetail (*Equisetum*). Its abundance in mature tissues and secondary cell walls has recently been recognized (Vega-Sanchez et al., 2013). Xylans consist of a β -(1-4)-linked xylose backbone with various substitutions. GXs are xylans substituted mostly by glucuronic acid and 4-O-methyl glucuronic acid through α -(1-2)-linkages. GAXs are not only substituted by glucuronic acid but also substituted by arabinofuranoses at the O-3, which can be further substituted by the phenylpropanoid acids, to form feruloyl- and *p*-coumaroyl esters linked at the O-5 (Scheller and Ulvskov, 2010). Acetyl groups are often attached to the O-3 of backbone xyloses but also attach to the O-2. Unlike xylans, which mainly consist of pentoses, mannans consist of hexoses like mannose, glucose, and galactose. GM and GGM have a β -(1-4)-linked backbone with mannose or a combination of glucose and mannose, respectively. Both GM and GGM can be acetylated and substituted by α -(1-6)-linked galactoses (Scheller and Ulvskov, 2010; Rodriguez-Gacio Mdel et al., 2012; Pauly et al., 2013). Relatively depleted in secondary walls, but rich in growing primary walls of dicot species, xyloglucan and pectins are two other polysaccharides in cell walls. Xyloglucan consists of β -(1-4)-linked glucose residues, modified by xylose and other sugar residues; and pectin is another branched or unbranched polymer that is rich in galacturonic acid, rhamnose, galactose, and several other monosaccharide residues (Somerville et al., 2004; Scheller and Ulvskov, 2010).

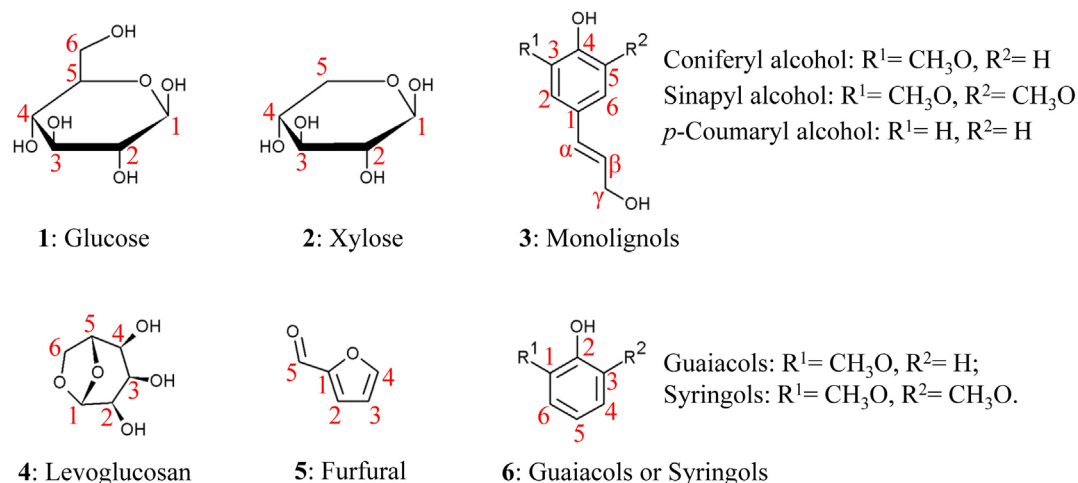


FIGURE 1 | Chemical structure of the major basic units of biomass polymers and related products.

Lignin is a cross-linked, heteropolyphenol mainly assembled from three monolignols – sinapyl (S), coniferyl (G), and *p*-coumaryl (H) alcohols. As waste products are often selected as biofuel feedstocks, it is also relevant to note that lignin derived from other monolignols such as caffeyl alcohol and 5-hydroxy-coniferyl have been found in the seedcoat of both monocots and dicots (Chen et al., 2012, 2013). Lignin structural heterogeneity and various types of incorporated groups can lead to a variety of different depolymerization reactions during pyrolysis (Kawamoto et al., 2007). Often traceable to the corresponding bio-oil components, the three major lignin units differ in the degree of methoxylation of their carbon ring. S-units are methoxylated at both O-3 and O-5 ring positions; G-unit have one methoxy group at the O-3 position; and H-units lack ring methoxy groups (3, Figure 1) (Boerjan et al., 2003). Lignin units undergo oxidative coupling in the cell wall to form many types of dimers, including β -O-4, β -5, β - β , 5-5, 5-O-4, and β -1, leaving other atoms free to further polymerize, which significantly increases the structural heterogeneity of lignin. Lignin units can also be esterified with *p*-coumaryl, *p*-hydroxybenzoyl, and acetyl groups, primarily at the γ position of terminal units (Petrik et al., 2014; Lu et al., 2015). Lignin compositions and the acylation groups vary among plant clades (Boerjan et al., 2003). Woody dicot lignins have G- and S-units and trace amount of H-units. Poplar wood, for example, has a G:S:H ratio of 55:45:1 (Vanholme et al., 2013). The lignin of many hardwoods is acylated by *p*-hydroxybenzoates (Lu et al., 2015) and acetyl groups in low amounts (Sarkanen et al., 1967). Biomass from other species, such as palms and kenaf, possess a high degree of lignin acetylation (Lu and Ralph, 2002). Grass lignins also contain G- and S-units with slightly higher amount of H-units than woody dicots. Wheat straw, for example, has a G:S:H ratio of 64:30:6 (Bule et al., 2013). Grass lignin possesses high levels of *p*-coumarate esters (Hatfield et al., 2008) and can also be etherified by triclin and ferulic acid (Ralph et al., 1995; Lan et al., 2015), as discussed further below. Woody gymnosperm

lignins are different from angiosperm lignins, being primarily composed of G-units and a lower amount of H-units (Boerjan et al., 2003).

Biomass also contains inorganic elements including Ca, K, Si, Mg, Al, S, Fe, P, Cl, and Na and some trace elements (<0.1%) such as Mn and Ti, according to ash analysis, formed by oxidation of biomass at 575°C (Masia et al., 2007; Vassilev et al., 2010). As with other biomass components, the abundance of mineral elements varies among species. In general, compared with grass biomass, woody biomasses contain less ash, Cl, K, N, S, and Si, but more Ca (Vassilev et al., 2010).

Plant biomass components do not accumulate independently of each other, though their relationships are still an active area of research (Dick-Perez et al., 2011; Tan et al., 2013; Mikkelsen et al., 2015). Biomass component amounts can correlate because they are physically bound to each other through covalent and non-covalent bonds or because they accumulate in the same plant organ or stage of plant development, though a physical interaction may not exist. Because the abundance of some biomass components is correlated, the thermal products from one biomass component may also correlate with other components. For example, the abundance of cellulose correlates with the abundance of lignin in five different biomass sources (Pearson's correlation coefficient = 0.83) and lignin-derived thermal products correlate with cellulosic glucose (Mante et al., 2014). Many mineral elements are also correlated with each other, for example, N, S, and Cl; Si, Al, Fe, Na, and Ti; Ca, Mg, and Mn; K, P, S, and Cl (Vassilev et al., 2010). Numerous interactions between lignins and hemicelluloses and among hemicelluloses have been observed. Among the best-studied examples, GAXs of grasses and other recently evolved monocot species covalently link to lignin through ether bonds with ferulate esters on arabinose moieties of arabinoxylan (Bunzel et al., 2004). In poplar and spruce wood, NMR results indicate that lignin and carbohydrates are directly bonded through several types of ether linkages (Yuan et al., 2011; Du et al., 2014). The data provide evidence for ether

bonds between lignin and C1, C5, and C6 atoms of pentoses and hexoses (Yuan et al., 2011). Generally, xylan is the most closely associated polysaccharide to lignin, and NMR studies have also clearly identified lignin–glucuronic acid ester bonds (Yuan et al., 2011). Also, MLGs closely coat low-substituted xylan regions, likely via non-covalent interactions (Carpita et al., 2001; Kozlova et al., 2014). Furthermore, some components can also affect the distribution of other components. For example, rice plants that overexpress an enzyme that cleaves MLG exhibit reduced MLG and have an altered distribution profile of Si though maintain the same total amount of Si (Kido et al., 2015). In sum, mounting evidence supports covalent and non-covalent interactions among cell wall polymers and components; however, these connections have been difficult to study with questions persisting related to how different cell wall preparations and manipulation may alter observations.

MODELS FOR RELATIONSHIPS BETWEEN BIOMASS COMPONENTS AND BIO-OIL PRODUCT COMPOSITION

Reaction pathways of individual biomass components to formation of thermal products have been described (Collard and Blin, 2014). However, the pyrolysis literature suggests that biomass components tend to have more complex effects on bio-oil yield and product composition than simply their quantity. Here, we introduce three possible “models” of how biomass components may influence the yield or composition of thermal products, and in Section “Evidence Relating Biomass Content and Bio-oil Composition,” we discuss evidence supporting each of them. **Figure 2** provides schematic representations of the following models:

Model 1: Biomass components are the direct sources of thermal products. Components are converted to products through depolymerization and secondary reactions such as cracking, i.e., splitting, and recombination (**Figure 2A**).

Model 2: Components or their derived products act as catalysts that accelerate thermal reactions of other components, altering product yields and ratios (**Figure 2B**).

Model 3: Chemical interactions or structural relationships among cell wall components alter bio-oil composition and/or yield (**Figure 2C**). This “indirect” model applies when variation in a biomass component alters the yield of a chemically unrelated product in a manner not easily explained by a catalytic effect. Chemical interactions that alter products may either be covalent or non-covalent chemical bonds between cell wall components. Structural relationships refer to correlations between components, often minor ones, and physical features of the biomass. For example, the abundance of a cell wall component may be indicative of the structure of the plant material, such as biomass bulk density differences caused by different leaf to stem ratios, but do not reflect chemical bonding between components. As of the preparation of this review, very little evidence addresses how biological correlations effect bio-oil products, so the discussion focuses on potential chemical interactions.

EVIDENCE RELATING BIOMASS CONTENT AND BIO-OIL COMPOSITION

Evidence in the literature for the three models described above is presented in **Table 3** and discussed below. In the reviewed experiments, relationships between biomass components and pyrolysis products have been identified by varying the starting biomass, either through experimentation on purified components, via naturally occurring variation among different biomass sources, or via pretreatment of the biomass. Most studies included in this discussion report the chemical products derived from pyrolysis of biomass or biomass components. Studies that only reported weight losses or elemental balances were not considered. The two dominant techniques present in this corpus of literature are either pyrolysis-gas chromatography/mass spectroscopy, where pyrolysis vapors from microgram- to milligram-scale samples are directly transported to a GC for analysis, or pyrolysis in a gram- to kilogram-scale reactor system followed by condensation of the vapors and subsequent chromatographic analysis of the liquid.

Model 1: Direct Products of Cellulose, Hemicellulose, and Lignin

Thermal breakdown of purified cellulose, hemicellulose, and lignin has been relatively well studied. Levoglucosan, a six-carbon 1,6-anhyrosugar (see **Figure 1**), was identified as the main product of cellulose pyrolysis nearly a century ago (Pictet and Sarasin, 1918). Levoglucosan is formed alongside other smaller decomposition products, with maximum levoglucosan production occurring at 500°C (Shafizadeh et al., 1979). Minor products of cellulose pyrolysis are dominated by other anhydrosugars that retain all six carbons of glucose, such as 1,6-anhydroglucofuranose and 5-hydroxymethyl furfural, but also smaller molecules, like furfural (**5**, **Figure 1**), formic acid, and glycolaldehyde, among others (Patwardhan et al., 2011b).

As with cellulose, hemicellulose pyrolysis products depend mostly on the number of carbons in the monosaccharide residues of the starting polymer (Shafizadeh et al., 1972). Pentoses and hexoses produce similar light C1–C3 oxygenates but differ in the types and selectivities (i.e., relative ratios) of heavier C4–C6 products. Consistent with expectations, pyrolysis of monosaccharides reveals that hexoses can form more unique compounds than pentoses, including pyranic species; additionally, pentoses yield more lighter fragmentation products than hexoses and only trace amounts of C6 and higher products (Raisanen et al., 2003).

Lignin thermal degradation products generally retain the characteristic ring decoration of the monolignols from which they originate (**3**, **Figure 1**). For example, syringol derivatives are bio-oil products derived from S-lignin units and guaiacols are products derived from G-lignin units (**6**, **Figure 1**). The derivative groups possess 1–3 carbons and/or oxygenate moieties at the fourth position (**6**, **Figure 1**). Consistent with expectations, softwood lignins yield almost exclusively guaiacyl derivatives, while hardwood lignins yield both guaiacyl and syringyl derivatives. Grasses yield not only guaiacyl, syringyl, and *p*-hydroxyphenyl derivatives but also vinylphenol, propenyl-phenols, and *p*-hydroxybenzaldehyde that are not produced during pyrolysis of

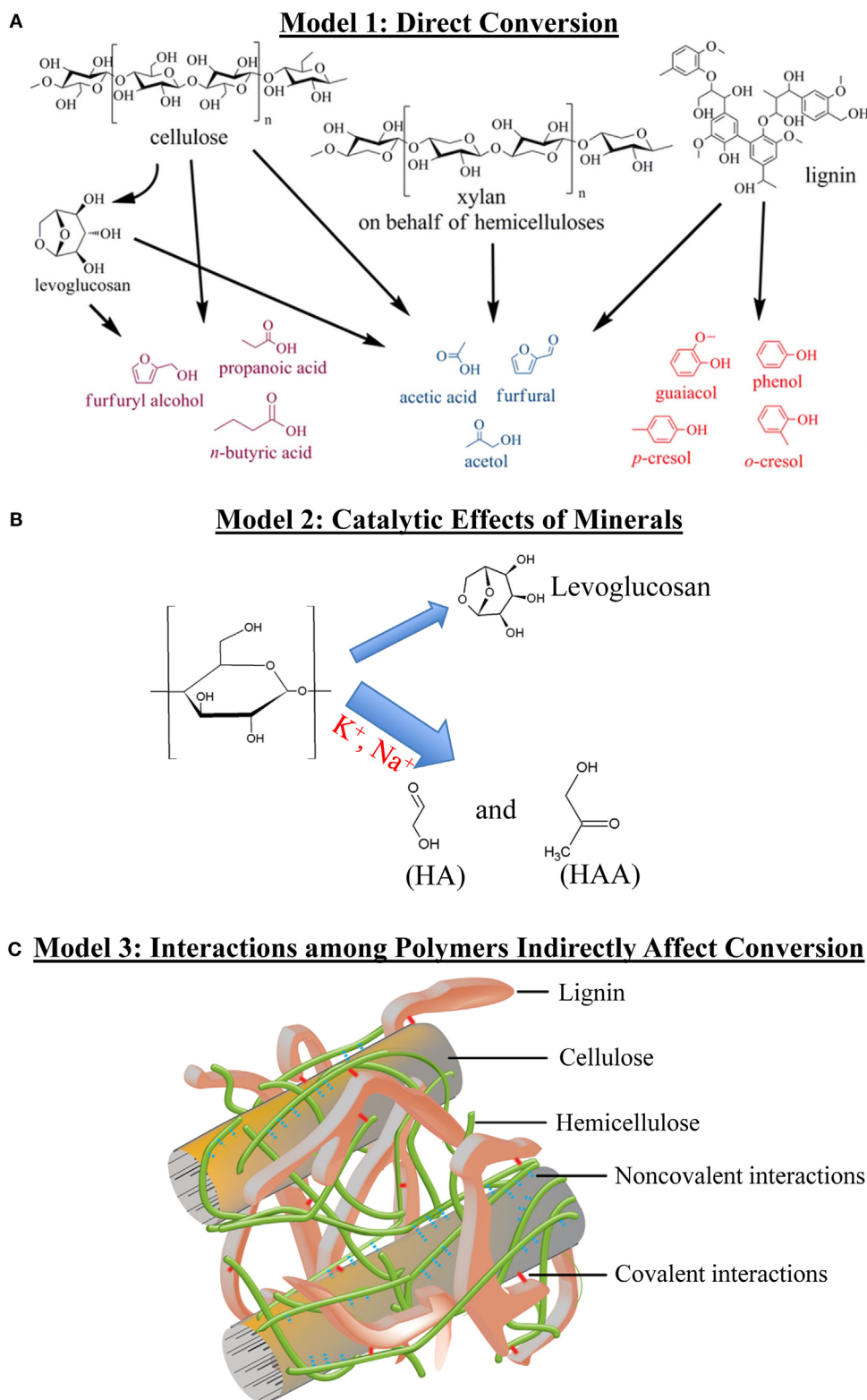


FIGURE 2 | Three models of how biomass components and their interactions affect the formation of thermal products. HA, hydroxyacetaldehyde; HAA, hydroxyacetone. **(A)** Direct conversion. **(B)** Catalytic effect of minerals. **(C)** Interactions among polymers indirectly affect conversion. **(A)** and **(C)** were adapted and modified with permission (Vanholme et al., 2010; Zhang et al., 2013).

TABLE 3 | Possible models of how biomass components affect the composition and yield of liquid thermal products.

Possible model	Biomass component	Variation of input material (independent variable)	Effect on products (dependent variable)	Sample type	Thermal conversion condition	Reference
1	Hemicellulose	Arabinose, xylose, mannose, arabinitol	Arabinose and xylose produce similar products but slightly different yields and product ratios. Mannose gives more decomposition products than arabinose and xylose and a unique product, 5-hydroxymethyl furfural	Powdered high-purity monosaccharides (>99%)	500 and 550°C, 10 s	Raisanen et al. (2003)
1	Hemicellulose or non-structural polysaccharides	Variation in sugar composition	Fermented grain samples produce less acetic acid, furfural, and acetone than hull and straw	Barley straw, hull and yeast-fermented grain	Fluidized bed, 500°C	Mullen et al. (2010)
1	Lignin	Lignin variation among hardwoods, softwoods, and grasses	Hardwood lignin produces guaiacyl and syringyl derivatives. Softwood lignin produces mostly guaiacyl derivatives but no hydroxyphenyl or syringyl compounds. Grass lignin uniquely produces vinylphenol, propenyl-phenols, and <i>p</i> -hydroxybenzaldehyde	Solvent extracted lignin from milled spruce, beech, aspen, and bamboo	510°C, 10 s	
1	Lignin	Natural variation and variation by hydrogen peroxide treatment in Klason lignin	Samples with high Klason lignin produce more 4-vinylguaiacol. Klason lignin positively correlates with 4-vinylguaiacol	Hybrid maize, <i>bm1</i> mutant, <i>bm3</i> mutant, and switchgrass biomass treated by different [H ₂ O ₂] to remove lignin	650°C, 20 s	Saizjimenez and Deleeuw (1986)
1	Lignin	Natural variation in lignin abundance	High-lignin endocarp biomasses produce more phenolic compounds, like phenol, 2-methoxyphenol, 2-methylphenol, 2-methoxy-4-methylphenol, and 4-ethyl-2-methoxyphenol, compared to switchgrass biomass	Walnut, olive, coconut husks, and peach drupe endocarp biomass and switchgrass biomass	650°C, 20 s	Li et al. (2012)
1	Lignin	Natural variation in lignin composition or structure	Compared with endocarp lignin, switchgrass lignin produces more acetic acid, toluene, furfural, and 4-methylphenol, and lower or undetectable amounts of 4-ethyl-2-methoxyphenol, 2-methoxy-4-vinylphenol and 2-methoxy-4-(2-propenyl)-phenol. Compared with walnut and olive endocarp lignin, coconut endocarp lignin produces strikingly more phenol and less 2-methoxy-4-methylphenol. Coconut shell lignin produces unique compounds among the analyzed feedstocks, such as 2,6-dimethoxyphenol, 2-methoxy-4-(2-propenyl)-phenol and vanillin, but less 2-methoxy-4-(1-propenyl)-phenol compared to walnut shell and olive lignin	Lignin extracted from various endocarps and switchgrass	650°C, 20 s	Mendu et al. (2011)
1	Lignin interunit linkages (lignin dimer)	Variation of lignin interunit linkages, including β-O-4' ethers, pinosresinols, phenylcoumaran, and dibenzodioxocins, among lignin-carbohydrate complexes with different compositions	Glucomannan-associated lignin produces more guaiacol than xylan-associated lignin and glucan-associated lignin	Lignin-carbohydrate complexes extracted from spruce wood	500°C, 1 min	Mendu et al. (2011)
2	Mineral content	Variation of K ⁺ , Mg ²⁺ , and Ca ²⁺ by demineralization and impregnation	Samples with more K ⁺ produce more glycolaldehyde, acetic acid, acetol, butanedial, guaiacol, syringol, and 4-vinylsyringol, but less levoglucosan, furans, and pyrans. Samples with more Mg ²⁺ produce less glycolaldehyde, levoglucosan, and 3-methoxycatechol. Samples with more Ca ²⁺ produce more pyrans and cyclopentenones but less levoglucosan and 3-methoxycatechol	Poplar wood powder demineralized with hydrofluoric acid and impregnated with KCl, MgCl ₂ , and CaCl ₂	550°C, 10 s	Eom et al. (2012)

(Continued)

TABLE 3 | Continued

Possible model	Biomass component	Variation of input material (independent variable)	Effect on products (dependent variable)	Sample type	Thermal conversion condition	Reference
2	Mineral content	Increase of minerals by adding 1% and 5% NaCl	Samples with more NaCl added produce more furans, acids, ketones, and phenols than the samples with no NaCl added	Rice straw and bamboo with NaCl	400–900°C, 4 min	Lou et al. (2013)
2	Mineral content	Decrease minerals by acid wash	Cellulose with the acid-washed char produces more levoglucosan but less formic acid, furane-type derivatives, and light oxygenates, than cellulose with unwashed char	Cellulose with acid-washed and unwashed red oak char	400°C, 15 s	Ronsse et al. (2012)
2	Mineral content	Decrease of K ⁺ and Na ⁺ by washing	Washed samples yield more levoglucosan, hydroxyacetaldehyde, and char than unwashed samples	Washed and unwashed lolium, festuca, willow, and switchgrass biomass	500°C, 10 s	Fahmi et al. (2007)
2	Mineral content	Increase mineral content by impregnation with NaCl, KCl, MgCl ₂ , CaCl ₂ , Ca(OH) ₂ , Ca(NO ₃) ₂ , CaCO ₃ , and CaHPO ₄	Impregnated samples produce less levoglucosan, but more glycolaldehyde and acetol, compared to untreated cellulose	Microcrystalline powdered cellulose impregnated with salt or switchgrass ash solution	500°C, 30 s and a range from 350 to 600°C with 50°C increments	Patwardhan et al. (2010)
2	Mineral content	Natural variation in ash	Ash negatively correlates with hydroxyacetaldehyde and phenolic compounds, such as trimethoxybenzene, syringol, 4-allyl-2,6-dimethoxyphenol, syringaldehyde, and 3,5-dimethoxy-4-hydroxycinnamaldehyde	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple	550°C, 20 s	Mante et al. (2014)
2	Mineral content	Increase mineral content by adding CaCl ₂ , NaCl, KCl, K ₂ SO ₄ , KHCO ₃ individually to washed wood	Washed wood with calcium salt added produces more C ₆ H ₈ O ₂ compound compared to washed wood with no additive. Samples with sodium salt and potassium salt added produce less 2-hydroxy-butanedial, 4-hydroxy-5,6-dihydro-(2H)-pyran-2-one, and levoglucosan, but more tetrahydro-4-hydroxy-pyran-2-one, compared to washed samples	Hornbeam wood washed and mixed with salts	280°C, 20 min, 100–500°C	Muller-Hagedorn et al. (2003)
2	Mineral content	Increase mineral content by adding 1% NaCl	Xylans with NaCl added yield more glycolaldehyde but less 1,4-anhydro- α -D-xylopyranose, 1,5-anhydro-4-deoxypent-1-en-3-ulose and char	Synthesized xylan	280°C, 30 min	Ponder and Richards (1991)
2	Mineral content	Replace the cations in biomass with Na ⁺ by ion-exchange or increase mineral content by adding Na ₂ SO ₄ or NaHCO ₃	Sodium-containing pulps produce less polysaccharide-derived products and similar amounts of lignin-derived products compared to untreated pulps	Spruce pulp, treated with ion exchange and salt impregnation	620°C, 2 s	Kleen and Gellerstedt (1995)
3	Intercomponent linkages	Biomass with hemicellulose selectively removed, compared with a physical mixtures of lignin and cellulose, and separate pyrolysis of lignin and cellulose	The native grass cellulose–lignin samples produce more C1 to C3 products and furan, but less pyrans and levoglucosan	Corn stover, red oak, pine, and switchgrass depleted of hemicellulose and compared with different extracted component mixtures	500°C	Zhang et al. (2015)

(Continued)

TABLE 3 | Continued

Possible model	Biomass component	Variation of input material (independent variable)	Effect on products (dependent variable)	Sample type	Thermal conversion condition	Reference
1 and 2 or 3	Acetyl content	Natural variation in acetyl content among five species	Acetyl content positively correlates with acetic acid, but also methyl pyruvate and 2-furanone	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple	550°C, 20 s	Mante et al. (2014)
1 and 3	Acetyl content	Variation of acetylation between acetylxylan and xylan	The acetylxylan produces more acetic acid, furan, and acetone, but less 2-furfural and acetaldehyde, than xylan	Xylans extracted from cotton wood	500°C	Shafizadeh et al. (1972)
1 and 3	Cellulose	Natural variation in glucans, mostly cellulose, among five species	Glucans positively correlate with levoglucosan and hydroxymethyl furfural, but also phenolic compounds like syringaldehyde, vanillin, and 3,5-dimethoxy-4-hydroxycinnamaldehyde	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple	550°C, 20 s	Mante et al. (2014)
1 and 3	Lignin	Natural variation in lignin among five species	Lignin positively correlates with hydroxyacetaldehyde and phenolic compounds, trimethoxybenzene; syringol; 4-allyl-2,6-dimethoxyphenol; syringaldehyde; and 3,5-dimethoxy-4-hydroxycinnamaldehyde	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple	550°C, 20 s	Mante et al. (2014)
1 and 3	Xylan	Natural variation in xylan among five species	Xylan positively correlates with hydroxyacetone, but negatively correlates with hydroxymethyl furfural and syringaldehyde	Poplar, willow, switchgrass, hot-water extracted sugar maple, and debarked sugar maple	550°C, 20 s	Mante et al. (2014)
2 and 3	Mineral content or other structure	Variation in weakly associated and strongly associated K, Na, Mg, and Ca by washing with deionized water or nitric acid	Acid-washed samples with less mineral content produce bio-oil with a greater fraction of water-insoluble content and produce more levoglucosan and sugar compounds but less monophenols like phenol, 2-methyl-phenol, syringol compared to unwashed samples	Eucalyptus wood	Fluidized bed, 500°C	Mourant et al. (2011)
3	Water	Natural variation in moisture content (0–20%)	Samples with greater moisture content produce more char and gas but less water compared to lower moisture samples	Pine wood	Fluidized bed, 480°C, 2 s	Westerhof et al. (2007)
3	Water	Natural variation in moisture content (2–55%)	Samples with greater moisture content produce more condensable products, but less char than dry samples	Norway spruce	500°C, 30 min	Burhenne et al. (2013)
3	Water	Natural variation in moisture content (5–15%)	Samples with greater moisture content produce less levoglucosan at 450 and 500°C than samples with lower moisture content. Other products are also significantly affected by moisture content, but trends depend on pyrolysis temperature	Switchgrass	Fluidized bed, 450, 500, and 550°C	He et al. (2009)
3	Intercomponent linkages	Variation in polysaccharide amounts in lignin–carbohydrate complexes by enzyme treatment	Samples with lignin-associated polysaccharides removed produce more coniferyl alcohol than untreated samples	Lignin–carbohydrate complexes extracted from spruce wood	500°C, 1 min	Du et al. (2014)

softwood and hardwood (Saizjimenez and Deleeuw, 1986; Mante et al., 2014) and are likely derived from ferulate and coumarate esters (Penning et al., 2014b). Phenol derivatives are the large majority of the products formed from lignin pyrolysis; aromatic hydrocarbons and some furan derivatives are also detectable, but at very low amounts that might represent lignin sample contaminants (Saizjimenez and Deleeuw, 1986). Lignins from spruce wood with different dimer compositions also show different product distributions, including variations in the yield of major products like guaiacol (Du et al., 2014). This suggests that bonds between lignin units and the lignin structure determined by those bonds may impact pyrolysis as well.

Model 2: Secondary Reactions Catalyzed by Inorganic Components

The biopolymers that make up the majority of the biomass by weight are established as the primary source of bio-oil products formed during thermal degradation. However, secondary reactions occur during the pyrolysis process involving other components present within the biomass (Ponder and Richards, 1991; Kleen and Gellerstedt, 1995; Muller-Hagedorn et al., 2003; Fahmi et al., 2007; Patwardhan et al., 2010; Ronsse et al., 2012; Lou et al., 2013; Mante et al., 2014). As products form, they can interact with catalytic minerals in the residual solid. For example, levoglucosan has been shown to react on minerals present in the residual char from pyrolysis of biomass. The products formed include levoglucosenone, furan derivatives, and lighter oxygenates such as acetic acid, acetone, and acetol. Demineralization prohibits the formation of these products (Fahmi et al., 2007; Ronsse et al., 2012).

Different inorganics are responsible for different kinds of secondary reactions. In general, the presence of metal cations enhances the homolytic cleavage of pyranose ring bonds over the heterolytic cleavage of glycosidic linkages, leading to the increased formation of light oxygenate decomposition products at the expense of levoglucosan formation. While Na^+ , K^+ , Mg^{2+} , and Ca^{2+} all catalyze levoglucosan decomposition, the effects of group 1 (alkali metals) and group 2 (alkaline) elements differ. Increased Na^+ and K^+ alkali metal loading increased formic acid, glycolaldehyde, and acetol more than similar amounts of the alkaline metals, Mg^{2+} and Ca^{2+} , though more furfural is produced with increasing concentrations of Mg^{2+} and Ca^{2+} . Additionally, the alkali metals reduce levoglucosan production at very low thresholds. This suggests that Na^+ and K^+ ultimately promote cracking reactions while Mg^{2+} and Ca^{2+} promote dehydration reactions (Muller-Hagedorn et al., 2003; Patwardhan et al., 2010; Eom et al., 2012).

Model 3: Interactions and Linkages Between Primary Components

While the first two models address the direct conversion of biopolymer organic components to related bio-oil products and their further reaction catalyzed by biomass inorganics, the third addresses compositional and structural relationships among cell wall components and their impact on products. Interactions between polysaccharides and lignin have been shown to alter pyrolysis products (Du et al., 2014; Zhang et al., 2015). The cellulose–lignin interaction can lead to a decrease in levoglucosan

yield and an increase in light (C1–C3) compounds, especially glycolaldehyde and furans. Based on the nature of the small products, Zhang et al. (2015) hypothesized that the cellulose–lignin interaction occupies the C6 position, disfavoring glycosidic bond cleavage that is required for the formation of levoglucosan and favoring light compound and furan formation through ring scission, rearrangement, and dehydration reactions. The strength of this effect on pyrolysis products is most pronounced in grasses, followed by softwood and then hardwood, possibly due to the increased prevalence of covalent bonds between cellulose and lignin in grass cell walls (Jin et al., 2006; Zhou et al., 2010). Hemicellulose–lignin interactions, especially the xylan–lignin interaction revealed in NMR experiments (Yuan et al., 2011), may also affect pyrolysis. Indeed, enzymatic removal of hemicelluloses from lignin–carbohydrate complexes increased coniferyl alcohol yields (Du et al., 2014).

An example of a compositional feature that may impact product distribution is the degree of acetylation of the biopolymers. As mentioned, acetyl groups decorate hemicellulose side chains and are also present in the lignin. The increased abundance of these groups in biomass correlates with increasing yields of acetic acid, methyl pyruvate, acetone, and furan; additionally, this acetylation correlates with decreasing yields of furfural and acetaldehyde (Shafizadeh et al., 1972; Mante et al., 2014). While the acetic acid and perhaps the methyl pyruvate can be explained by the direct production of these compounds upon pyrolytic decomposition (Model 1), the nature of the relationship between acetate and the furanic and other 4-carbon species has not been clearly defined. The production of the 4-carbon species may be due to an indirect effect (Model 3) or may be the result of catalytic reaction of acetate with itself (Model 2).

Several investigations (Westerhof et al., 2007; He et al., 2009; Burhenne et al., 2013) suggest that feedstock moisture content can also play a role in the yield and product distribution of the organic fraction of the bio-oil. As previously discussed, the presence of water in bio-oil prohibits its direct use and creates challenges to catalytic valorization. For these reasons, biomass is typically subjected to drying prior to pyrolysis, which both reduces the required energy of the pyrolysis step and limits the water in the liquid condensate to water produced by decomposition reactions. However, the degree to which the feedstock moisture content should be eliminated is still under investigation. Burhenne et al. (2013) found that higher feedstock moisture content led to slightly lower char and gas yields upon pyrolysis with minimal changes to the elemental composition of the char. However, this is in disagreement with Westerhof et al. (2007) who observed slightly higher char yields with increasing moisture content. The water weight fraction distribution of the feedstocks in the two studies were quite different, 2.4–55.4% in Burhenne et al. versus 0–20% in Westerhof et al. Beyond impacts to the yields, He et al. (2009) studied the change in selectivity to the organic fraction produced upon pyrolysis of switchgrass with 5, 10, and 15% feedstock moisture contents. The authors found that at 500°C, the lowest moisture content feedstock produced the highest amounts of levoglucosan and acetic acid. The authors note that while significant differences in pyrolysis products were observed, they could not identify clear trends in their data. Among these

studies, the observable but sometimes contradictory or unclear trends suggest that the feedstock moisture content may have multiple impacts on the pyrolysis process, possibly related to the physical location of the water in biomass.

In addition to compositional factors, morphological factors also influence the bio-oil product distributions. Biomass undergoing thermal decomposition retains its morphology even in harsh thermal treatment regimes (Pohlmann et al., 2014). Biomass is a poor conductor of heat (conductivity <0.1 W/m K) (Bridgwater et al., 1999), and large temperature gradients occur in heated biomass particles (Bryden et al., 2002). Most reactor systems for thermal degradation require size reduction of biomass particles; as an example, fluidized beds require particle sizes no larger than 2 mm (Bridgwater et al., 1999) to ensure rapid reaction. These particle sizes are larger than the tissue structures present in biomass. While the overall tissue and cellular morphology remain intact, micropore formation and shrinkage during the reaction process can occur in a non-uniform manner throughout the biomass (Davidsson and Pettersson, 2002; Pohlmann et al., 2014). Piskorz and colleagues observed decreasing liquid yields with increasing particle size, attributed to increasing incidence of secondary reactions with in wood particles (Scott and Piskorz, 1984). The principles of internal and external diffusion and the impacts of tortuosity, surface area, and diffusion path lengths are all fundamental to catalytic reaction engineering, and in the case of thermal biomass conversion, these important parameters are all dictated by the reacting feedstock (Fogler, 2006). Some evidence supports the notion that different plant developmental stages, which are related to the ratio of leaves to stems and biomass density, result in different pyrolysis products. For example, switchgrass harvested at later times during the growing season produced increased yields of condensable products, relative to that from younger, leafier material (Boateng et al., 2006), though compositional and developmental differences of the starting material were not carefully assessed.

CONCLUSION

Years of research have led to understanding of the direct pyrolysis conversion pathways of the major monomeric and polymeric constituents of biomass (Model 1, **Table 2**). The observation that these constituents often represent minor components in raw bio-oil (**Table 1**) highlights the importance of catalytic degradation (Model 2) and possibly indirect effects (Model 3) on pyrolysis products. The latter model is only recently receiving attention as knowledge of cell wall structures and analytical repertoires blossom (Mante et al., 2014; Zhang et al., 2015). Detailed examination of the relationships between components and products is still sparse, with the biological literature providing detailed characterization of cell wall components, while the engineering literature analyzes the chemical components, or often just total yields, of different pyrolysis fractions. We would argue that further investigations on the relationships between biomass components and thermal products will allow improvement of thermal product “quality.” Short of attaining (or improving on) petroleum fuel-like properties, even the criteria for a high-quality thermal product remain unclear. As discussed, this is, in part, because methods

for upgrading are so dependent on bio-oil composition. Thus, methods that economically separate and/or simplify the different product streams, while still maintaining C–C bonds and overall C-content, are more likely to be amenable to catalytic upgrading.

Greater and more systematic analysis of biomass composition and pyrolysis products within species that show significant compositional variation will aid in better understanding biomass–bio-oil relationships. Much of the existing literature relies on comparisons of thermal degradation products across diverse taxonomic groups that vary greatly in cell wall composition beyond the biomass components measured (**Table 3**). An analysis of more subtle compositional differences, in which compositional factors are varied across different samples, may aid in refining biomass–bio-oil relationships. For example, genetic mutants that vary in only one component relative to near isogenic, unmutated “wild-type” plants can directly address relationships between starting components and products (Li et al., 2012). In addition to genetically determined compositional differences, biomass composition also depends on growth conditions and developmental stage, which relates to harvest time. Taken together, the scale of the problem points to the value of developing high-throughput methods to help identify species and genotypes that are most suitable for production of specific thermal products and to guide the optimization of genetic stocks and growth condition for bio-energy crops. Methods available to identify such “high-quality” biomass include near-infrared reflectance spectroscopy (Vogel et al., 2011), Fourier transform near-infrared spectroscopy (Liu et al., 2010), and pyrolysis molecular beam mass spectrometry, at least for lignin components (Sykes et al., 2009; Penning et al., 2014b). In general, these methods can be trained, either rationally or in a model-independent manner, to detect spectroscopic or molecular signatures in biological materials with linear or non-linear relationships to thermal products.

Besides selecting or breeding for natural variation in biomass composition (Wegrzyn et al., 2010; Penning et al., 2014a), it is also possible to genetically modify biomass composition (Bartley and Ronald, 2009). Most simply, genetic engineering of bioenergy plants can be achieved by modifying the plant’s genome to (1) express genes from other organisms, (2) increase expression of native genes, or (3) reduce expression of native genes. More complex schemes are also possible, in which expression patterns of genes are altered through synthetic biology approaches that recombine various genetic elements (Yang et al., 2013). The most common method for plant genetic engineering co-opts the molecular machinery of a bacterial pathogen that introduces genes into plant chromosomes to facilitate its pathogenesis.

Genetic engineering to improve bio-oil production would aim to increase biomass components that enhance the yield of favored products and/or to decrease components that produce disfavored products or interfere with upgrading strategies. Advances in understanding cell wall biosynthesis, including genes responsible for synthesizing the major polymer classes (Bonawitz and Chapple, 2010; Scheller and Ulvskov, 2010; Pauly et al., 2013) and covalent interactions among them (Chiniquy et al., 2012; Bartley et al., 2013; Schultink et al., 2015); regulation of expression of the cell wall biosynthesis genes (Zhao and Dixon, 2011); and metal ion transport proteins that determine the abundance and

distribution of plant mineral content (Ma et al., 2006; Yamaji and Ma, 2009; Zhong and Ye, 2015), lay the foundation for genetically engineering bioenergy crop cell wall content and structure. For example, lignin is an important target for genetic engineering for pyrolysis since the major lignin-derived products have a lower O:C ratio, a higher energy value, and are more stable than sugar-derived products (Tanger et al., 2013; Mante et al., 2014). Some important genes that participate in or regulate lignin synthesis have already been modified in energy crops without major interference with plant biomass yield (Baxter et al., 2014, 2015; reviewed in Bartley et al., 2014). However, current genetic engineering strategies are focused on developing low lignin biomass for saccharification and biochemical conversion to fuels. Therefore, more work is required to develop biomass with high-lignin content for thermal conversion. Producing corrosive acetic acid in bio-oil (Mante et al., 2014), acetyl groups on cell wall polymers are another potential target for genetic engineering of “pyrolysis crops.” Three enzyme classes, including the reduced wall acetylation (RWA) proteins, Trichome birefringence-like (TBL) and Altered Xyloglucan (AXY) proteins acetylate cell wall polysaccharides (Lee et al., 2011; Xiong et al., 2013; Schultink et al., 2015). A mutant of the dicot reference plant, *Arabidopsis thaliana*, which lacks expression of all four RWA genes, shows a 40% reduction in secondary wall-associated acetyl groups (Lee et al., 2011). Reducing expression of this family of genes in bioenergy crops may help to solve the problems caused by acetic acid in bio-oil produced from such plants.

Pretreatments such as washing/leaching and torrefaction are another class of strategies to improve biomass quality by changing biomass composition (Zheng et al., 2013; Banks et al., 2014).

REFERENCES

- Adler, P. R., Sanderson, M. A., Boateng, A. A., Weimer, P. I., and Jung, H. J. G. (2006). Biomass yield and biofuel quality of switchgrass harvested in fall or spring. *Agron. J.* 98, 1518–1525. doi:10.2134/agronj2005.0351
- Banks, S. W., Nowakowski, D. J., and Bridgwater, A. V. (2014). Fast pyrolysis processing of surfactant washed *Miscanthus*. *Fuel Process. Technol.* 128, 94–103. doi:10.1016/j.fuproc.2014.07.005
- Bartley, L. E., Peck, M. L., Kim, S. R., Ebert, B., Manisseri, C., Chiniy, D. M., et al. (2013). Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiol.* 161, 1615–1633. doi:10.1104/pp.112.208694
- Bartley, L. E., and Ronald, P. C. (2009). Plant and microbial research seeks bio-fuel production from lignocellulose. *Calif. Agric.* 63, 178–184. doi:10.3733/ca.v063n04p178
- Bartley, L. E., Tao, X., Zhang, C., Nguyen, H., and Zhou, J. (2014). “Switchgrass biomass content, synthesis, and biochemical conversion to biofuels,” in *Compendium of Bioenergy Plants*, eds Luo H., and Wu Y. (Boca Raton, FL: Science), 109–169.
- Baxter, H. L., Mazarei, M., Labbe, N., Kline, L. M., Cheng, Q., Windham, M. T., et al. (2014). Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnol. J.* 12, 914–924. doi:10.1111/pbi.12195
- Baxter, H. L., Poovaiah, C. R., Yee, K. L., Mazarei, M., Rodriguez, M., Thompson, O. A., et al. (2015). Field evaluation of transgenic switchgrass plants overexpressing *PvMYB4* for reduced biomass recalcitrance. *Bioenerg. Res.* 8, 910–921. doi:10.1007/s12155-014-9570-1
- Boateng, A. A., Hicks, K. B., and Vogel, K. P. (2006). Pyrolysis of switchgrass (*Panicum virgatum*) harvested at several stages of maturity. *J. Anal. Appl. Pyrolysis* 75, 55–64. doi:10.1016/j.jaap.2005.03.005
- For example, by washing biomass with detergent (Triton) or acid to remove minerals, the yield of bio-oil is increased and reaction water content is reduced (Banks et al., 2014). Coupling biochemical conversion of biomass, which depletes the polysaccharide fraction, with pyrolysis of the resulting residue, or bagasse, is another avenue to explore further (Islam et al., 2010; Cunha et al., 2011). Torrefaction is a low-temperature (200–400°C) thermal pretreatment that decomposes hemicellulose and may segregate disfavored products such as water and acid into intermediate streams before the next stage of pyrolysis (Zheng et al., 2013). More efficient torrefaction may be achieved by changing the composition or chemical structure of hemicellulose through genetic methods to further separate the decomposition temperatures of hemicellulose from lignin and cellulose. By identifying and studying the roles of key biomass components during thermal conversion, it will be possible to maximize the economic and environmental benefits of plant biomass-derived biofuels in the future.

AUTHOR CONTRIBUTIONS

FL and CW drafted and revised the manuscript. LB, LL, and RM revised the manuscript.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2010-38502-21836, through the South Central Sun Grant Program.

- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi:10.1146/annurev.arplant.54.031902.134938
- Bonawitz, N. D., and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Ann. Rev. Genet.* 44, 337–363. doi:10.1146/annurev-genet-102209-163508
- Bridgwater, A. V. (1994). Catalysis in thermal biomass conversion. *Appl. Catal. A Gen.* 116, 5–47. doi:10.1016/0926-860x(94)80278-5
- Bridgwater, A. V., Meier, D., and Radlein, D. (1999). An overview of fast pyrolysis of biomass. *Org. Geochem.* 30, 1479–1493. doi:10.1016/S0146-6380(99)00120-5
- Bryden, K. M., Ragland, K. W., and Rutland, C. J. (2002). Modeling thermally thick pyrolysis of wood. *Biomass Bioenergy* 22, 41–53. doi:10.1016/S0961-9534(01)00060-5
- Bule, M. V., Gao, A. H., Hiscox, B., and Chen, S. (2013). Structural modification of lignin and characterization of pretreated wheat straw by ozonation. *J. Agric. Food Chem.* 61, 3916–3925. doi:10.1021/jf4001988
- Bunzel, M., Ralph, J., Lu, F., Hatfield, R. D., and Steinhart, H. (2004). Lignins and ferulate-coniferyl alcohol cross-coupling products in cereal grains. *J. Agric. Food Chem.* 52, 6496–6502. doi:10.1021/jf040204p
- Burhenne, L., Damiani, M., and Aicher, T. (2013). Effect of feedstock water content and pyrolysis temperature on the structure and reactivity of spruce wood char produced in fixed bed pyrolysis. *Fuel* 107, 836–847. doi:10.1016/j.fuel.2013.01.033
- Carlson, T. R., Tompsett, G. A., Conner, W. C., and Huber, G. W. (2009). Aromatic production from catalytic fast pyrolysis of biomass-derived feedstocks. *Top. Catal.* 52, 241–252. doi:10.1007/s11244-008-9160-6
- Carlson, T. R., Vispute, T. R., and Huber, G. W. (2008). Green gasoline by catalytic fast pyrolysis of solid biomass derived compounds. *ChemSusChem* 1, 397–400. doi:10.1002/cssc.200800018

- Carpita, N. C., Defernez, M., Findlay, K., Wells, B., Shoue, D. A., Catchpole, G., et al. (2001). Cell wall architecture of the elongating maize coleoptile. *Plant Physiol.* 127, 551–565. doi:10.1104/pp.010146
- Chen, F., Tobimatsu, Y., Havkin-Frenkel, D., Dixon, R. A., and Ralph, J. (2012). A polymer of caffeyl alcohol in plant seeds. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1772–1777. doi:10.1073/pnas.1120992109
- Chen, F., Tobimatsu, Y., Jackson, L., Nakashima, J., Ralph, J., and Dixon, R. A. (2013). Novel seed coat lignins in the Cactaceae: structure, distribution and implications for the evolution of lignin diversity. *Plant J.* 73, 201–211. doi:10.1111/tpj.12012
- Chiniquy, D., Sharma, V., Schultink, A., Baidoo, E. E., Rautengarten, C., Cheng, K., et al. (2012). XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17117–17122. doi:10.1073/pnas.1202791109
- Collard, F. X., and Blin, J. (2014). A review on pyrolysis of biomass constituents: mechanisms and composition of the products obtained from the conversion of cellulose, hemicelluloses and lignin. *Renew. Sustain. Energ. Rev.* 38, 594–608. doi:10.1016/j.rser.2014.06.013
- Cunha, J. A., Pereira, M. M., Valente, L. M. M., De La Piscina, P. R., Homs, N., and Santos, M. R. L. (2011). Waste biomass to liquids: low temperature conversion of sugarcane bagasse to bio-oil. The effect of combined hydrolysis treatments. *Biomass Bioenergy* 35, 2106–2116. doi:10.1016/j.biombioe.2011.02.019
- Davidsson, K. O., and Pettersson, J. B. C. (2002). Birch wood particle shrinkage during rapid pyrolysis. *Fuel* 81, 263–270. doi:10.1016/S0016-2361(01)00169-7
- Demirbas, A. (2011). Competitive liquid biofuels from biomass. *Appl. Energy* 88, 17–28. doi:10.1016/j.apenergy.2010.07.016
- Dick-Perez, M., Zhang, Y., Hayes, J., Salazar, A., Zabolina, O. A., and Hong, M. (2011). Structure and interactions of plant cell-wall polysaccharides by two- and three-dimensional magic-angle-spinning solid-state NMR. *Biochemistry* 50, 989–1000. doi:10.1021/bi101795q
- Du, X., Perez-Boada, M., Fernandez, C., Rencoret, J., Del Rio, J. C., Jimenez-Barbero, J., et al. (2014). Analysis of lignin-carbohydrate and lignin-lignin linkages after hydrolase treatment of xylan-lignin, glucomannan-lignin and glucan-lignin complexes from spruce wood. *Planta* 239, 1079–1090. doi:10.1007/s00425-014-2037-y
- El-Nashaar, H. M., Banowetz, G. M., Griffith, S. M., Casler, M. D., and Vogel, K. P. (2009). Genotypic variability in mineral composition of switchgrass. *Bioresour. Technol.* 100, 1809–1814. doi:10.1016/j.biortech.2008.09.058
- Eom, I. Y., Kim, J. Y., Kim, T. S., Lee, S. M., Choi, D., Choi, I. G., et al. (2012). Effect of essential inorganic metals on primary thermal degradation of lignocellulosic biomass. *Bioresour. Technol.* 104, 687–694. doi:10.1016/j.biortech.2011.10.035
- Fahmi, R., Bridgwater, A. V., Darvell, L. I., Jones, J. M., Yates, N., Thain, S., et al. (2007). The effect of alkali metals on combustion and pyrolysis of *Lolium* and *Festuca* grasses, switchgrass and willow. *Fuel* 86, 1560–1569. doi:10.1016/j.fuel.2006.11.030
- Faix, O., Fortmann, I., Bremer, J., and Meier, D. (1991a). Thermal-degradation products of wood – a collection of electron-impact (EI) mass-spectra of polysaccharide derived products. *Holz Roh Werkst.* 49, 299–304. doi:10.1007/Bf02663795
- Faix, O., Fortmann, I., Bremer, J., and Meier, D. (1991b). Thermal-degradation products of wood – gas-chromatographic separation and mass-spectrometric characterization of polysaccharide derived products. *Holz Roh Werkst.* 49, 213–219. doi:10.1007/Bf02613278
- Fogler, H. S. (2006). *Elements of Chemical Reaction Engineering*. Upper Saddle River, NJ: Prentice Hall PTR.
- Furimsky, E. (2000). Catalytic hydrodeoxygenation. *Appl. Catal. A Gen.* 199, 147–190. doi:10.1016/S0926-860X(99)00555-4
- Gonzalez-Borja, M. A., and Resasco, D. E. (2015). Reaction pathways in the liquid phase alkylation of biomass-derived phenolic compounds. *AIChE J.* 61, 598–609. doi:10.1002/aic.14658
- Hatfield, R. D., Marita, J. M., and Frost, K. (2008). Characterization of *p*-coumarate accumulation, *p*-coumaroyl transferase, and cell wall changes during the development of corn stems. *J. Sci. Food Agric.* 88, 2529–2537. doi:10.1002/jsfa.3376
- He, R., Ye, X. P., English, B. C., and Satrio, J. A. (2009). Influence of pyrolysis condition on switchgrass bio-oil yield and physicochemical properties. *Bioresour. Technol.* 100, 5305–5311. doi:10.1016/j.biortech.2009.02.069
- Huber, G. W., Iborra, S., and Corma, A. (2006). Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem. Rev.* 106, 4044–4098. doi:10.1021/cr068360d
- Islam, M. R., Parveen, M., and Haniyu, H. (2010). Properties of sugarcane waste-derived bio-oils obtained by fixed-bed fire-tube heating pyrolysis. *Bioresour. Technol.* 101, 4162–4168. doi:10.1016/j.biortech.2009.12.137
- Jin, Z., Katsumata, K. S., Lam, T. B., and Iiyama, K. (2006). Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods. *Biopolymers* 83, 103–110. doi:10.1002/bip.20533
- Kawamoto, H., Horigoshi, S., and Saka, S. (2007). Pyrolysis reactions of various lignin model dimers. *J. Wood Sci.* 53, 168–174. doi:10.1007/s10086-006-0834-z
- Kido, N., Yokoyama, R., Yamamoto, T., Furukawa, J., Iwai, H., Satoh, S., et al. (2015). The matrix polysaccharide (1;3,1;4)-beta-D-glucan is involved in silicon-dependent strengthening of rice cell wall. *Plant Cell Physiol.* 56, 268–276. doi:10.1093/pcp/pcu162
- Kleen, M., and Gellerstedt, G. (1995). Influence of inorganic species on the formation of polysaccharide and lignin degradation products in the analytical pyrolysis of pulps. *J. Anal. Appl. Pyrolysis* 35, 15–41. doi:10.1016/0165-2370(95)00893-J
- Kozlova, L. V., Ageeva, M. V., Ibragimova, N. N., and Gorskova, T. A. (2014). Arrangement of mixed-linkage glucan and glucuronoarabinoxylan in the cell walls of growing maize roots. *Ann. Bot.* 114, 1135–1145. doi:10.1093/aob/mcu125
- Lan, W., Lu, F., Regner, M., Zhu, Y., Rencoret, J., Ralph, S. A., et al. (2015). Tricin, a flavonoid monomer in monocot lignification. *Plant Physiol.* 167, 1284–1295. doi:10.1104/pp.114.253757
- Lee, C., Teng, Q., Zhong, R., and Ye, Z. H. (2011). The four *Arabidopsis* reduced wall acetylation genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. *Plant Cell Physiol.* 52, 1289–1301. doi:10.1093/pcp/pcr075
- Li, M., Foster, C., Kelkar, S., Pu, Y., Holmes, D., Ragauskas, A., et al. (2012). Structural characterization of alkaline hydrogen peroxide pretreated grasses exhibiting diverse lignin phenotypes. *Biotechnol. Biofuels* 5, 38. doi:10.1186/1754-6834-5-38
- Liu, L., Ye, X. P., Womac, A. R., and Sokhansanj, S. (2010). Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydr. Polym.* 81, 820–829. doi:10.1016/j.carbpol.2010.03.058
- Lou, R., Wu, S. B., Lv, G. J., and Zhang, A. L. (2013). Factors related to minerals and ingredients influencing the distribution of pyrolysates derived from herbaceous biomass. *Bioresources* 8, 1345–1360. doi:10.15376/biores.8.1.1345-1360
- Lu, F., Karlen, S. D., Regner, M., Kim, H., Ralph, S. A., Sun, R.-C., et al. (2015). Naturally *p*-hydroxybenzoylated lignins in palms. *Bioenergy Res.* 8, 934–952. doi:10.1007/s12155-015-9583-4
- Lu, F. C., and Ralph, J. (2002). Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf. *Chem. Commun.* 90–91. doi:10.1039/b109876d
- Ma, J. F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M., et al. (2006). A silicon transporter in rice. *Nature* 440, 688–691. doi:10.1038/nature04590
- Mante, O. D., Babu, S. P., and Amidon, T. E. (2014). A comprehensive study on relating cell-wall components of lignocellulosic biomass to oxygenated species formed during pyrolysis. *J. Anal. Appl. Pyrolysis* 108, 56–67. doi:10.1016/j.jaap.2014.05.016
- Masia, A. A. T., Buhre, B. J. P., Gupta, R. P., and Wall, T. F. (2007). Characterising ash of biomass and waste. *Fuel Process. Technol.* 88, 1071–1081. doi:10.1016/j.fuproc.2007.06.011
- McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresour. Technol.* 83, 37–46. doi:10.1016/S0960-8524(01)00118-3
- Mendu, V., Harman-Ware, A. E., Crocker, M., Jae, J., Stork, J., Morton, S. III, et al. (2011). Identification and thermochemical analysis of high-lignin feedstocks for biofuel and biochemical production. *Biotechnol. Biofuels* 4, 43. doi:10.1186/1754-6834-4-43
- Mettler, M. S., Vlachos, D. G., and Dauenhauer, P. J. (2012). Top ten fundamental challenges of biomass pyrolysis for biofuels. *Energy Environ. Sci.* 5, 7797–7809. doi:10.1039/C2EE21679E
- Mikkelsen, D., Flanagan, B. M., Wilson, S. M., Bacic, A., and Gidley, M. J. (2015). Interactions of arabinoxylan and (1,3)(1,4)-beta-glucan with cellulose networks. *Biomacromolecules* 16, 1232–1239. doi:10.1021/acs.biomac.5b00009
- Mohan, D., Pittman, C. U., and Steele, P. H. (2006). Pyrolysis of wood/biomass for bio-oil: a critical review. *Energy Fuel* 20, 848–889. doi:10.1021/ef0502397
- Mourant, D., Wang, Z. H., He, M., Wang, X. S., Garcia-Perez, M., Ling, K. C., et al. (2011). Mallee wood fast pyrolysis: effects of alkali and alkaline earth metallic species on the yield and composition of bio-oil. *Fuel* 90, 2915–2922. doi:10.1016/j.fuel.2011.04.033

- Mullen, C. A., Boateng, A. A., Hicks, K. B., Goldberg, N. M., and Moreau, R. A. (2010). Analysis and comparison of bio-oil produced by fast pyrolysis from three barley biomass/byproduct streams. *Energy Fuel* 24, 699–706. doi:10.1021/ef900912s
- Muller-Hagedorn, M., Bockhorn, H., Krebs, L., and Muller, U. (2003). A comparative kinetic study on the pyrolysis of three different wood species. *J. Anal. Appl. Pyrolysis* 6, 231–249. doi:10.1016/S0165-2370(03)00065-2
- Newman, R. H., Hill, S. J., and Harris, P. J. (2013). Wide-angle x-ray scattering and solid-state nuclear magnetic resonance data combined to test models for cellulose microfibrils in mung bean cell walls. *Plant Physiol.* 163, 1558–1567. doi:10.1104/pp.113.228262
- NSF. (2008). “Selective thermal processing of cellulosic biomass and lignin” in *Breaking the Chemical and Engineering Barriers to Lignocellulosic Biofuels: Next Generation Hydrocarbon Biorefineries*. ed. Huber G. W. (Washington, DC: University of Massachusetts, Amherst), 30–47.
- Oasmaa, A., and Czernik, S. (1999). Fuel oil quality of biomass pyrolysis oils state of the art for the end users. *Energy Fuel* 13, 914–921. doi:10.1021/ef980272b
- O'Neill, M. A., and York, W. S. (2009). “The plant cell wall,” in *Annual Plant Reviews*. ed. Rose J. K. C. (Blackwell: Wiley), 1–44.
- Patwardhan, P. R., Brown, R. C., and Shanks, B. H. (2011a). Understanding the fast pyrolysis of lignin. *ChemSusChem* 4, 1629–1636. doi:10.1002/cssc.201100133
- Patwardhan, P. R., Dalluge, D. L., Shanks, B. H., and Brown, R. C. (2011b). Distinguishing primary and secondary reactions of cellulose pyrolysis. *Bioresour. Technol.* 102, 5265–5269. doi:10.1016/j.biortech.2011.02.018
- Patwardhan, P. R., Satrio, J. A., Brown, R. C., and Shanks, B. H. (2010). Influence of inorganic salts on the primary pyrolysis products of cellulose. *Bioresour. Technol.* 101, 4646–4655. doi:10.1016/j.biortech.2010.01.112
- Pauly, M., Gille, S., Liu, L., Mansoori, N., De Souza, A., Schultink, A., et al. (2013). Hemicellulose biosynthesis. *Planta* 238, 627–642. doi:10.1007/s00425-013-1921-1
- Pauly, M., and Keegstra, K. (2008). Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54, 559–568. doi:10.1111/j.1365-3113X.2008.03463.x
- Penning, B. W., Sykes, R. W., Babcock, N. C., Dugard, C. K., Held, M. A., Klimek, J. F., et al. (2014a). Genetic determinants for enzymatic digestion of lignocellulosic biomass are independent of those for lignin abundance in a maize recombinant inbred population. *Plant Physiol.* 165, 1475–1487. doi:10.1104/pp.114.242446
- Penning, B. W., Sykes, R. W., Babcock, N. C., Dugard, C. K., Klimek, J. F., Gamblin, D., et al. (2014b). Validation of PyMBMS as a high-throughput screen for lignin abundance in lignocellulosic biomass of grasses. *Bioenerg. Res.* 7, 899–908. doi:10.1007/s12155-014-9410-3
- Perlack, R. D., Wright, L. L., Turhollow, A., Graham, R. L., Stokes, B. J., and Erbach, D. C. (2005). *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*. Oak Ridge, TN: Oak Ridge National Laboratory.
- Petrik, D. L., Karlen, S. D., Cass, C. L., Padmakshan, D., Lu, F., Liu, S., et al. (2014). p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in *Brachypodium distachyon*. *Plant J.* 77, 713–726. doi:10.1111/tpj.12420
- Pham, T. N., Sooknoi, T., Crossley, S. P., and Resasco, D. E. (2013). Ketoneization of carboxylic acids: mechanisms, catalysts, and implications for biomass conversion. *ACS Catal.* 3, 2456–2473. doi:10.1021/cs400501h
- Pictet, A., and Sarasin, J. (1918). Sur la distillation de la cellulose et de l'amidon sous pression réduite. *Helv. Chim. Acta* 1, 87–96. doi:10.1002/hlca.19180010109
- Pohlmann, J. G., Osorio, E., Vilela, A. C. F., Diez, M. A., and Borrego, A. G. (2014). Integrating physicochemical information to follow the transformations of biomass upon torrefaction and low-temperature carbonization. *Fuel* 131, 17–27. doi:10.1016/j.fuel.2014.04.067
- Ponder, G. R., and Richards, G. N. (1991). Mechanisms of pyrolysis of polysaccharides. 4. Thermal synthesis and pyrolysis of a xylan. *Carbohydr. Res.* 218, 143–155. doi:10.1016/0008-6215(91)84093-T
- Raisanen, U., Pitkanen, I., Halttunen, H., and Hurtt, M. (2003). Formation of the main degradation compounds from arabinose, xylose, mannose and arabinol during pyrolysis. *J. Therm. Anal. Calorim.* 72, 481–488. doi:10.1023/A:1024557011975
- Ralph, J., Grabber, J. H., and Hatfield, R. D. (1995). Lignin-ferulate cross-links in grasses – active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydr. Res.* 275, 167–178. doi:10.1016/0008-6215(95)00237-N
- Resasco, D. E. (2011). What should we demand from the catalysts responsible for upgrading biomass pyrolysis oil? *J. Phys. Chem. Lett.* 2, 2294–2295. doi:10.1021/jz201135x
- Rodriguez-Gacio Mdel, C., Iglesias-Fernandez, R., Carbonero, P., and Matilla, A. J. (2012). Softening-up mannan-rich cell walls. *J. Exp. Bot.* 63, 3976–3988. doi:10.1093/jxb/ers096
- Ronsse, F., Bai, X. L., Prins, W., and Brown, R. C. (2012). Secondary reactions of levoglucosan and char in the fast pyrolysis of cellulose. *Environ. Prog. Sustain.* 31, 256–260. doi:10.1002/ep.11633
- Saizjimeñez, C., and Deleeuw, J. W. (1986). Lignin pyrolysis products – their structures and their significance as biomarkers. *Org. Geochem.* 10, 869–876. doi:10.1016/S0166-6380(86)80024-9
- Sarkanen, K. V., Chang, H., and Allan, G. G. (1967). Species variation in lignins. 3. Hardwood lignins. *Tappi* 50, 587.
- Scheller, H. V., and Ulvskov, P. (2010). Hemicelluloses. *Annu. Rev. Plant Biol.* 61, 263–289. doi:10.1146/annurev-arplant-042809-112315
- Schultink, A., Naylor, D., Dama, M., and Pauly, M. (2015). The role of the plant-specific ALTERED XYLOGLUCAN9 protein in *Arabidopsis* cell wall polysaccharide O-acetylation. *Plant Physiol.* 167, 1271–U1243. doi:10.1104/pp.114.256479
- Scott, D. S., and Piskorz, J. (1984). The continuous flash pyrolysis of biomass. *Can. J. Chem. Eng.* 62, 404–412. doi:10.1016/j.biortech.2012.11.114
- Shafizadeh, F., McGinnis, G. D., and Philpot, C. W. (1972). Thermal degradation of xylan and related model compounds. *Carbohydr. Res.* 25, 23–33. doi:10.1016/S0008-6215(00)82742-1
- Shafizadeh, F., Furneaux, R. H., Cochran, T. G., Scholl, J. P., and Sakai, Y. (1979). Production of levoglucosan and glucose from pyrolysis of cellulosic materials. *J. Appl. Polym. Sci.* 23, 3525–3539. doi:10.1002/app.1979.070231209
- Singh, M. P., Erickson, J. E., Sollenberger, L. E., Woodard, K. R., Vendramini, J. M. B., and Fedenko, J. R. (2012). Mineral composition and biomass partitioning of sweet sorghum grown for bioenergy in the southeastern USA. *Biomass Bioenergy* 47, 1–8. doi:10.1016/j.biombioe.2012.10.022
- Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78. doi:10.1146/annurev.cellbio.22.022206.160206
- Somerville, C. (2007). Biofuels. *Curr. Biol.* 17, R115–R119. doi:10.1016/j.cub.2007.01.010
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., et al. (2004). Toward a systems approach to understanding plant cell walls. *Science* 306, 2206–2211. doi:10.1126/science.1102765
- Sykes, R., Yung, M., Novaes, E., Kirst, M., Peter, G., and Davis, M. (2009). “High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy,” in *Biofuels*, ed. Mielenz J. R. (New York City: Humana Press), 169–183.
- Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., et al. (2013). An *Arabidopsis* cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* 25, 270–287. doi:10.1105/tpc.112.107334
- Tanger, P., Field, J. L., Jahn, C. E., Defoort, M. W., and Leach, J. E. (2013). Biomass for thermochemical conversion: targets and challenges. *Front. Plant Sci.* 4:218. doi:10.3389/fpls.2013.00218
- Vanhohle, B., Cesarino, I., Goeminne, G., Kim, H., Marroni, F., Van Acker, R., et al. (2013). Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. *New Phytol.* 198, 765–776. doi:10.1111/nph.12179
- Vanhohle, R., Van Acker, R., and Boerjan, W. (2010). Potential of *Arabidopsis* systems biology to advance the biofuel field. *Trends Biotechnol.* 28, 543–547. doi:10.1016/j.tibtech.2010.07.008
- Vassilev, S. V., Baxter, D., Andersen, L. K., and Vassileva, C. G. (2010). An overview of the chemical composition of biomass. *Fuel* 89, 913–933. doi:10.1016/j.fuel.2009.10.022
- Vega-Sanchez, M. E., Verherbruggen, Y., Christensen, U., Chen, X., Sharma, V., Varanasi, P., et al. (2012). Loss of cellulose synthase-like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties, and defense responses in vegetative tissues of rice. *Plant Physiol.* 159, 56–69. doi:10.1104/pp.112.195495
- Vega-Sanchez, M. E., Verherbruggen, Y., Scheller, H. V., and Ronald, P. C. (2013). Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signal. Behav.* 8, e23143. doi:10.4161/psb.23143
- Vogel, J. (2008). Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* 11, 301–307. doi:10.1016/j.pbi.2008.03.002

- Vogel, K. P., Dien, B. S., Jung, H. G., Casler, M. D., Masterson, S. D., and Mitchell, R. B. (2011). Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses. *Bioenergy Res.* 4, 96–110. doi:10.1007/s12155-010-9104-4
- Wan, S. L., Waters, C., Stevens, A., Gumidyal, A., Jentoft, R., Lobban, L., et al. (2015). Decoupling HZSM-5 catalyst activity from deactivation during upgrading of pyrolysis oil vapors. *ChemSusChem* 8, 552–559. doi:10.1002/cssc.201402861
- Wegrzyn, J. L., Eckert, A. J., Choi, M., Lee, J. M., Stanton, B. J., Sykes, R., et al. (2010). Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, Salicaceae) secondary xylem. *New Phytol.* 188, 515–532. doi:10.1111/j.1469-8137.2010.03415.x
- Westerhof, R. J. M., Kuipers, N. J. M., Kersten, S. R. A., and Van Swaaij, W. P. M. (2007). Controlling the water content of biomass fast pyrolysis oil. *Ind. Eng. Chem. Res.* 46, 9238–9247. doi:10.1021/ie070684k
- Xiong, G. Y., Cheng, K., and Pauly, M. (2013). Xylan O-acetylation impacts xylem development and enzymatic recalcitrance as indicated by the *Arabidopsis* mutant tbl29. *Mol. Plant.* 6, 1373–1375. doi:10.1093/mp/sst014
- Yamaji, N., and Ma, J. F. (2009). A transporter at the node responsible for inter-vascular transfer of silicon in rice. *Plant Cell* 21, 2878–2883. doi:10.1105/tpc.109.069831
- Yang, F., Mitra, P., Zhang, L., Prak, L., Verhertbruggen, Y., Kim, J. S., et al. (2013). Engineering secondary cell wall deposition in plants. *Plant Biotechnol. J.* 11, 325–335. doi:10.1111/Pbi.12016
- Youngs, H., and Somerville, C. (2012). Growing better biofuel crops. *Scientist* 26, 46–52. doi:10.2134/jeq2013.05.0171
- Yuan, T. Q., Sun, S. N., Xu, F., and Sun, R. C. (2011). Characterization of lignin structures and lignin-carbohydrate complex (LCC) linkages by quantitative ¹³C and 2D HSQC NMR spectroscopy. *J. Agric. Food Chem.* 59, 10604–10614. doi:10.1021/jf2031549
- Zapata, P. A., Faria, J., Ruiz, M. P., and Resasco, D. E. (2012). Condensation/hydrogenation of biomass-derived oxygenates in water/oil emulsions stabilized by nanohybrid catalysts. *Top. Catal.* 55, 38–52. doi:10.1007/s11244-012-9768-4
- Zhang, J., Choi, Y. S., Yoo, C. G., Kim, T. H., Brown, R. C., and Shanks, B. H. (2015). Cellulose-hemicellulose and cellulose-lignin interactions during fast pyrolysis. *ACS Sustain. Chem. Eng.* 3, 293–301. doi:10.1021/sc500664h
- Zhang, X. S., Yang, G. X., Jiang, H., Liu, W. J., and Ding, H. S. (2013). Mass production of chemicals from biomass-derived oil by directly atmospheric distillation coupled with co-pyrolysis. *Sci. Rep.* 3, 1120. doi:10.1038/srep01120
- Zhao, Q., and Dixon, R. A. (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 16, 227–233. doi:10.1016/j.tplants.2010.12.005
- Zhao, Y. L., Steinberger, Y., Shi, M., Han, L. P., and Xie, G. H. (2012). Changes in stem composition and harvested produce of sweet sorghum during the period from maturity to a sequence of delayed harvest dates. *Biomass Bioenergy* 39, 261–273. doi:10.1016/j.biombioe.2012.01.020
- Zheng, A., Zhao, Z., Chang, S., Huang, Z., Wang, X., He, F., et al. (2013). Effect of torrefaction on structure and fast pyrolysis behavior of corn cobs. *Bioresour. Technol.* 128, 370–377. doi:10.1016/j.biortech.2012.10.067
- Zhong, R., and Ye, Z. H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 56, 195–214. doi:10.1093/pcp/pcu140
- Zhou, Y., Stuart-Williams, H., Farquhar, G. D., and Hocart, C. H. (2010). The use of natural abundance stable isotopic ratios to indicate the presence of oxygen-containing chemical linkages between cellulose and lignin in plant cell walls. *Phytochemistry* 71, 982–993. doi:10.1016/j.phytochem.2010.03.001
- Zhu, X. L., Lobban, L. L., Mallinson, R. G., and Resasco, D. E. (2011). Bifunctional transalkylation and hydrodeoxygenation of anisole over a Pt/Hβ catalyst. *J. Catal.* 281, 21–29. doi:10.1016/j.jcat.2011.03.030

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Lin, Waters, Mallinson, Lobban and Bartley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Potential for Genetic Improvement of Sugarcane as a Source of Biomass for Biofuels

Nam V. Hoang^{1,2*}, Agnelo Furtado¹, Frederik C. Botha^{1,3}, Blake A. Simmons^{1,4} and Robert J. Henry¹

¹ Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia, QLD, Australia,

² College of Agriculture and Forestry, Hue University, Hue, Vietnam, ³ Sugar Research Australia, Indooroopilly, QLD, Australia,

⁴ Joint BioEnergy Institute, Emeryville, CA, USA

OPEN ACCESS

Edited by:

P. C. Abhilash,
Banaras Hindu University, India

Reviewed by:

Yu-Shen Cheng,
National Yunlin University of Science
and Technology, Taiwan
Tianju Chen,
Chinese Academy of Sciences, China

*Correspondence:

Nam V. Hoang
hoang.nam@uq.net.au

Specialty section:

This article was submitted to
Bioenergy and Biofuels, a section of
the journal *Frontiers in Bioengineering
and Biotechnology*

Received: 25 July 2015

Accepted: 26 October 2015

Published: 17 November 2015

Citation:

Hoang NV, Furtado A, Botha FC,
Simmons BA and Henry RJ (2015)
Potential for Genetic Improvement
of Sugarcane as a Source
of Biomass for Biofuels.
Front. Bioeng. Biotechnol. 3:182.
doi: 10.3389/fbioe.2015.00182

Sugarcane (*Saccharum* spp. hybrids) has great potential as a major feedstock for biofuel production worldwide. It is considered among the best options for producing biofuels today due to an exceptional biomass production capacity, high carbohydrate (sugar + fiber) content, and a favorable energy input/output ratio. To maximize the conversion of sugarcane biomass into biofuels, it is imperative to generate improved sugarcane varieties with better biomass degradability. However, unlike many diploid plants, where genetic tools are well developed, biotechnological improvement is hindered in sugarcane by our current limited understanding of the large and complex genome. Therefore, understanding the genetics of the key biofuel traits in sugarcane and optimization of sugarcane biomass composition will advance efficient conversion of sugarcane biomass into fermentable sugars for biofuel production. The large existing phenotypic variation in *Saccharum* germplasm and the availability of the current genomics technologies will allow biofuel traits to be characterized, the genetic basis of critical differences in biomass composition to be determined, and targets for improvement of sugarcane for biofuels to be established. Emerging options for genetic improvement of sugarcane for the use as a bioenergy crop are reviewed. This will better define the targets for potential genetic manipulation of sugarcane biomass composition for biofuels.

Keywords: sugarcane, biofuels, biomass for biofuels, biofuel traits, association studies

INTRODUCTION

Plant biomass from grasses such as sugarcane or woody species contains mostly cellulose, hemicellulose, and lignin (also referred to as lignocellulosic biomass), which can be converted to biofuels as a source of renewable energy. At the moment, plant biomass-derived biofuels have great potential in countries that have limited oil resources because they reduce the dependence on fossil fuel, mitigate air pollution by cutting down greenhouse gas emissions, and can be produced from a wide range

Abbreviations: AFLP, amplified fragment length polymorphism; BAC, bacterial artificial chromosome; CAD, cinnamyl alcohol dehydrogenase (EC 1.1.1.195); cDNA, complementary DNA; COMT, caffeic acid O-methyltransferase (EC 2.1.1.68); DArT, diversity array technology; EST, expressed sequence tag; Gb/Mb, gigabase/megabase; LD, linkage disequilibrium; Lignin G, lignin guaiacyl; Lignin H, lignin hydroxyphenyl; Lignin S, lignin syringyl; NGS, next-generation sequencing; QTL, quantitative trait locus; RFLP, restricted fragment length polymorphism; RNAi, RNA interference; S/G ratio, syringyl/guaiacyl ratio; SNP, single nucleotide polymorphism; SUCEST, sugarcane EST database; TF, transcription factor; TIGR, the institute for genome research.

of abundant sources (Matsuoka et al., 2009). Biofuels generated from plant lignocellulosic biomass (also known as the second generation of biofuels) have been shown to be advantageous over the first generation (from plant starches, sugar, and oil) in terms of net energy and CO₂ balance and, more importantly, they do not compete with food industries for supplies (Yuan et al., 2008). To date, producing bioethanol from the sugar in sugarcane has been one of the world's most commercially successful biofuel production systems, with the potential to deliver second-generation fuels with a high positive energy balance and at a relatively low production cost (Yuan et al., 2008; Botha, 2009; Matsuoka et al., 2009). The rapid growth and high yield of sugarcane compared to other grasses and woody plants makes it a good candidate for ethanol processing platform and the second generation of biofuels in general (Pandey et al., 2000). Sugarcane has an exceptional ability to produce biomass as a C4 plant with the potential of a perennial grass crop allowing harvest four to five times by using ratoons without requiring replanting (Verheye, 2010), resulting in a lower cost of energy production from sugarcane than for most of the other potential sources of biomass (Botha, 2009). Brazil is the world's first country to launch a national fuel alcohol program (ProAlcool). This program is based on the use of sugarcane and substitutes the usage of gasoline by ethanol (Dias De Oliveira et al., 2005). Approximately, 23.4 billion liters (6.19 billion U.S. liquid gallons) of ethanol was produced in Brazil in the year 2014 (Renewable Fuels Association, 2015). As of 2009, sugarcane bagasse contributes to about 15% of the total electricity consumed in Brazil, and it is predicted that energy generated from sugarcane stalks could supply more than 30% of the country energy needs by 2020 and will be equal to or more than the electricity produced from hydropower (Matsuoka et al., 2009).

Conventionally, sugarcane bagasse is usually burned to produce fertilizer or steam and electricity to fuel the boilers in sugar mills (Pandey et al., 2000). Recently, it has been used for biofuel production; however, the production cost of biofuels from lignocellulosic biomass is still considered to be relatively high, which makes it difficult to be price-competitive and commercialized on a large scale (Halling and Simms-Borre, 2008). At the moment, the cost of bagasse pretreatment (to remove or separate its recalcitrant components before converting to biofuels) and microbial enzymes contributes mostly to the total production cost, resulting in reducing the incentive to transition from first generation to second generation of biofuels in sugarcane (Yuan et al., 2008). To maximize the efficiency of conversion of sugarcane biomass into biofuels, it is imperative to generate improved sugarcane cultivars with not only high biomass yield and fiber content but also better biomass degradability for conversion to biofuels in addition to improving the pretreatment and enzyme digestion technologies.

This review focuses on the potential for the genetic improvement of sugarcane as a source of biomass for biofuels, exploring the beneficial characteristics of sugarcane, the available genetic resources and germplasm, the potential of cell wall modification by breeding and biotechnology, and the potential of whole genome/transcriptome sequencing applications in dissecting important biofuel traits to improve sugarcane biomass composition. This will define the targets for potential genetic manipulation and better exploitation of sugarcane biomass for biofuels.

SUGARCANE AT A QUICK GLANCE

Biology and Genetics

Taxonomically, sugarcane belongs to the genus *Saccharum* (established by Carl Linnaeus in 1753), in the grass family *Poaceae* (or *Gramineae*), subfamily *Panicoideae*, tribe *Andropogoneae*, subtribe *Sacharinae*, under the group *Saccharastrae* and has a very close genetic relationship to sorghum and other grass family members such as *Erianthus* and *Miscanthus* (Amalraj and Bala-sundaram, 2006). Typically, the genus is divided into six different species namely *Saccharum barberi*, *Saccharum edule*, *Saccharum officinarum*, *Saccharum robustum*, *Saccharum sinense*, and *Saccharum spontaneum* (Daniels and Roach, 1987; Amalraj and Bala-sundaram, 2006), in which *S. spontaneum* and *S. robustum* are wild species; *S. officinarum*, *S. barberi*, and *S. sinense* are early cultivars while *S. edule* is a marginal specialty cultivar. All genotypes of the *Saccharum* genus are reported to be polyploid with the ploidy level ranging from 5× to 16× and are considered as among the most complex plant genomes (Manners et al., 2004). The cytotype (2n, the number of chromosomes in the cell) was reported to be different in each species as follows: *S. officinarum* (2n = 80), *S. spontaneum* (2n = 40–128), *S. barberi* (2n = 111–120), *S. sinense* (2n = 81–124), *S. edule* (2n = 60–80), and *S. robustum* (2n = 60, 80); hence, the basic chromosome number (x, the monoploid set of chromosomes in the cell) ranges from 5, 6, 8, 10 to 12 (Sreenivasan et al., 1987). The basic chromosome number of *S. spontaneum* is 8 (even though a number of very variable cytotype is observed) and of *S. officinarum* and *S. robustum* is 10 [Panje and Babu, 1960; D'Hont et al. (1998), and Piperidis et al. (2010)]. For the other three species, *S. sinense*, *S. barberi*, and *S. edule*, due to the fact that these are early interspecific hybrid cultivars, there have not been a consensus reported, but a study by Ming et al. (1998) suggested that the basic chromosome number for these three species could also be 10.

Hybrid sugarcane was derived from crosses between a female *S. officinarum* (2n = 80) and a male *S. spontaneum* (2n = 40–128). Due to the female restitution phenomenon, at first, the F1 hybrid conserves the whole *S. officinarum* chromosome set and half of the *S. spontaneum* which was 2n + n, then a few backcrosses later, this hybrid breaks down to n + n, establishing the hybrid chromosome set of modern sugarcane hybrid (Bremer, 1961). For this reason, current sugarcane cultivars (*Saccharum* spp. hybrids) have a combination of a highly aneuploid and interspecific set of chromosomes. By using genomic *in situ* hybridization (GISH) and fluorescent *in situ* hybridization (FISH), it is revealed that among chromosomes in the nucleus of modern hybrid sugarcane, approximately 80% are contributed by *S. officinarum*, 10–20% from *S. spontaneum*, and less than 5–17% from recombination of chromosomes of the two species (D'Hont et al., 1996; Piperidis et al., 2001; Cuadrado et al., 2004). Modern sugarcane hybrids are normally crosses between varieties/clones, which makes the combination of the chromosomes in each offspring unique and unpredictable due to the random sorting of the chromosomes in the genome (Grivet and Arruda, 2002). The first sugarcane breeding program, which started more than one century ago, generated a few interspecific hybrids and constitutes the basic germplasm used by sugarcane breeding programs (Ming et al., 2010). Modern sugarcane cultivars are derived from the basic germplasm, but

there has been only a few generations for chromosome recombination opportunities (the number of meiosis that chromosomes have undergone is mainly about 2–7) as the sugarcane breeding processes normally take between 10 and 15 years (Raboin et al., 2008; Ming et al., 2010). As a result, the modern sugarcane population has a narrow genetic basis and high linkage disequilibrium (Roach, 1989; Lima et al., 2002; Raboin et al., 2008).

The Nature of a Complex, Polyploid, and Repetitive Genome

The complex and polyploid genome of sugarcane makes the process of analyzing and understanding difficult by normal methods applied to diploid plants. The size of the sugarcane genome is about 10 Gb while its genome complexity is due to the mixture of euploid and aneuploid chromosome sets with homologous genes present in from 8 to 12 copies (Souza et al., 2011). The estimated monoploid genome size is about 750–930 Mb (the monoploid genome size of the two parental species, *S. officinarum* and *S. spontaneum*, are 930 Mb and 750 Mb, respectively), not much larger than the sorghum genome (~730 Mb) and about twice the size of the rice genome (~380 Mb) (D'Hont and Glaszmann, 2001). On the other hand, studies revealed that despite this complex and polyploid genome, sugarcane showed synteny with other grasses, especially sorghum (collinear, due to the limited divergence time) and maize (orthologous but altered loci collinearity) [reviewed in Grivet and Arruda (2002)]. It was thought that the sugarcane genome contains roughly the same amount of repetitive DNA as in the sorghum genome (Jannoo et al., 2007); however, studies on BAC-end sequences by Wang et al. (2010), Figueira et al. (2012), and Kim et al. (2013) suggested that there is less repetitive content in the sugarcane genome (e.g., 45.2% and 42.8% repetitive sequences observed in large BAC collections in comparison to 61% in the sorghum genome). More recently, using the *k*-mer approach, Berkman et al. (2014) found that the repetitive proportion in three sugarcane hybrid cultivars ranges from 63.74 to 78.37% and higher than that in the sorghum genome (55.5%) using the same approach. The authors postulated that the increased proportion could be attributed to ploidy level rather than repetitive content in the sugarcane genome. A high gene-copy number, the integration of two chromosome sets from two different species, and a significant repeat content hinder the understanding of how the genome functions and obtaining a genuine assembled monoploid genome (Souza et al., 2011; Figueira et al., 2012).

Candidate Crop for Future Biofuels

To date, sugarcane is among the most efficient crops in the world together with other C4 grasses such as switch grass (*Panicum virgatum*), *Miscanthus* species (*Miscanthus x giganteus*), and *Erianthus* species (*Erianthus arundinaceus* Retz.) in terms of converting solar energy into stored chemical energy and biomass accumulation (Tew and Cobill, 2008; Furtado et al., 2014). In general, C4 plants outperform C3 plants in biomass yield, including grain, stem, and leaf yield (Jakob et al., 2009; Wang and Pateron, 2013). Sugarcane and other C4 grasses are the highest yield potential feedstocks (Table 1), and for sugarcane, the potential yield can exceed 100 tons dry matter per hectare per year (Jakob

TABLE 1 | Average lignocellulosic biomass yield (dry matter) from sugarcane compared to other sources.

Plant name	Yield (tons/ha/year)	Reference
Sugarcane	22.9 ^a	Van Der Weijde et al. (2013)
Switch grass	7–35	Reviewed in Hattori and Morita (2010)
Miscanthus	12–40	Reviewed in Hattori and Morita (2010)
Erianthus	40–60	Reviewed in Hattori and Morita (2010)
Eucalyptus	15–40	Reviewed in Johansson and Burnham (1993)

^aAverage total cane biomass dry matter is 39 tons/ha/year (Moore, 2009).

et al., 2009; Moore, 2009; Henry, 2010a). At present, the most suitable energy crop is probably sugarcane because of its high biomass yield and the potential for production on other than prime agricultural land avoiding competing with the land used for food industries (Waclawovsky et al., 2010). Globally, sugarcane is the most important crop in about 100 countries with a production area of 26.9 million hectares, total production of ~1.9 billion tons, and yield of 70.9 tons of fresh cane per hectare (FAOSTAT, 2015). At present, Brazil is the world's largest sugarcane producer followed by India, China, Thailand, Pakistan, Mexico, Colombia, Indonesia, Philippines, U.S., and Australia (FAOSTAT, 2015). In sugarcane internodal tissue, sucrose concentration ranges from 14 to 42% of the dry weight (Whittaker and Botha, 1997), while the rest of dry biomass comes from the cell wall lignocellulose, mostly containing cellulose, hemicellulose, lignin, and ash (Pereira et al., 2015). Biofuels from sugarcane can be produced extensively not only from its soluble sugar but also from main residues in sugarcane production, bagasse and trash, on the same production area (Seabra et al., 2010; Alonso Pippo et al., 2011a,b; Macrelli et al., 2012). The total estimated available lignocellulosic biomass from sugarcane worldwide was 584 million dry tons per year, with an average lignocellulosic biomass yield of 22.9 dry tons per hectare per year (Van Der Weijde et al., 2013). Sugarcane bioethanol yield from bagasse is estimated at about 3,000 L per hectare in a total yield of 9,950 L per hectare from sugar and bagasse (Somerville et al., 2010).

AVAILABLE SUGARCANE GENETIC RESOURCES FOR BIOFUELS

Existing Variations within *Saccharum* Germplasm

Genetically diverse sugarcane germplasm may play a key role in improving sugarcane for biofuels through breeding and biotechnological approaches. Genetic variation may be found in biomass yield, fiber content, and sugar composition in the *Saccharum* germplasm. This includes the diversity among the cultivars within one species and also diversity among species within the genus. A relatively high genetic variability within sugarcane hybrid cultivars was reported thanks to their heterozygosity and high polyploidy despite their originating from a few clones of a narrow genetic base (Aitken and McNeil, 2010). There is also great genetic and morphological diversity within *Saccharum* species, *Miscanthus* species, and *Erianthus* species to be potentially exploited and incorporated to broaden the genetic base in

breeding programs (Harvey et al., 1994; Aitken and McNeil, 2010). To date, the genetic diversity of *S. officinarum* has been exploited in breeding programs; however, the diversity of *S. spontaneum* and other species have not been used much (Aitken and McNeil, 2010). *Saccharum* species have also been shown to have varied genome size, *S. officinarum* genome is about 7.50–8.55 Gb, *S. robustum* ranging from 7.56 to 11.78 Gb, and *S. spontaneum* ranging from 3.36 to 12.64 Gb, whereas the other three species – *S. sinense*, *S. barberi*, *S. edule* – and modern sugarcane are interspecific hybrids whose genome size depends upon each cross (Zhang et al., 2012).

There are two world largest collections of germplasm of *Saccharum* species, one is located in Florida (USA) while the other is in Kerala (India), containing approximately 1,200 accessions collected from 45 countries (Tai and Miller, 2001; Todd et al., 2014). These collections could be potentially selected and utilized for breeding purpose to improve sugarcane germplasm for new biofuel traits (Todd et al., 2014). The wild sugarcane species show wider variability in comparison to the domesticated species. In the *Saccharum* genus, *S. spontaneum* has the widest range of morphological variability, ratoon yielding, as well as biotic and abiotic stress tolerance (Tai and Miller, 2001; Aitken and McNeil, 2010; Govindaraj et al., 2014). The coefficient of variation (CV%) for some of the traits such as internode length, midrib width, leaf width, plant height, and stalk height studied by Govindaraj et al. (2014) were reported to be between 15 and 30%, which indicates a very high variability within the collection. It has been shown that the diversity within modern sugarcane hybrids was mostly contributed by the introgression from *S. spontaneum* (D'Hont et al., 1996). On the other hand, *S. robustum* also possesses a large amount of phenotypic variations in many traits studied (Aitken and McNeil, 2010). Sugarcane parental species (*S. officinarum*, *S. spontaneum*, and *S. robustum*), *Miscanthus* species, *Erianthus* species, and sorghum species with their diversity in genome content, structure, and tremendous allelic variation are a valuable and significant genetic reservoir which could be exploited for improving sugarcane biomass.

Genetic Markers and Maps

To support the effort of understanding the sugarcane genome, many physical maps, molecular markers, and resources such as RFLP, RAPD, AFLP, SSR, and ESTs have been developed over time. These common markers have been applied for genetic studies such as diversity, mapping, quantitative trait loci (QTL), and synteny definition; however, these systems have been developed mostly for well-established diploid species and are less effective for polyploidy plants (Garcia et al., 2013). Markers like AFLP, SSR, and RFLP are unable to estimate the number of allelic copies and level of polyploidy in complicated genomes such as potato, strawberry, and sugarcane (Garcia et al., 2013). More recently, the use of SNPs markers, which are distributed at high density across the genome, for complex genomes can allow estimation of the number of allelic copies and the ploidy level of genomes (Zhu et al., 2008; Hall et al., 2010). The currently available genetic maps and markers have been generated for sugarcane by using low-throughput methods, providing limited information on genome organization due to the low density of markers and coverage (most of them have less than 1,000 markers) (Aitken et al., 2014).

Therefore, it is difficult to allocate these markers into linkage groups or cosegregation groups or sugarcane expected chromosome number (Souza et al., 2011). More detailed linkage maps of *S. officinarum* cultivar IJ76-545 (534 markers in 123 linkage groups) and cultivar Green German (615 markers in 72 linkage groups); *S. spontaneum* cultivar IND (536 markers in 69 linkage groups); and the hybrid cultivars R570 and Q165 (with 2,000 markers placed in more than 100 linkage groups) have been constructed (Souza et al., 2011; Aitken et al., 2014). Most recently, using Diversity Array Technology (DArT), Aitken et al. (2014) integrated DArT markers, RFLPs, AFLPs, SSRs, and SNPs into the largest marker collection for sugarcane, which contains 2,467 single-dose markers for the cross between Q165 and IJ76-514 (a *S. officinarum* accession) and 2,267 markers from the cultivar Q165. These were placed into 160 linkage groups and eight homology groups, with some uncategorized linkage groups indicating that more markers are required. There is still a need to develop high-throughput marker arrays for sugarcane association studies, to generate more markers, and also to make use of the available markers. These markers will be a valuable resource in facilitating and unraveling the complex genome structure of sugarcane. It is worth considering that information on DNA-based molecular markers of progenitor plants can potentially reveal available genetic polymorphism for the analysis of their progenies (Henry et al., 2012). This could be a useful strategy in the case of sugarcane, where the genomes of the parental species are less complex than that of the hybrids.

Transcriptome Sequences and Transcription Factors

Expressed sequence tags (ESTs) and complementary DNA (cDNA) sequences provide direct evidence of the genes present in the samples, and this sequence information is very useful for genome exploration, gene prediction/discovery, genome structure identification, SNP characterization, and transcriptome and proteome analysis (Nagaraj et al., 2007). As of May 2015, the GenBank EST database (dbEST) was composed of 75,906,308 ESTs from different organisms of which 284,818 hits were detected under the search term sugarcane ("*S. officinarum*" or "*Saccharum* hybrid cultivar" or sugarcane). In the last 20 years, sugarcane ESTs have been used for gene discovery, BAC clone selection, and dissecting the coding regions of the genome, involving many projects in South Africa, France, U.S., Australia, and Brazil (Carson and Botha, 2000, 2002; Vettore et al., 2001; Casu et al., 2003, 2004; Grivet et al., 2003; Pinto et al., 2004; Bower et al., 2005). The largest collection of sugarcane ESTs was generated by SUCEST, which is composed of approximate 238,208 ESTs from 26 diverse cDNA libraries of different tissues of sugarcane cultivars, e.g., SP80-3280, SP70-1143, RB845205, RB845298, and RB805028 (Vettore et al., 2001, 2003; Souza et al., 2011). These sequences were assembled into 42,982 sugarcane assembled sequences representing more than 30,000 unique genes (~90% of the estimated genes, about 43,141, of *S. officinarum*) (Vettore et al., 2003; Hotta et al., 2010; Grassius: Grass Regulatory Information Server, 2015). There are other sugarcane EST collections containing less EST entries generated by Casu et al. (2003, 2004) (8,342 ESTs), Ma et al. (2004) (7,993 ESTs), Gupta et al. (2010) (~35,000 ESTs) and small number of ESTs by Carson and Botha (2000, 2002).

Due to the homology between genomes, genome-wide mapping of ESTs of one species provides an important framework for the genome structure of other related species (Sato et al., 2011). However, it is noteworthy that the discovery of the ESTs may be restricted to specific cultivars, as within sugarcane germplasm each cultivar has been shown to have different gene expression level [reviewed in Hotta et al. (2010)]. Moreover, for bio-fuel trait analysis, the TFs regulating monolignol biosynthesis in lignin pathway have received attention as understanding this allows reducing and modifying lignin content and composition which are essential in addressing the recalcitrant problem in biomass conversion (Santos Brito et al., 2015). It is shown that the lignin regulation can be species specific and information on TFs obtained from model plants such as *Arabidopsis* may require to be validated in other species (Santos Brito et al., 2015). A limited number of TFs in grass and sugarcane have been preliminarily characterized recently including those involved in monolignol biosynthesis, for example, in grass (Handakumbura and Hazen, 2012), rice (Yoshida et al., 2013), sorghum (Yan et al., 2013), and sugarcane (Santos Brito et al., 2015). Gene discovery of sugarcane has progressed to some extent despite the complexity of the genome. The valuable information of ESTs, TFs, full-length cDNAs, and BACs provides an understanding of allelic variations in the genome while a full-genome sequence is not available.

BAC Libraries to Construct a Reference Genome for Sugarcane

Sugarcane cultivar R570 and other cultivars including ones from the parental species *S. officinarum* and *S. spontaneum* have been used for constructing of bacterial artificial chromosome (BAC) libraries (Hotta et al., 2010). BAC libraries from the sugarcane include hybrid cultivar R570 (103,296 clones, average insert size of 130 kb and two other libraries of 100,000 clones) (Tomkins et al., 1999; Grivet and Arruda, 2002), *S. spontaneum* cultivar SES208 (38,400 clones, average insert size of 120 kb), and *S. officinarum* cultivar LA Purple (74,880 clones, average insert size of 150 kb) generated from different restriction enzymes, e.g., *HindIII* and *BamHI* [reviewed in Souza et al. (2011)]. BAC sequencing in sugarcane is currently based on the sequencing of BAC clones anchored to an available physical map. Even though it requires a higher cost compared to the whole-genome shotgun sequencing (using high-throughput platforms, Illumina, for example), it is a reliable approach for reference construction, especially, for highly repetitive genomes which cannot be sequenced and resolved by a short-read method (Eversole et al., 2009; Steuernagel et al., 2009). This BAC sequencing approach has been used successfully in sequencing of *Arabidopsis*, rice, and maize genomes and producing the barley reference genome [reviewed in Steuernagel et al. (2009)]. The ongoing Sugarcane Genome Sequencing Initiative (SUGESI) has selected 5,000 BAC clones for sequencing from a library by Tomkins et al. (1999) of cultivar R570, the most intensively characterized cultivar to date, to help assembly of the monoploid coverage (monoploid tiling path) of the sugarcane genome using the sorghum sequence as the guide (Souza et al., 2011; Sugesi, 2015).

Sorghum bicolor Genome as the Closest Related Reference Genome

Sorghum is the most closely related species to sugarcane (Grivet et al., 1994; Dillon et al., 2007). The sorghum genome sequencing project was initiated and completed in 2007 with the total genome size of ~730 Mb, and 34,496 protein-coding loci, at the coverage of 8.5× using whole-genome shotgun sequencing by standard Sanger methodologies (Paterson et al., 2009). The sequenced genome is composed of 10 pairs of chromosomes and 3,294 supercontigs (most of these have been placed into chunks on 10 chromosomes), covering 90% of the genome and 99% of protein-coding regions (including the majority of available non-repetitive markers, known sorghum protein-coding genes, and the majority of ESTs) (Paterson et al., 2009). The sorghum genome has approximately 61% repetitive DNA, a low level of gene duplication compared to other C4 grasses, and a high degree of gene parallelism with sugarcane, even though the sugarcane genome is much more polyploid (Paterson et al., 2009, 2010). Microcollinearity between sugarcane and sorghum genomes indicated that sorghum is suitable as the template for sugarcane genome assembly (Ming et al., 1998; Wang et al., 2010; Figueira et al., 2012). It has been suggested that the sugarcane genome could be 20–30% smaller than that of sorghum despite the estimated monoploid genome size of sugarcane being about 760–930 Mb, at approximately the size of the sorghum genome (Figueira et al., 2012).

BIOMASS-DERIVED BIOFUELS AND THE CHALLENGING ISSUES IN BIOMASS CONVERSION TO BIOFUELS

The Second Generation of Biofuels – Cell Walls for Fuels

Due to the depletion of fossil fuel sources, the potential for oil to become more expensive, and the raising awareness of the negative impact of fossil fuels on the environment, biomass-derived biofuels have been investigated and developed recently as an alternative source of renewable, sufficient, and clean energy (Botha, 2009). The demand for renewable biofuels is predicted to be increasing (Fedenko et al., 2013). The first generation of biofuels from plant biomass involved the process of conversion of stored polysaccharides, non-structural carbohydrates, and oils from plants (starchy, sugary, and oily parts of plants such as corn starch, sugarcane molasses, soybeans, canola seeds, and palm oil) into fuels like ethanol and diesel (Schubert, 2006; Yuan et al., 2008). However, these sources are also used as food supplies and are limited due to the increasing demand from the growing world's population (Schubert, 2006). The second generation of biofuels can be generated by using the non-food parts of plants such as cell walls, composed of structural polysaccharides, such as cellulose and hemicellulose (Schubert, 2006; Yuan et al., 2008; Henry, 2010a). This is considered to be advantageous over the first generation of biofuels as it has a higher energy production potential, lower cost, sustainable CO₂ balance, no competition with the food production, and a wide range of plant biomass sources are available at costs affordable to a biorefinery (Yuan et al., 2008; Henry, 2010a). As of 2009, sugarcane biomass as

sucrose accounted for about 40% of biofuels feedstock worldwide for first-generation biofuel production (Lam et al., 2009). Using sugarcane bagasse as a feedstock for second-generation biofuels would lead to doubling the current output of biofuel production from sugarcane (Halling and Simms-Borre, 2008).

Sugarcane Cell Wall and Biomass Composition

Physically, sugarcane biomass can be divided into four major fractions, whose content depends on the industrial process: fiber (heterogeneous organic solid fraction), non-soluble solids (inorganic substances), soluble solids (sucrose, waxes, and other chemicals), and water (Canilha et al., 2012; Shi et al., 2013). Second generation of biofuels focuses on using the fiber fraction especially the cell wall constituents of the plant to produce biofuels (Schubert, 2006; Henry, 2010a). This approach may be made more efficient by optimizing the composition of the biomass source for biofuel production. This could be achieved by advances in pretreatment methods or biotechnological modification of cell wall synthesis pathways to create a biomass that can be more efficiently processed (Sims et al., 2006; Yuan et al., 2008; Simpson, 2009; Viikari et al., 2012). Three major components make up the fiber fraction of sugarcane, namely, cellulose, hemicellulose (or non-cellulosic polysaccharide components), and lignin. Cellulose constitutes around 50% of the dry weight sugarcane bagasse while hemicellulose and lignin each account for about 25% (Loureiro et al., 2011). These three components are biosynthesized through different complex pathways (Higuchi, 1981; Whetten and Ron, 1995; Saxena and Brown, 2000; Mutwil et al., 2008; Harris and DeBolt, 2010; Pauly et al., 2013). Cellulose and hemicellulose molecules form the cell walls which act as the skeleton of plants and are strengthened by lignin and phenolic cross-linkages (Carpita, 1996; Henry, 2010b). The complex interlinking between cell wall components plays an important role in grass defense and yet challenges the biofuel production by requiring the pretreatment to separate them (De O. Buanafina, 2009).

The sugarcane and grass cell wall are categorized as type II cell wall, which differs from the type I and type III cell walls of other plants [reviewed in Souza et al. (2013)]. In general, there is little pectin, less lignin, and less structural proteins in grass cell walls than that in the non-grasses (Carpita, 1996; Henry, 2010b; Saathoff et al., 2011). There is similar cellulose content between grass and non-grass primary and secondary cell walls; however, hemicellulose composition is different between two groups. Grass cell walls have four to eight times more xylans, higher mixed linkage glucans, and lower levels of xyloglucans, mannans, glucomannans, and pectin in primary cell wall, but higher phenolics and lignin in the secondary cell wall (Loureiro et al., 2011). Grassy lignin is composed of three monolignols (lignin syringyl – S, lignin guaiacyl – G and lignin hydroxyphenyl – H subunits) forming various ratios of them and normally has more H subunit (more coumaryl derivatives) than in non-grasses (Vogel, 2008). A recent study by Bottcher et al. (2013) showed that sugarcane lignin content and composition are varied depending on tissue types and stem positions on the plant. Within one plant, the bottom internode has higher lignin accumulation than the top internode,

and the inner part of stem has higher syringyl/guaiacyl (S/G) ratio than the outer part. Polysaccharides found in sugarcane leaf and culm walls were similar but different in the proportions of xyloglucan and arabinoxylan (Souza et al., 2013). The major monosaccharides released from sugarcane cell walls were glucose, xylose, and arabinose (Loureiro et al., 2011; Rabemanolontsoa and Saka, 2013; Souza et al., 2013). Understanding the fine structure and detailed composition of sugarcane cell wall will assist in optimizing the tissue pretreatment and cell wall hydrolysis protocol. At present, converting sugarcane lignocellulosic biomass to ethanol includes (1) pretreatment to remove the lignin and other recalcitrant cellular constituents (or hemicellulose) to free cellulose, (2) enzyme-mediated action to depolymerize carbohydrates to simple sugars, and (3) fermentation of sugars and distillation of ethanol as the end product (Canilha et al., 2012).

Dealing with the Conversion Issues

Even though sugarcane biomass is less resistant to enzymatic digestion compared to that from woody plants, it is reported that biomass recalcitrant components impede the efficiency of the conversion to ethanol (Jung, 1989; Anterola and Lewis, 2002; Chen and Dixon, 2007; Himmel et al., 2007; Balat et al., 2008; Li et al., 2013). Biomass recalcitrance is caused by many factors such as the presence of epidermal and sclerenchyma tissues, vascular bundle density and arrangement, degree of lignification, heterogeneity and complexity of cell wall constituents, insoluble matter, natural inhibitors, and cellulose crystallinity (Himmel et al., 2007). Most approaches for producing biofuels from biomass at the moment rely on the disruption of the biomass, to separate lignocellulose and remove lignin in the biomass, and then conversion using microbial enzymes (Sticklen, 2006). In general, overcoming the recalcitrant issue can be addressed by physical, chemical, and genetic approaches. Physical and chemical strategies deal mainly with the pretreatment and involve loosening the cell wall structure, lowering the biomass heterogeneity, providing the enzymes access to the cellulose, cleaving the crossing linkages, and removing enzymatic inhibitors (Balat et al., 2008; Saathoff et al., 2011). To make the physical and chemical changes in plant biomass, pretreatment processing conditions must be tailored to the specific chemical and structural composition of the various and variable sources of lignocellulosic biomass (Mosier et al., 2005). Currently available physical and chemical pretreatment methods are varied and can be listed as uncatalyzed steam explosion, flow-through acid, liquid hot water, pH-controlled hot water, dilute acid, ammonia, lime and, more recently, the method using ionic liquids (Mosier et al., 2005; Shi et al., 2013; Sun et al., 2013). Genetic approaches involve genetic enhancement, molecular biology, and plant breeding efforts to improve biomass sources by having crops with less lignin, modified lignin, crops that self-produced enzymes, and crops with increased cellulose and biomass overall [reviewed in Sticklen (2006)]. The costs of the enzymatic pretreatment of cellulosic biomass (which accounts for about 25% of total processing expenses), biomass conversion, and microbial tanks limit the price-competitiveness of biofuel from lignocellulosic biomass in comparison to fossil fuel (Gnansounou and Dauriat, 2010; Macrelli et al., 2012, 2014; Van Der Weijde

et al., 2013). This emphasizes the value of genetic improvement of biomass composition to reduce processing costs.

POTENTIAL IMPROVEMENT OF SUGARCANE BY BREEDING FOR BIOFUELS

The complex and highly polyploid genome of sugarcane poses a great challenge in unraveling and studying its functions. Each cross of modern sugarcane cultivar has a unique set of chromosomes due to the random sorting of chromosomes and recombination of alleles from two progenitor species (Grivet and Arruda, 2002). There are several distinct alleles at each locus in sugarcane chromosomes, making the characteristics of the offspring unpredictable and requiring evaluation of thousands of lines from many parents to gather sufficient information in breeding programs (Matsuoka et al., 2009). In conventional breeding, after crossing and obtaining the F1 generation, hundreds of thousands of F1 seedlings are used for screening for the desired traits such as disease resistance, sugar content, agronomic characteristics, and adaptability (Matsuoka et al., 2009). The process is normally repeated for some vegetatively propagated generations to obtain the required stability of the traits. For industrial purpose, after a long process of selection, from hundreds of thousands seedlings at the beginning, breeders normally end up at a limited number of clones for release as commercial lines or cultivars.

To facilitate the second generation of biofuels, sugarcane breeding programs need to be focusing not only on important traits such as total biomass yield, sugar yield adaptability to local environment, and resistance to major pathogens but also on biofuel traits (e.g., less lignin, improve biomass composition for conversion) as a whole (Matsuoka et al., 2009; Waclawovsky et al., 2010). In sugarcane breeding, to maximize heterosis, the parents are usually selected from divergent genotypes of genetic background (Tabasum et al., 2010). Increasing sugarcane biomass yield and productivity is getting more and more difficult to achieve by conventional methods; hence, broadening the sugarcane genetic basis by introgression of its ancestors or closely related species such as *Miscanthus* and *Erianthus* is being explored in sugarcane improvement [reviewed in Dal-Bianco et al. (2012) and De Siqueira Ferreira et al. (2013)]. This is normally done by crossing *S. officinarum* and *Erianthus*, *Miscanthus*, or backcrossing the hybrids to *S. spontaneum* (Matsuoka et al., 2009). Dual-purpose cane and energy cane, sugarcane lines for lignocellulosic biomass production, have been derived from two sugarcane species, *S. spontaneum* and *S. robustum*, by crossing to develop lines with a high ability to accumulate fiber and high biomass content in addition to accumulating soluble sugars (De Siqueira Ferreira et al., 2013). Another case is *Miscane*, which was the result of crossing between *Saccharum* x *Miscanthus*. This produces cane varieties with more biomass (lignocellulose and total fermentable sugars), disease resistance, and cold tolerance. This effectively adapts *Miscanthus* to a tropical climate and expands sugarcane production to temperate, dry, and cold conditions (Alexander, 1985; Burner et al., 2009; Lam et al., 2009). Recently, using molecular markers in sugarcane breeding program (marker-assisted selection) allows the direct comparison of DNA genetic diversity and provides a precise tool in assessing

the genetic diversity of the germplasm (Tabasum et al., 2010; Berkman et al., 2012). The use of markers associated with the desired traits in combination with the advances in next-generation sequencing (NGS) technology, bioinformatics tools, and high-throughput phenotyping methods will significantly improve the sugarcane breeding programs (Lam et al., 2009). NGS will allow a great number of markers such as SNPs to be generated, which could be used to obtain a high density of marker at high coverage across the genome, to dissect the important traits they associate with. These sources of markers will be essential in breeding programs for screening of the parental plants from germplasm collection and of progenies derived from the crosses, selecting traits where the phenotypic methods are not practical (Berkman et al., 2012). High-throughput phenotyping methods will collect data from a large number of samples to overcome the small effects of genes, especially the QTL, controlling the traits (Lam et al., 2009).

POTENTIAL IMPROVEMENT BY MOLECULAR GENETICS FOR BIOFUELS

The competitiveness of biofuels over other options relies on biotechnology advancement. Efficient conversion of plant biomass to biofuels requires the supply of appropriate feedstocks that can be sustainably produced in large quantities at high yields. The efficient conversion of the biomass in these feedstocks will be facilitated by having a composition that is optimized for efficient processing to deliver high yields of the desired end products. Manipulating of the carbohydrates of the cell walls is the key of improving the biomass composition for biofuels (Harris and DeBolt, 2010). Powerful tools of biotechnology could aim to produce genetically modified sugarcane plants with a favorable ratio of cellulose to non-cellulose content; with *in planta* enzymes that can digest the biomass or degrade the lignin prior to its conversion to ethanol; with pest and disease resistance, flower inhibition, abiotic resistance; or incorporate them into elite sugarcane cultivars for better agronomic performance (Sticklen, 2006; Yuan et al., 2008; Matsuoka et al., 2009; Arruda, 2012).

Among the grasses potentially used for biofuel production such as sugarcane, switch grass, *Miscanthus*, and *Erianthus*, sugarcane has been used more for gene transformation studies (Falco et al., 2000; Manickavasagam et al., 2004; Basnayake et al., 2011) and the first transgenic sugarcane was established by Bower and Birch (1992). The current status of improving sugarcane biomass by using the genetic tools is hindered by its genome complexity, low transformation efficiency, transgene inactivation (gene silencing and regulation), somaclonal variation, and difficulty in backcrossing (Ingelbrecht et al., 1999; Hotta et al., 2010; Arruda, 2012; Dal-Bianco et al., 2012). Targets tackled so far on sugarcane include sucrose and biomass yield increase [i.e., in Ma et al. (2000) and Botha et al. (2001)], downregulation of lignin content or monolignol changes in lignin to lower biomass recalcitrance (described later), expression and accumulation of microbial cellulosic enzymes in leaf [i.e., in Harrison et al. (2011)], herbicide tolerance [i.e., in Gallo-Meagher and Irvine (1996) and Enríquez-Obregón et al. (1998)], disease or pest resistance [i.e., in Joyce et al. (1998), Arencibia et al. (1999), and Zhang et al.

(1999)], flowering inhibition [reviewed in Matsuoka et al. (2009) and Hotta et al. (2010)], and drought tolerance [i.e., in Zhang et al. (2006)]. Genetically modified sugarcane has great potential to contribute to biofuel production, with new varieties incorporating these characteristics (Arruda, 2012). Unexploited genes not only from the *Saccharum* germplasm but also in other related species, such as cold-tolerant genes in *S. spontaneum* and *Miscanthus* or drought-tolerant genes in sorghum, once identified would allow their integration into the sugarcane genome, facilitating the production of more sugarcane biomass in temperate areas or under dry conditions (Lam et al., 2009).

Increasing plant cellulose and total biomass content may be achieved by using approaches such as manipulation of growth regulators or key nutrients, increasing the ability of the plant to fix carbon by increasing atmospheric CO₂ and also manipulating some key metabolic enzymes in biomass synthesis pathways [reviewed in Sticklen (2006)]. Reduction of the cross-links of the maize cell walls (including ferulate and diferulate cross-links; benzyl ether and ester cross-links) has been shown to increase the initial hydrolysis of its cell wall polysaccharides by up to 46% (Grabber, 2005). In general, selection of grasses with less ferulate cross-linking or potent microbial xylanases by breeding or engineering tools are more attractive than pretreatment of the cell wall with a feruloyl esterase (Grabber, 2005).

Lignin content accounts for about 25% of sugarcane total lignocellulosic biomass and is probably the main obstacle affecting the efficiency of saccharification during conversion to ethanol (Canilha et al., 2012, 2013). Lignin and other recalcitrant components in cell walls prevent cellulase accessing the cellulose molecules and need to be removed before further processing (Sticklen, 2006). Lignin biosynthesis pathways are complicated and at least 10 different enzymes have been found involved in the lignin pathway in sugarcane (Higuchi, 1981; Whetten and Ron, 1995) and a total of 28 unigenes associated with monolignol biosynthesis were identified in sugarcane using SUCEST database and annotated genes from closely related species such as sorghum, maize, and rice (Bottcher et al., 2013). Tailoring sugarcane biomass composition for biofuels can be achieved by manipulating some of the key genes in lignin pathway (downregulation of some key enzymes), mostly targeting genes which encode the terminal enzymes such as caffeic acid O-methyltransferase (COMT) and cinnamyl alcohol dehydrogenase (CAD), to minimize the impact of the modifications on growth and development of the plant [as reviewed in Sticklen (2006), Jung et al. (2012), and Furtado et al. (2014)]. Not only lignin content but also the lignin S/G ratio is a very important aspect to consider in terms of modifying the lignin content because these two are both associated with biomass recalcitrance (Chen and Dixon, 2007; Li et al., 2010). Sugarcane lignin content was reduced by 3.9–13.7% using RNA interference (RNAi) suppression to downregulate the COMT gene [which has at least 31 different ESTs involved (Ramos et al., 2001)] by 67–97% and at the same time, the lignin S/G ratio was reduced from 1.47 to 1.27–0.79 (Jung et al., 2012). This resulted in an increase of up to 29% in total sugar yield without pretreatment (34% with dilute acid pretreatment). This study suggests that RNAi-mediated gene suppression is a promising method for suppression of target genes not only in lignin pathway but also for cell

wall constituent biosynthesis (Jung et al., 2012; Bottcher et al., 2013).

Producing enzymes *in planta* is another way to cut the cost of biofuel production as it reduces the expense of enzymes and enzyme treatment. Cellulase has been produced within the plant (in the apoplast) of *Arabidopsis*, rice, and maize without effects on the growth and development of the host plants [reviewed in Sticklen (2006)]. *In planta* enzyme expression in sugarcane is still in its infancy; however, a high-yield biofuel plant such as sugarcane must be a target for the production of enzymes within the biomass. Recombinant protein enzymes have been targeted to organelles such as chloroplasts, vacuoles, and the endoplasmic reticulum to separate the enzymes produced and their substrates (Harrison et al., 2011). In sugarcane, thanks to its well-established transformation methods via *Agrobacterium*, the expression of enzymes in leaves and other tissues is feasible (Manickavasagam et al., 2004; Taylor et al., 2008). Endoglucanases and exoglucanases have been overexpressed in sugarcane leaves by using the maize *PepC* promoter achieving an accumulation level of 0.05% of total soluble proteins (endoglucanase, in chloroplast) and less of exoglucanases without altering the phenotype (Harrison et al., 2011). In the future, enzymes might be synthesized in specific energy cane plants that could be coprocessed with other biomass sources from sugarcane for sugar and biomass production (e.g., bagasse from sugar mills) (Arruda, 2012).

POTENTIAL OF SUGARCANE WHOLE GENOME AND TRANSCRIPTOME SEQUENCING FOR BIOFUELS

The advent of NGS technology and a sharp reduction in per-base cost in the past decade [as reviewed in Van Dijk et al. (2014)] allows us to sequence the whole genome of a species, even a complex genome such as sugarcane, at a relatively low price within a relatively short time. At present, the cost of sequencing of a human genome at 30× coverage using the latest Illumina's HiSeq X is around US \$1,000. Since the first plant genome was completely sequenced (*Arabidopsis thaliana* in 2000) using the traditional Sanger sequencing platform, the sequencing strategies have moved to high-throughput and cost-effective approaches (Henry et al., 2012). High-throughput genome sequencing platforms have recently advanced and facilitated improved genotyping, allowing huge data output to be generated for polymorphism detection (especially SNPs) and marker discovery.

Potential Strategies in Dissection of Biofuel Traits in Sugarcane

At present, a whole-genome sequence of sugarcane is not available to support its biofuel trait analysis. However, a strategy to overcome this using the currently available resources, for dissecting biofuel traits, for example, in sugarcane biomass, is to carry on the association studies, in which a population of genetic variability is selected, phenotyped, and genotyped. Association studies use the molecular markers from the genetic variability to detect the association between markers and traits of interest in order to validate the location of the genes, especially for

quantitative traits (Huang et al., 2010). This strategy has been used for human and animal genetic studies since it was first established and more recently also for plants. To date, association studies have been applied successfully to many different plants including *Arabidopsis*, wheat, barley, rice, cotton, maize, potato, soybean, sugar beet, *Pinus*, *Eucalyptus*, ryegrass [also Zhu et al. (2008); for a review, see Hall et al. (2010)], and sugarcane (Aitken et al., 2005; Wei et al., 2006) for important traits like pathogen resistance, flowering time, grain composition, and quality. Association studies differ from traditional QTL studies, where in QTL analysis the linkage disequilibrium between markers and QLTs from a segregating population is established in a cross of different genotypes, whereas in association studies a non-structured population is used (Neale and Savolainen, 2004; Ingvarsson and Street, 2011). Therefore, association studies investigate variations of the whole population not just variations between parents. Association studies analyze the direct linkage disequilibrium between genetic markers and traits to overcome the limitations of the traditional QTL in sample size, low variation, and recombination in the population (Ingvarsson and Street, 2011). In sugarcane, association studies are a powerful method for understanding the complex traits which are controlled by many loci and dosage effects (i.e., Ming et al., 2001; Wei et al., 2006; Banerjee et al., 2015). In general, association studies involve population selection, phenotyping, genotyping, population structure, and statistical testing for the association. For these, there is a requirement to have a population with genetic variability and high linkage disequilibrium; and for sugarcane, the most important aspect of doing association studies is having marker data and a breeding population of elite varieties (Huang et al., 2010). Due to the limited number of generations, low recombination rate between chromosomes, and strong founder effect, it is expected that sugarcane has an extensive linkage disequilibrium despite the large number of chromosomes and being an outcrossing species (Huang et al., 2010). In fact, attaining a F₂ population (such as inbred backcrosses or recombinant inbred lines and double haploid lines) in sugarcane is not practical due to its clonal propagation, high heterozygosity, and inbreeding depression (Aitken and McNeil, 2010; Sreedhar and Collins, 2010). Therefore, more commonly, a segregating F₁ population from biparental crosses or self-pollinated progenies from heterozygous parents (as the *pseudo* F₂ population) are used, and hence, most of sugarcane linkage maps (as AFLP, RAPD, isozyme, and SSR) were developed on this type of F₁ population (Sreedhar and Collins, 2010). To date, most of these maps have low coverage and a limited number of markers because of the genome complexity and high cost of marker generation (Aitken et al., 2014). The high redundancy of the chromosomes in the sugarcane genome implies that with conventional approaches only the single-dose markers (present on only one of the homologous/homoeologous haplotype) can be used to obtain a high-resolution mapping (Hoarau et al., 2002; Le Cunff et al., 2008).

The potential applications of the current genotyping technologies to sugarcane association studies employ both whole-genome sequencing and whole transcriptome sequencing technologies. Genotyping is normally either by analysis candidate genes or genome-wide approaches, in which the candidate gene approach

is restricted to genes which are likely thought to be associated with traits of interest based on prior knowledge (Hirschhorn and Daly, 2005; Ingvarsson and Street, 2011). At present, whole-genome sequencing based on the random sequencing of fragments of whole genomic DNA has been successfully applied to medium-size genomes with limited amount of repetitive elements, genome resequencing with the guide of a reference sequence, or *de novo* assembly of small genomes (Steuernagel et al., 2009; Henry et al., 2012; Xu et al., 2012; Edwards et al., 2013). The large genome size of sugarcane is partially attributable to sugarcane being a polyploid and the genome having a significant amount of repetitive sequences (Berkman et al., 2014). As a result, the current short reads generated from NGS technologies cannot resolve completely the challenges in the sugarcane genomes. For highly repetitive genomes, the genomic complexity will be lost or reduced by using the *de novo* assembly approaches of NGS-derived short reads as the identical repeat sequences in the genome will be collapsed (Green, 2002). Therefore, it is required to develop efficient genotyping strategies using whole-genome sequencing data for sugarcane system to overcome the challenges. Moreover, whole transcriptome sequencing gives details of the entire transcript expressed in the samples across the whole genome and could be applicable to the sugarcane genome in identifying biological significant variations (SNPs) between different developmental stages, between varieties, or for transcripts *de novo* assembly and gene discovery (Henry et al., 2012).

For large and polyploid genomes, there are still requirements to enrich the genomic DNA to capture the coding regions to ensure the depth of coverage, resolve the variable short reads, and lessen the effect of repetitive sequences in the genome on discovery of polymorphisms (Bundock et al., 2012; Henry et al., 2012). Selective sequencing of genomic loci of interest (genes or exomes) can reduce the cost compared to whole-genome sequencing and therefore simplify the data interpretation since non-coding regions are not abundant in the data. The enrichment techniques can be hybrid capture (e.g., Agilent SureSelect, NimbleGen, FlexGen) or selective circularization (e.g., Selector probes) or PCR amplification (e.g., Raindance). Hybrid capture supported by a microarray platform has been applied to sugarcane and other complex genomes due to its high capacity to enrich large regions of interest (1–50 Mb), the possibility of multiplexing, the availability of kits, and a the small amount of input DNA required (<1–3 µg) (Mertes et al., 2011). This approach uses a selection library of fragmented DNA or RNA representing the targets (normally oligonucleotides from 80 to 180 bases produced from known information such as gene indices, ESTs) to capture the cDNA fragments from a shotgun DNA library based on the hybridization and then sequence the captured fragments (Mertes et al., 2011; Bundock et al., 2012). Bundock et al. (2012) conducted the solution-based hybridization (Agilent SureSelect) to capture the exome regions of sugarcane using sorghum and sugarcane coding probes, enriched the genome 10–11 folds, and detected 270,000–280,000 SNPs in each genotype of the material tested. At the moment, a great number of SNPs from a genome or haplotype can be generated by using high-capacity genome sequencing instruments or high-density oligonucleotide arrays (Zhu et al., 2008). The continuous advancement in genotyping technology

allows generation of up to 1 million SNPs spanning across the entire genome in one reaction (e.g., using SNP chip), and the newest SNP chip can measure the copy number as well as the allelic variation. Examples of available platforms are Affymetrix (e.g., Affymetrix Genome-Wide Human SNP Array 6.0) and Illumina (e.g., Illumina's WGGT Infinium BeadChips). Due to the multiple chromosomes in the homologous groups of sugarcane genome and the number of alleles at each locus (and the SNPs numbers consequently), an allele would likely to be defined by a combination of SNPs, not just a single SNP (McIntyre et al., 2006, 2015). SNP genotyping including SNP calling and statistical methods to estimate the ploidy level and allele dosage within homologous groups have been developed for sugarcane by Garcia et al. (2013) to allow in-depth association analysis of the genome. In this study, SNPs were developed by SEQUENOM iPLEX MassARRAY and capture primers and then discovered by QualitySNP software, mass-based procedures, and the SuperMASSA software. For whole transcriptome sequencing, Cardoso-Silva et al. (2014) identified 5,106 SSRs and 708,125 SNPs from the unigenes assembled from RNA-seq data of contrasting sugarcane varieties. These advances in sugarcane genotyping technology, together with well-developed high-throughput phenotyping methods for biofuel traits [reviewed in Lupoi et al. (2013) and Lupoi et al. (2015)] and bioinformatics tools, could accelerate sugarcane analysis while a reference genome is not available.

Some of the association studies have been carried out on sugarcane recently such as those for QTLs which control the *Pachymetra* root rot and brown rust resistance on 154 genotypes (McIntyre et al., 2005); genetics of root rot, leaf scald, Fiji leaf gall, cane sugar, and yield using 1,068 AFLP, 141 SRR (on 154 genotypes), and 1,531 DArT markers (on 480 genotypes) (Wei et al., 2006, 2010); smut and *eldana* stalk borer using 275 RFLP and 1,056 AFLP markers on 77 genotypes (Butterfield, 2007); resistance to sugarcane yellow leaf virus using 3,949 polymorphic markers (DArT and AFLP) on 189 genotypes (Debibakas et al., 2014); markers agro-morphological traits, sugar yield disease resistance, and bagasse content using 3,327 DArT, AFLP, and SSR markers on 183 genotypes (Gouy et al., 2015); and sucrose and yield contributing traits using 989 SSR markers on 108 genotypes (Banerjee et al., 2015). Using the Affymetrix GeneChip Sugarcane Genome Array, Casu et al. (2007) identified 119 transcripts associated with enzymes of cell wall metabolism and development on sugarcane variety Q177. These promising preliminary studies were carried out on small sample sizes and limited numbers of markers (even though a small number of significant associations have been identified) while the polyploid sugarcane genome and small effect of quantitative traits requires larger sample sizes and more markers (e.g., genome-wide markers) so that significant association can be detected (Huang et al., 2010; Gouy et al., 2015).

The Reference Sequence Matters

As mentioned earlier, construction of a sugarcane nuclear genome reference sequence is an important objective, even though it might take some time to finish. However, in the meantime, sugarcane genome analysis still can exploit the currently available genetic

resources such as the sorghum gene indices (sorghum gene models), sugarcane gene indices (DFCI Sugarcane Gene Index version 3.0, an integrated collection of sugarcane ESTs, complete cDNA sequences, non-redundant data of all sugarcane genes and their related information), transcription factors (TFs), and sugarcane tentative consensus/assembled sequences. For example, in the study mentioned earlier, Bundock et al. (2012), based on the gene sequences in the sorghum genome and sugarcane gene indices, captured the exomic regions of two sugarcane genotypes Q165 and IJ76-514, detected SNPs present in 13,000–16,000 targeted genes from Illumina short read data of these samples, and 87–91% of SNPs were validated and confirmed by 454 sequencing. For transcript profiling, the reference transcriptome sequence can be constructed for specific tissues using *de novo* assembly such as in Vargas et al. (2014) and Cardoso-Silva et al. (2014) and validated to find suitable reference gene sets to be used for gene expression normalization as in Guo et al. (2014). The currently available resources, on the other hand, are also utilized. Park et al. (2015) used the Sugarcane Assembled Sequences from SUCEST-FUN database as reference sequences in a study on cold-responsive gene expression profiling of sugarcane hybrids and *S. spontaneum* and found that more than 600 genes are differentially expressed in each genotype after applying stress.

CONCLUSION

Sugarcane has been shown to be a good candidate for use as a lignocellulosic biomass feedstock for second-generation biofuel production. However, its genome complexity still remains a great bottleneck restricting the dissection of biofuel traits. The most significant achievements in improving sugarcane biomass for biofuels so far have been the establishment of the high fiber cane varieties to generate more lignocellulosic biomass, and preliminary results in modifying biomass with more cellulose, less lignin content, a preferable lignin S/G ratio, and enzyme expressed in *planta* (in leaves) for easy conversion to biofuels. The improvement of sugarcane biomass has been by traditional breeding, molecular genetics approaches and, more recently, accelerated with the use of NGS technology. The future of second-generation biofuel production using sugarcane lignocellulosic biomass will depend greatly on advances in understanding of the key biofuel traits required to deliver more efficient and price-competitive biofuels. This objective will be facilitated once the whole genome of sugarcane is fully sequenced. Optimizing sugarcane lignocellulosic bagasse composition may result in biomass with better digestibility, modified carbohydrates, and reduction of cross-linking or self-produced enzymes (*in planta*). Currently available sugarcane genetic resources include diverse germplasm in the genus *Saccharum*, genetic markers and maps, ESTs, and the sequence of a closely related species genome. However, novel strategies need to be developed to overcome the challenges posed by the complex genetics. Traditional approaches using breeding and molecular genetics have potential for wider use improving sugarcane while the advent of NGS technology and high-throughput phenotyping technologies will accelerate the process of dissection of biofuel traits, genome-wide. By using these approaches, the loci of interest will be defined for use to improve sugarcane

biomass. Once a better understanding of the genes controlling cell wall biosynthesis is achieved, breeding programs will be able to accelerate the selection and development of varieties with optimized biomass composition to generate better sugarcane biomass sources to meet the demand of biofuel production.

AUTHOR CONTRIBUTIONS

NVH wrote the paper. AF, FCB, BAS, and RJH discussed and edited the manuscript. All authors read and approved the final manuscript.

REFERENCES

- Aitken, K., Jackson, P., and McIntyre, C. (2005). A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theor. Appl. Genet.* 110, 789–801. doi:10.1007/s00122-004-1813-7
- Aitken, K., and McNeil, M. (2010). "Diversity analysis," in *Genetics, Genomics and Breeding of Sugarcane*, eds R. Henry and C. Kole (Enfield, NH: Science Publishers), 19–42.
- Aitken, K. S., McNeil, M. D., Hermann, S., Bundock, P. C., Kilian, A., Heller-Uszynska, K., et al. (2014). A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput diversity array technology (DArT) markers. *BMC Genomics* 15:152. doi:10.1186/1471-2164-15-152
- Alexander, A. G. (1985). *The Energy Cane Alternative*. Amsterdam: Elsevier Science Publishers BV.
- Alonso Pippo, W., Luengo, C. A., Alonsoamador Morales Alberteris, L., Garzone, P., and Cornacchia, G. (2011a). Energy recovery from sugarcane-trash in the light of 2nd generation biofuel. Part 2: socio-economic aspects and techno-economic analysis. *Waste Biomass Valorization* 2, 257–266. doi:10.1007/s12649-011-9069-3
- Alonso Pippo, W., Luengo, C. A., Alonsoamador Morales Alberteris, L., Garzone, P., and Cornacchia, G. (2011b). Energy recovery from sugarcane-trash in the light of 2nd generation biofuels. Part 1: current situation and environmental aspects. *Waste Biomass Valorization* 2, 1–16. doi:10.1007/s12649-010-9048-0
- Amalraj, V. A., and Balasundaram, N. (2006). On the taxonomy of the members of 'Saccharum complex'. *Genet. Resour. Crop Evol.* 53, 35–41. doi:10.1007/s10722-004-0581-1
- Anterola, A. M., and Lewis, N. G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221–294. doi:10.1016/S0031-9422(02)00211-X
- Arencibia, A., Carmona, E., Cornide, M., Castiglione, S., O'reilly, J., Chinae, A., et al. (1999). Somaclonal variation in insect-resistant transgenic sugarcane (*Saccharum* hybrid) plants produced by cell electroporation. *Transgenic Res.* 8, 349–360. doi:10.1023/A:1008900230144
- Arruda, P. (2012). Genetically modified sugarcane for bioenergy generation. *Curr. Opin. Biotechnol.* 23, 315–322. doi:10.1016/j.copbio.2011.10.012
- Balat, M., Balat, H., and Öz, C. (2008). Progress in bioethanol processing. *Prog. Energy Combust. Sci.* 34, 551–573. doi:10.1016/j.pecs.2007.11.001
- Banerjee, N., Siraree, A., Yadav, S., Kumar, S., Singh, J., Kumar, S., et al. (2015). Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica* 205, 185–201. doi:10.1007/s10681-015-1422-3
- Basnayake, S. W. V., Moyle, R., and Birch, R. G. (2011). Embryogenic callus proliferation and regeneration conditions for genetic transformation of diverse sugarcane cultivars. *Plant Cell Rep.* 30, 439–448. doi:10.1007/s00299-010-0927-4
- Berkman, P. J., Bundock, P. C., Casu, R. E., Henry, R. J., Rae, A. L., and Aitken, K. S. (2014). A survey sequence comparison of *Saccharum* genotypes reveals allelic diversity differences. *Trop. Plant Biol.* 7, 71–83. doi:10.1007/s12042-014-9139-3
- Berkman, P. J., Lai, K., Lorenc, M. T., and Edwards, D. (2012). Next-generation sequencing applications for wheat crop improvement. *Am. J. Bot.* 99, 365–371. doi:10.3732/ajb.1100309
- Botha, F. (2009). "Energy yield and cost in a sugarcane biomass system," in: *Proceedings Australian Society Sugarcane Technologists*, 1–9.
- Botha, F., Sawyer, B., and Birch, R. (2001). "Sucrose metabolism in the culm of transgenic sugarcane with reduced soluble acid invertase activity," in *Proceedings of the International Society of Sugar Cane Technologists*, 588–591.
- Bottcher, A., Cesarino, I., Santos, A. B. D., Vicentini, R., Mayer, J. L. S., Vanholme, R., et al. (2013). Lignification in sugarcane: biochemical characterization, gene discovery, and expression analysis in two genotypes contrasting for lignin content. *Plant Physiol.* 163, 1539–1557. doi:10.1104/pp.113.225250
- Bower, N. I., Casu, R. E., Maclean, D. J., Reverter, A., Chapman, S. C., and Manners, J. M. (2005). Transcriptional response of sugarcane roots to methyl jasmonate. *Plant Sci.* 168, 761–772. doi:10.1016/j.plantsci.2004.10.006
- Bower, R., and Birch, R. G. (1992). Transgenic sugarcane plants via microprojectile bombardment. *Plant J.* 2, 409–416. doi:10.1111/j.1365-313X.1992.00409.x
- Bremer, G. (1961). Problems in breeding and cytology of sugar cane. *Euphytica* 10, 59–78. doi:10.1007/BF00037206
- Bundock, P. C., Casu, R. E., and Henry, R. J. (2012). Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid crop plant. *Plant Biotechnol. J.* 10, 657–667. doi:10.1111/j.1467-7652.2012.00707.x
- Burner, D. M., Tew, T. L., Harvey, J. J., and Belesky, D. P. (2009). Dry matter partitioning and quality of *Miscanthus*, *Panicum*, and *Saccharum* genotypes in Arkansas, USA. *Biomass Bioenergy* 33, 610–619. doi:10.1016/j.biombioe.2008.10.002
- Butterfield, M. K. (2007). *Marker Assisted Breeding in Sugarcane: A Complex Polyploidy*. PhD Thesis, Univ Stellenbosch, Matieland, Stellenbosch, South Africa.
- Canilha, L., Kumar Chandel, A., Dos Santos Milessi, T. S., Fernandes Antunes, F. A., Da Costa Freitas, W. L., Das Gracas Almeida Felipe, M., et al. (2012). Bioconversion of sugarcane biomass into ethanol: an overview about composition, pretreatment methods, detoxification of hydrolysates, enzymatic saccharification, and ethanol fermentation. *J. Biomed. Biotechnol.* 2012, 989572. doi:10.1155/2012/989572
- Canilha, L., Rodrigues, R., Antunes, F. A. E., Chandel, A. K., Milessi, T., Felipe, M. D. G. A., et al. (2013). "Bioconversion of hemicellulose from sugarcane biomass into sustainable products," in *Sustainable Degradation of Lignocellulosic Biomass-Techniques, Applications and Commercialization*, eds A. K. Chandel and S. S. da Silva (Rijeka: InTech), 15–45. doi:10.5772/53832
- Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., et al. (2014). *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS ONE* 9:e88462. doi:10.1371/journal.pone.0088462
- Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 445–476. doi:10.1146/annurev.arplant.47.1.445
- Carson, D., and Botha, F. (2002). Genes expressed in sugarcane maturing internodal tissue. *Plant Cell Rep.* 20, 1075–1081. doi:10.1007/s00299-002-0444-1
- Carson, D. L., and Botha, F. C. (2000). Preliminary analysis of expressed sequence tags for sugarcane. *Crop Sci.* 40, 1769–1779. doi:10.2135/cropsci2000.4061769x
- Casu, R., Dimmock, C., Chapman, S., Grof, C. L., McIntyre, C. L., Bonnett, G., et al. (2004). Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. *Plant Mol. Biol.* 54, 503–517. doi:10.1023/B:PLAN.0000038255.96128.41
- Casu, R., Grof, C. L., Rae, A., McIntyre, C. L., Dimmock, C., and Manners, J. (2003). Identification of a novel sugar transporter homologue strongly expressed

ACKNOWLEDGMENTS

We are grateful to the Australian Agency for International Development (AusAID) for financial support through an Australian Development Scholarship (ADS) to NVH and to the Queensland Government for funding this research. This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

- in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol. Biol.* 52, 371–386. doi:10.1023/A:1023957214644
- Casu, R. E., Jarney, J. M., Bonnett, G. D., and Manners, J. M. (2007). Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. *Funct. Integr. Genomics* 7, 153–167. doi:10.1007/s10142-006-0038-z
- Chen, F., and Dixon, R. A. (2007). Lignin modification improves fermentable sugar yields for biofuel production. *Nat. Biotechnol.* 25, 759–761. doi:10.1038/nbt1316
- Cuadrado, A., Acevedo, R., Moreno Diaz De La Espina, S., Jouve, N., and De La Torre, C. (2004). Genome remodelling in three modern *S. officinarum* *S. spontaneum* sugarcane cultivars. *J. Exp. Bot.* 55, 847–854. doi:10.1093/jxb/erh093
- Dal-Bianco, M., Carneiro, M. S., Hotta, C. T., Chapola, R. G., Hoffmann, H. P., Garcia, A. A. F., et al. (2012). Sugarcane improvement: how far can we go? *Curr. Opin. Biotechnol.* 23, 265–270. doi:10.1016/j.copbio.2011.09.002
- Daniels, J., and Roach, B. T. (1987). “Chapter 2 – taxonomy and evolution,” in *Developments in Crop Science*, ed. J. H. Don (Amsterdam: Elsevier), 7–84.
- De O. Buana, M. M. (2009). Feruloylation in grasses: current and future perspectives. *Mol. Plant* 2, 861–872. doi:10.1093/mp/ssp067
- De Siqueira Ferreira, S., Nishiyama, M. Y. Jr., Paterson, A. H., and Souza, G. M. (2013). Biofuel and energy crops: high-yield Saccharinae late center stage in the post-genomics era. *Genome Biol.* 14, 210. doi:10.1186/gb-2013-14-6-210
- Debabakas, S., Rocher, S., Garsmeur, O., Toubi, L., Roques, D., D'Hont, A., et al. (2014). Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide association. *Theor. Appl. Genet.* 127, 1719–1732. doi:10.1007/s00122-014-2334-7
- D'Hont, A., and Glaszmann, J. C. (2001). “Sugarcane genome analysis with molecular markers, a first decade of research,” in *Proceedings of International Society of Sugar Cane Technologists*, 556–559.
- D'Hont, A., Grivet, L., Feldmann, P., Rao, S., Berding, N., and Glaszmann, J. C. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet.* 250, 405–413. doi:10.1007/s004380050092
- D'Hont, A., Ison, D., Alix, K., Roux, C., and Glaszmann, J. C. (1998). Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41, 221–225. doi:10.1139/g98-023
- Dias De Oliveira, M. E., Vaughan, B. E., and Rykiel, E. J. (2005). Ethanol as fuel: energy, carbon dioxide balances, and ecological footprint. *Bioscience* 55, 593–602. doi:10.1641/0006-3568(2005)055[0593:EAFEDC]2.0.CO;2
- Dillon, S. L., Shapter, F. M., Henry, R. J., Cordeiro, G., Izquierdo, L., and Lee, L. S. (2007). Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (*Andropogoneae*). *Ann. Bot.* 100, 975–989. doi:10.1093/aob/mcm192
- Edwards, D., Batley, J., and Snowdon, R. J. (2013). Accessing complex crop genomes with next-generation sequencing. *Theor. Appl. Genet.* 126, 1–11. doi:10.1007/s00122-012-1964-x
- Enriquez-Obregón, G. A., Vázquez-Padrón, R. I., Prieto-Samsonov, D. L., De La Riva, G. A., and Selman-Housein, G. (1998). Herbicide-resistant sugarcane (*Saccharum officinarum* L.) plants by *Agrobacterium*-mediated transformation. *Planta* 206, 20–27. doi:10.1007/s004250050369
- Eversole, K., Graner, A., and Stein, N. (2009). “Wheat and barley genome sequencing,” in *Genetics and Genomics of the Triticeae*, eds G. J. Muehlbauer and C. Feuillet (New York, NY: Springer), 713–742.
- Falco, M. C., Tulmann Neto, A., and Ulian, E. C. (2000). Transformation and expression of a gene for herbicide resistance in a Brazilian sugarcane. *Plant Cell Rep.* 19, 1188–1194. doi:10.1007/s002990000253
- FAOSTAT. (2015). Available at: <http://faostat3.fao.org/home/E> [accessed May 25, 2015].
- Fedenko, J., Erickson, J., Woodard, K., Sollenberger, L., Vendramini, J. B., Gilbert, R., et al. (2013). Biomass production and composition of perennial grasses grown for bioenergy in a subtropical climate across Florida, USA. *BioEnergy Res.* 6, 1082–1093. doi:10.1007/s12155-013-9342-3
- Figueira, T., Okura, V., Rodrigues Da Silva, F., Jose Da Silva, M., Kudrna, D., Ammiraju, J. S. et al. (2012). A BAC library of the SP80-3280 sugarcane variety (*Saccharum* sp.) and its inferred microsynteny with the *Sorghum* genome. *BMC Res. Notes* 5:185. doi:10.1186/1756-0500-5-185
- Furtado, A., Lupoi, J. S., Hoang, N. V., Healey, A., Singh, S., Simmons, B. A., et al. (2014). Modifying plants for biofuel and biomaterial production. *Plant Biotechnol. J.* 12, 1246–1258. doi:10.1111/pbi.12300
- Gallo-Meagher, M., and Irvine, J. E. (1996). Herbicide resistant transgenic sugarcane plants containing the bar gene. *Crop Sci.* 36, 1367–1374. doi:10.2135/cropsci1996.0011183X003600050047x
- Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., et al. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep.* 3, 3399. doi:10.1038/srep03399
- Gnansounou, E., and Dauriat, A. (2010). Techno-economic analysis of lignocellulosic ethanol: a review. *Bioresour. Technol.* 101, 4980–4991. doi:10.1016/j.biortech.2010.02.009
- Gouy, M., Rousselle, Y., Thong Chane, A., Anglade, A., Royaert, S., Nibouche, S., et al. (2015). Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica* 202, 269–284. doi:10.1007/s10681-014-1294-y
- Govindaraj, P., Amalraj, V. A., Mohanraj, K., and Nair, N. V. (2014). Collection, characterization and phenotypic diversity of *Saccharum spontaneum* L. from arid and semi arid zones of northwestern India. *Sugar Tech* 16, 36–43. doi:10.1007/s12355-013-0255-4
- Grabber, J. H. (2005). How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. *Crop Sci.* 45, 820. doi:10.2135/cropsci2004.0191
- Grassius: Grass Regulatory Information Server. (2015). Available at: <http://grassius.org/> [accessed May 25, 2015].
- Green, P. (2002). Whole-genome disassembly. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4143–4144. doi:10.1073/pnas.082095999
- Grivet, L., and Arruda, P. (2002). Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr. Opin. Plant Biol.* 5, 122–127. doi:10.1016/S1369-5266(02)00234-0
- Grivet, L., Dhont, A., Dufour, P., Hamon, P., Roques, D., and Glaszmann, J. C. (1994). Comparative genome mapping of sugar-cane with other species within the *Andropogoneae* tribe. *Heredity* 73, 500–508. doi:10.1038/hdy.1994.148
- Grivet, L., Glaszmann, J. C., Vincentz, M., Da Silva, F., and Arruda, P. (2003). ESTs as a source for sequence polymorphism discovery in sugarcane: example of the *Adh* genes. *Theor. Appl. Genet.* 106, 190–197. doi:10.1007/s00122-002-1075-1
- Guo, J., Ling, H., Wu, Q., Xu, L., and Que, Y. (2014). The choice of reference genes for assessing gene expression in sugarcane under salinity and drought stresses. *Sci. Rep.* 4, 7042. doi:10.1038/srep07042
- Gupta, V., Raghuvanshi, S., Gupta, A., Saini, N., Gaur, A., Khan, M. S., et al. (2010). The water-deficit stress- and red-rot-related genes in sugarcane. *Funct. Integr. Genomics* 10, 207–214. doi:10.1007/s10142-009-0144-9
- Hall, D., Tegstrom, C., and Ingvarsson, P. K. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. *Brief. Funct. Genomics* 9, 157–165. doi:10.1093/bfpg/elp048
- Halling, P., and Simms-Borre, P. (2008). Overview of lignocellulosic feedstock conversion into ethanol-focus on sugarcane bagasse. *Int. Sugar J.* 110, 191.
- Handakumbura, P. P., and Hazen, S. P. (2012). Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with *Arabidopsis thaliana*. *Front. Plant Sci.* 3:74. doi:10.3389/fpls.2012.00074
- Harris, D., and DeBolt, S. (2010). Synthesis, regulation and utilization of lignocellulosic biomass. *Plant Biotechnol. J.* 8, 244–262. doi:10.1111/j.1467-7652.2009.00481.x
- Harrison, M. D., Geijskes, J., Coleman, H. D., Shand, K., Kinkema, M., Palupe, A., et al. (2011). Accumulation of recombinant cellobiohydrolase and endoglucanase in the leaves of mature transgenic sugar cane. *Plant Biotechnol. J.* 9, 884–896. doi:10.1111/j.1467-7652.2011.00597.x
- Harvey, M., Hockett, B., and Botha, F. (1994). “Use of polymerase chain reaction (PCR) and random amplification of polymorphic DNAs (RAPDs) for the determination of genetic distances between 21 sugarcane varieties,” in *Proceedings of the South African Sugar Technologists Association* (Citeseer), 36–40.
- Hattori, T., and Morita, S. (2010). Energy crops for sustainable bioethanol production; which, where and how? *Plant Prod. Sci.* 13, 221–234. doi:10.1626/pps.13.221
- Henry, R. J. (2010a). Evaluation of plant biomass resources available for replacement of fossil oil. *Plant Biotechnol. J.* 8, 288–293. doi:10.1111/j.1467-7652.2009.00482.x
- Henry, R. J. (2010b). *Plant Resources for Food, Fuel and Conservation*. London: Earthscan.

- Henry, R. J., Edwards, M., Waters, D. L., Gopala Krishnan, S., Bundock, P., Sexton, T. R., et al. (2012). Application of large-scale sequencing to marker discovery in plants. *J. Biosci.* 37, 829–841. doi:10.1007/s12038-012-9253-z
- Higuchi, T. (1981). "Biosynthesis of lignin," in *Plant Carbohydrates II*, eds W. Tanner and F. Loewus (Berlin: Springer), 194–224.
- Himmel, M., Ding, S., Johnson, D., Adney, W., Nimlos, M., Brady, J., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi:10.1126/science.1137016
- Hirschhorn, J. N., and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* 6, 95–108. doi:10.1038/nrg1521
- Hoarau, J.-Y., Grivet, L., Offmann, B., Raboin, L.-M., Diorflar, J.-P., Payet, J., et al. (2002). Genetic dissection of a modern sugarcane cultivar (*Saccharum spp.*). II. Detection of QTLs for yield components. *Theor. Appl. Genet.* 105, 1027–1037. doi:10.1007/s00122-002-1047-5
- Hotta, C., Lembke, C., Domingues, D., Ochoa, E., Cruz, G. Q., Melotto-Passarin, D., et al. (2010). The biotechnology roadmap for sugarcane improvement. *Trop. Plant Biol.* 3, 75–87. doi:10.1007/s12042-010-9050-5
- Huang, E., Aitken, K., and George, A. (2010). "Association Studies," in *Genetics, Genomics and Breeding of Sugarcane*, eds R. Henry and C. Kole (Enfield, NH: Science Publishers), 43–68.
- Ingelbrecht, I. L., Irvine, J. E., and Mirkov, T. E. (1999). Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. *Plant Physiol.* 119, 1187–1198. doi:10.1104/pp.119.4.1187
- Ingvarsson, P. K., and Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytol.* 189, 909–922. doi:10.1111/j.1469-8137.2010.03593.x
- Jakob, K., Zhou, F., and Paterson, A. (2009). Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. *In vitro Cell. Dev. Biol. Plant* 45, 291–305. doi:10.1007/s11627-009-9214-x
- Jannoo, N., Grivet, L., Chantret, N., Garsmeur, O., Glaszmann, J. C., Arruda, P., et al. (2007). Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant J.* 50, 574–585. doi:10.1111/j.1365-3113.2007.03082.x
- Johansson, T. B., and Burnham, L. (1993). *Renewable Energy: Sources for Fuels and Electricity*. Washington, DC: Island Press.
- Joyce, P., McQualter, R., Handley, J., Dale, J., Harding, R., and Smith, G. (1998). "Transgenic sugarcane resistant to sugarcane mosaic virus," in *Proceedings-Australian Society of Sugar Cane Technologists* (Salisbury: Watson Ferguson and Company), 204–210.
- Jung, H. G. (1989). Forage lignins and their effects on fiber digestibility. *Agronomy Journal* 81, 33–38. doi:10.2134/agronj1989.00021962008100010006x
- Jung, J. H., Fouad, W. M., Vermerris, W., Gallo, M., and Altpeter, F. (2012). RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass. *Plant Biotechnol. J.* 10, 1067–1076. doi:10.1111/j.1467-7652.2012.00734.x
- Kim, C., Lee, T. H., Compton, R. O., Robertson, J. S., Pierce, G. J., and Paterson, A. H. (2013). A genome-wide BAC end-sequence survey of sugarcane elucidates genome composition, and identifies BACs covering much of the euchromatin. *Plant Mol. Biol.* 81, 139–147. doi:10.1007/s11103-012-9987-x
- Lam, E., Shine, J., Da Silva, J., Lawton, M., Bonos, S., Calvino, M., et al. (2009). Improving sugarcane for biofuel: engineering for an even better feedstock. *Glob. Change Biol. Bioenergy* 1, 251–255. doi:10.1111/j.1757-1707.2009.01016.x
- Le Cunff, L., Garsmeur, O., Raboin, L. M., Pauquet, J., Telismart, H., Selvi, A., et al. (2008). Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (*Bru1*) in highly polyploid sugarcane (2n ~ 12x ~ 115). *Genetics* 180, 649–660. doi:10.1534/genetics.108.091355
- Li, C., Sun, L., Simmons, B., and Singh, S. (2013). Comparing the recalcitrance of *Eucalyptus*, *Pine*, and *Switchgrass* using ionic liquid and dilute acid pretreatments. *Bioenerg. Res.* 6, 14–23. doi:10.1007/s12155-012-9220-4
- Li, X., Ximenes, E., Kim, Y., Slininger, M., Meilan, R., Ladisch, M., et al. (2010). Lignin monomer composition affects *Arabidopsis* cell-wall degradability after liquid hot water pretreatment. *Biotechnol. Biofuels* 3, 27. doi:10.1186/1754-6834-3-27
- Lima, M. L., Garcia, A. A., Oliveira, K. M., Matsuoaka, S., Arizono, H., De Souza, C. L. Jr., et al. (2002). Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum spp.*). *Theor. Appl. Genet.* 104, 30–38. doi:10.1007/s001220200003
- Loureiro, M., Barbosa, M. P., Lopes, F. F., and Silvério, F. (2011). "Sugarcane breeding and selection for more efficient biomass conversion in cellulosic ethanol," in *Routes to Cellulosic Ethanol*, eds M. S. Buckeridge and G. H. Goldman (New York, NY: Springer), 199–239.
- Lupoi, J., Singh, S., Simmons, B., and Henry, R. (2013). Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high-throughput techniques. *Bioenerg. Res.* 7, 1–23. doi:10.1007/s12155-013-9352-1
- Lupoi, J. S., Singh, S., Parthasarathi, R., Simmons, B. A., and Henry, R. J. (2015). Recent innovations in analytical methods for the qualitative and quantitative assessment of lignin. *Renew. Sustain. Energ. Rev.* 49, 871–906. doi:10.1016/j.rser.2015.04.091
- Ma, H., Albert, H. H., Paull, R., and Moore, P. H. (2000). Metabolic engineering of invertase activities in different subcellular compartments affects sucrose accumulation in sugarcane cells. *Funct. Plant Biol.* 27, 1021–1030. doi:10.1071/PP00029
- Ma, H. M., Schulze, S., Lee, S., Yang, M., Mirkov, E., Irvine, J., et al. (2004). An EST survey of the sugarcane transcriptome. *Theor. Appl. Genet.* 108, 851–863. doi:10.1007/s00122-003-1510-y
- Macrelli, S., Galbe, M., and Wallberg, O. (2014). Effects of production and market factors on ethanol profitability for an integrated first and second generation ethanol plant using the whole sugarcane as feedstock. *Biotechnol. Biofuels* 7, 26. doi:10.1186/1754-6834-7-26
- Macrelli, S., Mogensen, J., and Zacchi, G. (2012). Techno-economic evaluation of 2nd generation bioethanol production from sugar cane bagasse and leaves integrated with the sugar-based ethanol process. *Biotechnol. Biofuels* 5, 22. doi:10.1186/1754-6834-5-22
- Manickavasagam, M., Ganapathi, A., Anbazhagan, V. R., Sudhakar, B., Selvaraj, N., Vasudevan, A., et al. (2004). *Agrobacterium*-mediated genetic transformation and development of herbicide-resistant sugarcane (*Saccharum species* hybrids) using axillary buds. *Plant Cell Rep.* 23, 134–143. doi:10.1007/s00299-004-0794-y
- Manners, J., McIntyre, L., Casu, R., Cordeiro, G., Jackson, M., Aitken, K., et al. (2004). "Can genomics revolutionize genetics and breeding in sugarcane," in *Proceedings of the 4th International Crop Science Congress* (Brisbane, QLD).
- Matsuoka, S., Ferro, J., and Arruda, P. (2009). The Brazilian experience of sugarcane ethanol industry. *In vitro Cell. Dev. Biol. Plant* 45, 372–381. doi:10.1007/s11627-009-9220-z
- McIntyre, C. L., Goode, M. L., Cordeiro, G., Bundock, P., Elliott, F., Henry, R. J., et al. (2015). Characterisation of alleles of the sucrose phosphate synthase gene family in sugarcane and their association with sugar-related traits. *Mol. Breed.* 35, 1–14. doi:10.1007/s11032-015-0286-5
- McIntyre, C. L., Jackson, M., Cordeiro, G. M., Amouyal, O., Hermann, S., Aitken, K. S., et al. (2006). The identification and characterisation of alleles of sucrose phosphate synthase gene family III in sugarcane. *Mol. Breed.* 18, 39–50. doi:10.1007/s11032-006-9012-7
- McIntyre, C. L., Whan, V. A., Croft, B., Magarey, R. and Smith, G. R. (2005). Identification and validation of molecular markers associated with pachymetra root rot and brown rust resistance in sugarcane using Map- and association-based approaches. *Molecular Breeding* 16, 151–161. doi:10.1007/s11032-005-7492-5
- Mertes, F., Elsharawy, A., Sauer, S., Van Helvoort, J. M. L. M., Van Der Zaag, P. J., Franke, A., et al. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief. Funct. Genomics* 10, 374–386. doi:10.1093/bfgp/blr033
- Ming, R., Liu, S.-C., Lin, Y.-R., Da Silva, J., Wilson, W., Braga, D., et al. (1998). Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* 150, 1663–1682.
- Ming, R., Liu, S. C., Moore, P. H., Irvine, J. E., and Paterson, A. H. (2001). QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. *Genome Res.* 11, 2075–2084. doi:10.1101/gr.198801
- Ming, R., Moore, P. H., Wu, K.-K., D'Hont, A., Glaszmann, J. C., Tew, T. L., et al. (2010). "Sugarcane improvement through breeding and biotechnology," in *Plant Breeding Reviews*, ed. J. Janick (Oxford, UK: John Wiley & Sons, Inc), 15–118. doi:10.1002/9780470650349.ch2
- Moore, P. H. (2009). "Sugarcane biology, yield, and potential for improvement," in *Workshop BIOEN on Sugarcane Improvement* (San Pablo).

- Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y., Holtzapfle, M., et al. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* 96, 673–686. doi:10.1016/j.biortech.2004.06.025
- Mutwil, M., Debolt, S., and Persson, S. (2008). Cellulose synthesis: a complex complex. *Curr. Opin. Plant Biol.* 11, 252–257. doi:10.1016/j.pbi.2008.03.007
- Nagaraj, S. H., Gasser, R. B., and Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief. Bioinformatics* 8, 6–21. doi:10.1093/bib/bbl015
- Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi:10.1016/j.tplants.2004.05.006
- Pandey, A., Soccol, C. R., Nigam, P., and Soccol, V. T. (2000). Biotechnological potential of agro-industrial residues. I: sugarcane bagasse. *Bioresour. Technol.* 74, 69–80. doi:10.1016/S0960-8524(99)00142-X
- Panje, R. R. and Babu, C. N. (1960). Studies in *Saccharum spontaneum* distribution and geographical association of chromosome numbers. *Cytologia* 25, 152–72. doi:10.1508/cytologia.25.152
- Park, J. W., Benatti, T. R., Marconi, T., Yu, Q., Solis-Gracia, N., Mora, V., et al. (2015). Cold responsive gene expression profiling of sugarcane and *Saccharum spontaneum* with functional analysis of a cold inducible *Saccharum* homolog of NOD26-like intrinsic protein to salt and water stress. *PLoS ONE* 10:e0125810. doi:10.1371/journal.pone.0125810
- Paterson, A., Bowers, J., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. doi:10.1038/nature07723
- Paterson, A., Souza, G., Sluys, M. V., Ming, R., D'Hont, A., Henry, R., et al. (2010). "Structural genomics and genome sequencing," in *Genetics, Genomics and Breeding of Sugarcane*, eds. R. Henry and C. Kole (Enfield, NH: Science Publishers), 149–165.
- Pauly, M., Gille, S., Liu, L., Mansoori, N., De Souza, A., Schultink, A., et al. (2013). Hemicellulose biosynthesis. *Planta* 238, 627–642. doi:10.1007/s00425-013-1921-1
- Pereira, S. C., Maehara, L., Machado, C. M. M., and Farinas, C. S. (2015). 2G ethanol from the whole sugarcane lignocellulosic biomass. *Biotechnol. Biofuels* 8, 44. doi:10.1186/s13068-015-0224-0
- Pinto, L. R., Oliveira, K. M., Ulian, E. C., Garcia, A. A., and De Souza, A. P. (2004). Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome* 47, 795–804. doi:10.1139/g04-055
- Piperidis, G., D'Hont, A., and Hogarth, D. (2001). "Chromosome composition analysis of various *Saccharum* interspecific hybrids by genomic in situ hybridisation (GISH)," in *International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress*, Vol. 2 ed. D. M. Hogarth (Brisbane: Australian Society of Sugar Cane Technologists), 565–566.
- Piperidis, G., Piperidis, N., and D'Hont, A. (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. Genomics* 284, 65–73. doi:10.1007/s00438-010-0546-3
- Rabemanolontsoa, H., and Saka, S. (2013). Comparative study on chemical composition of various biomass species. *RSC Adv.* 3, 3946–3956. doi:10.1039/c3ra22958k
- Raboin, L. M., Pauquet, J., Butterfield, M., D'Hont, A., and Glaszmann, J. C. (2008). Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theor. Appl. Genet.* 116, 701–714. doi:10.1007/s00122-007-0703-1
- Ramos, R. L. B., Tovar, F. J., Junqueira, R. M., Lino, F. B., and Sachetto-Martins, G. (2001). Sugarcane expressed sequence tags (ESTs) encoding enzymes involved in lignin biosynthesis pathways. *Genet. Mol. Biol.* 24, 235–241. doi:10.1590/S1415-47572001000100031
- Renewable Fuels Association. (2015). *World Fuel Ethanol Production*. Available at: <http://ethanolrfa.org/pages/World-Fuel-Ethanol-Production>. Accessed 25 May 2015
- Roach, B. (1989). "Origin and improvement of the genetic base of sugarcane," in *Proceedings Australian Society of Sugar Cane Technologists*, eds. B. T. Egan (Salisbury: Ferguson and Company), 34–47.
- Saathoff, A., Sarath, G., Chow, E., Dien, B., and Tobias, C. (2011). Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PLoS ONE* 6:e16416. doi:10.1371/journal.pone.0016416
- Santos Brito, M., Nobile, P., Bottcher, A., Dos Santos, A., Creste, S., De Landell, M., et al. (2015). Expression profile of sugarcane transcription factor genes involved in lignin biosynthesis. *Trop. Plant Biol.* 8, 19–30. doi:10.1007/s12042-015-9147-y
- Sato, K., Motoi, Y., Yamaji, N., and Yoshida, H. (2011). 454 sequencing of pooled BAC clones on chromosome 3H of barley. *BMC Genomics* 12:246. doi:10.1186/1471-2164-12-246
- Saxena, I. M., and Brown, R. M. Jr. (2000). Cellulose synthases and related enzymes. *Curr. Opin. Plant Biol.* 3, 523–531. doi:10.1016/S1369-5266(00)00125-4
- Schubert, C. (2006). Can biofuels finally take center stage? *Nat. Biotechnol.* 24, 777–784. doi:10.1038/nbt0706-777
- Seabra, J. E. A., Tao, L., Chum, H. L., and Macedo, I. C. (2010). A techno-economic evaluation of the effects of centralized cellulosic ethanol and co-products refinery options with sugarcane mill clustering. *Biomass Bioenergy* 34, 1065–1078. doi:10.1016/j.biombioe.2010.01.042
- Shi, J., Gladden, J. M., Sathitsuksanoh, N., Kambam, P., Sandoval, L., Mitra, D., et al. (2013). One-pot ionic liquid pretreatment and saccharification of switchgrass. *Green Chem.* 15, 2579–2589. doi:10.1039/C3GC40545A
- Simpson, T. (2009). Biofuels: the past, present, and a new vision for the future. *Bioscience* 59, 926–927. doi:10.1525/bio.2009.59.11.2
- Sims, R. E. H., Hastings, A., Schlamadinger, B., Taylor, G., and Smith, P. (2006). Energy crops: current status and future prospects. *Glob. Chang. Biol.* 12, 2054–2076. doi:10.1111/j.1365-2486.2006.01163.x
- Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science* 329, 790–792. doi:10.1126/science.1189268
- Souza, A., Leite, D. C., Pattathil, S., Hahn, M., and Buckeridge, M. (2013). Composition and structure of sugarcane cell wall polysaccharides: implications for second-generation bioethanol production. *Bioenerg. Res.* 6, 564–579. doi:10.1007/s12155-012-9268-1
- Souza, G. M., Berges, H., Bocs, S., Casu, R., D'Hont, A., Ferreira, J. E., et al. (2011). The sugarcane genome challenge: strategies for sequencing a highly complex genome. *Trop. Plant Biol.* 4, 145–156. doi:10.1007/s12042-011-9079-0
- Sreedhar, A., and Collins, A. K. (2010). "Molecular genetic linkage mapping in *Saccharum*," in *Genetics, Genomics and Breeding of Sugarcane*, eds. R. Henry and C. Kole (Enfield, NH: Science Publishers), 69–96. doi:10.1201/EBK1578086849-6
- Sreenivasan, T. V., Ahloowalia, B. S., and Heinz, D. J. (1987). "Chapter 5 – cytogenetics," in *Developments in Crop Science*, ed. J. H. Don (Amsterdam: Elsevier), 211–253.
- Steuernagel, B., Taudien, S., Gundlach, H., Seidel, M., Ariyadasa, R., Schulte, D., et al. (2009). De novo 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. *BMC Genomics* 10:547. doi:10.1186/1471-2164-10-547
- Sticklen, M. (2006). Plant genetic engineering to improve biomass characteristics for biofuels. *Curr. Opin. Biotechnol.* 17, 315–319. doi:10.1016/j.copbio.2006.05.003
- Sugesi. (2015). *The Sugarcane Genome Sequencing Initiative*. Available at: <http://cnrgv.toulouse.inra.fr/fr/Projets/Canne-a-sucre/The-Sugarcane-Genome-Sequencing-Initiative-SUGESI-Strategies-for-Sequencing-a-Highly-Complex-Genome>. Accessed on 25 May 2015
- Sun, N., Liu, H., Sathitsuksanoh, N., Stavila, V., Sawant, M., Bonito, A., et al. (2013). Production and extraction of sugars from switchgrass hydrolyzed in ionic liquids. *Biotechnol. Biofuels* 6, 39. doi:10.1186/1754-6834-6-39
- Tabasum, S., Khan, F. A., Nawaz, S., Iqbal, M. Z., and Saeed, A. (2010). DNA profiling of sugarcane genotypes using randomly amplified polymorphic DNA. *Genet. Mol. Res.* 9, 471–483. doi:10.4238/vol9-1gmr709
- Tai, P., and Miller, J. (2001). A core collection for *Saccharum spontaneum* L. from the world collection of sugarcane. *Crop Sci.* 41, 879–885. doi:10.2135/cropsci2001.413879x
- Taylor, L. E. II, Dai, Z., Decker, S. R., Brunecky, R., Adney, W. S., Ding, S. Y., et al. (2008). Heterologous expression of glycosyl hydrolases in planta: a new departure for biofuels. *Trends Biotechnol.* 26, 413–424. doi:10.1016/j.tibtech.2008.05.002
- Tew, T., and Cobill, R. (2008). "Genetic improvement of sugarcane (*Saccharum* spp.) as an energy crop," in *Genetic Improvement of Bioenergy Crops*, ed. W. Vermerris (New York, NY: Springer), 273–294.
- Todd, J., Wang, J., Glaz, B., Sood, S., Ayala-Silva, T., Nayak, S., et al. (2014). Phenotypic characterization of the Miami world collection of sugarcane (*Saccharum* spp.) and related grasses for selecting a representative core. *Genet. Resour. Crop Evol.* 61, 1581–1596. doi:10.1007/s10722-014-0132-3
- Tomkins, J. P., Yu, Y., Miller-Smith, H., Frisch, D. A., Woo, S. S., and Wing, R. A. (1999). A bacterial artificial chromosome library for sugarcane. *Theor. Appl. Genet.* 99, 419–424. doi:10.1007/s001220051252

- Van Der Weijde, T., Alvim Kamei, C. L., Torres, A. F., Vermerris, W., Dolstra, O., Visser, R. G., et al. (2013). The potential of C4 grasses for cellulosic biofuel production. *Front. Plant Sci.* 4:107. doi:10.3389/fpls.2013.00107
- Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi:10.1016/j.tig.2014.07.001
- Vargas, L., Santa Brígida, A. B., Mota Filho, J. P., De Carvalho, T. G., Rojas, C. A., Vanechoutte, D., et al. (2014). Drought tolerance conferred to sugarcane by association with *Gluconacetobacter diazotrophicus*: a transcriptomic view of hormone pathways. *PLoS ONE* 9:e114744. doi:10.1371/journal.pone.0114744
- Verheye, W. (2010). "Growth and production of sugarcane," in *Soils, Plant Growth and Production Volume II in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO*, Vol. 2. Paris: Eolss Publishers. Available at: www.eolss.net/sample-chapters/c10/e1-05a-22-00.pdf [Accessed 15 July 15].
- Vettore, A. L., Da Silva, F. R., Kemper, E. L., Souza, G. M., Da Silva, A. M., Ferro, M. I., et al. (2003). Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res.* 13, 2725–2735. doi:10.1101/gr.1532103
- Vettore, A. L., Silva, F. R. D., Kemper, E. L., and Arruda, P. (2001). The libraries that made SUCEST. *Genet. Mol. Biol.* 24, 1–7. doi:10.1590/S1415-47572001000100002
- Viikari, L., Vehmaanpera, J., and Koivula, A. (2012). Lignocellulosic ethanol: from science to industry. *Biomass Bioeng.* 46, 13–24. doi:10.1016/j.biombioe.2012.05.008
- Vogel, J. (2008). Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* 11, 301–307. doi:10.1016/j.pbi.2008.03.002
- Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H., and Souza, G. M. (2010). Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* 8, 263–276. doi:10.1111/j.1467-7652.2009.00491.x
- Wang, J., Roe, B., Macmil, S., Yu, Q., Murray, J. E., Tang, H., et al. (2010). Micro-collinearity between autopolyploid sugarcane and diploid *Sorghum* genomes. *BMC Genomics* 11:261. doi:10.1186/1471-2164-11-261
- Wang, X., and Paterson, A. (2013). "Comparative genomic analysis of C4 photosynthesis pathway evolution in grasses," in *Genomics of the Saccharinae*, ed. A. H. Paterson (New York, NY: Springer), 447–477.
- Wei, X., Jackson, P. A., Hermann, S., Kilian, A., Heller-Uszynska, K., and Deomano, E. (2010). Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 53, 973–981. doi:10.1139/g10-050
- Wei, X., Jackson, P. A., McIntyre, C. L., Aitken, K. S., and Croft, B. (2006). Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor. Appl. Genet.* 114, 155–164. doi:10.1007/s00122-006-0418-8
- Whetten, R., and Ron, S. (1995). Lignin biosynthesis. *Plant Cell* 7, 1001–1013. doi:10.2307/3870053
- Whittaker, A., and Botha, F. C. (1997). Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. *Plant Physiol.* 115, 1651–1659.
- Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B. (2012). Whole-genome strategies for marker-assisted plant breeding. *Mol. Breed.* 29, 833–854. doi:10.1007/s11032-012-9699-6
- Yan, L., Xu, C., Kang, Y., Gu, T., Wang, D., Zhao, S., et al. (2013). The heterologous expression in *Arabidopsis thaliana* of *Sorghum* transcription factor SbbHLH1 downregulates lignin synthesis. *J. Exp. Bot.* 64, 3021–3032. doi:10.1093/jxb/ert150
- Yoshida, K., Sakamoto, S., Kawai, T., Kobayashi, Y., Sato, K., Ichinose, Y., et al. (2013). Engineering the *Oryza sativa* cell wall with rice NAC transcription factors regulating secondary wall formation. *Front. Plant Sci.* 4:383. doi:10.3389/fpls.2013.00383
- Yuan, J. S., Tiller, K. H., Al-Ahmad, H., Stewart, N. R., and Stewart, C. N. Jr. (2008). Plants to power: bioenergy to fuel the future. *Trends Plant Sci.* 13, 421–429. doi:10.1016/j.tplants.2008.06.001
- Zhang, J. S., Nagai, C., Yu, Q. Y., Pan, Y. B., Ayala-Silva, T., Schnell, R. J., et al. (2012). Genome size variation in three *Saccharum* species. *Euphytica* 185, 511–519. doi:10.1007/s10681-012-0664-6
- Zhang, L., Xu, J., and Birch, R. G. (1999). Engineered detoxification confers resistance against a pathogenic bacterium. *Nat. Biotechnol.* 17, 1021–1024. doi:10.1038/13721
- Zhang, S.-Z., Yang, B.-P., Feng, C.-L., Chen, R.-K., Luo, J.-P., Cai, W.-W., et al. (2006). Expression of the *Grifola frondosa* trehalose synthase gene and improvement of drought-tolerance in sugarcane (*Saccharum officinarum* L.). *J. Integr. Plant Biol.* 48, 453–459. doi:10.1111/j.1744-7909.2006.00246.x
- Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi:10.3835/plantgenome2008.02.0089

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hoang, Furtado, Botha, Simmons and Henry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification and molecular characterization of the switchgrass AP2/ERF transcription factor superfamily, and overexpression of *PvERF001* for improvement of biomass characteristics for biofuel

OPEN ACCESS

Edited by:

Robert Henry,
The University of Queensland,
Australia

Reviewed by:

Jaime Puna,
Instituto Superior Engenharia Lisboa,
Portugal
Arumugam Muthu,
Council of Scientific and Industrial
Research, India

*Correspondence:

C. Neal Stewart Jr.,
Department of Plant Sciences,
University of Tennessee, 2431 Joe
Johnson Drive, 252 Ellington Plant
Sciences, Knoxville,
TN 37996-4561, USA
nealstewart@utk.edu

Specialty section:

This article was submitted to
Bioenergy and Biofuels, a section of
the journal *Frontiers in Bioengineering
and Biotechnology*

Received: 29 May 2015

Accepted: 29 June 2015

Published: 20 July 2015

Citation:

Wuddineh WA, Mazarei M, Turner GB,
Sykes RW, Decker SR, Davis MF and
Stewart CN Jr. (2015) Identification
and molecular characterization of the
switchgrass AP2/ERF transcription
factor superfamily, and
overexpression of *PvERF001* for
improvement of biomass
characteristics for biofuel.
Front. Bioeng. Biotechnol. 3:101.
doi: 10.3389/fbioe.2015.00101

Wegi A. Wuddineh^{1,2}, Mitra Mazarei^{1,2}, Geoffrey B. Turner^{2,3}, Robert W. Sykes^{2,3},
Stephen R. Decker^{2,3}, Mark F. Davis^{2,3} and C. Neal Stewart Jr.^{1,2*}

¹ Department of Plant Sciences, University of Tennessee, Knoxville, TN, USA, ² Bioenergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA, ³ National Renewable Energy Laboratory, Golden, CO, USA

The APETALA2/ethylene response factor (AP2/ERF) superfamily of transcription factors (TFs) plays essential roles in the regulation of various growth and developmental programs including stress responses. Members of these TFs in other plant species have been implicated to play a role in the regulation of cell wall biosynthesis. Here, we identified a total of 207 AP2/ERF TF genes in the switchgrass genome and grouped into four gene families comprised of 25 AP2-, 121 ERF-, 55 DREB (dehydration responsive element binding)-, and 5 RAV (related to API3/VP) genes, as well as a singleton gene not fitting any of the above families. The ERF and DREB subfamilies comprised seven and four distinct groups, respectively. Analysis of exon/intron structures of switchgrass AP2/ERF genes showed high diversity in the distribution of introns in AP2 genes versus a single or no intron in most genes in the ERF and RAV families. The majority of the subfamilies or groups within it were characterized by the presence of one or more specific conserved protein motifs. *In silico* functional analysis revealed that many genes in these families might be associated with the regulation of responses to environmental stimuli via transcriptional regulation of the response genes. Moreover, these genes had diverse endogenous expression patterns in switchgrass during seed germination, vegetative growth, flower development, and seed formation. Interestingly, several members of the ERF and DREB families were found to be highly expressed in plant tissues where active lignification occurs. These results provide vital resources to select candidate genes to potentially impart tolerance to environmental stress as well as reduced recalcitrance. Overexpression of one of the ERF genes (*PvERF001*) in switchgrass was associated with increased biomass yield and sugar release efficiency in transgenic lines, exemplifying the potential of these TFs in the development of lignocellulosic feedstocks with improved biomass characteristics for biofuels.

Keywords: AP2, ethylene response factors, stress response, transcription factors, biofuel, *PvERF001*, overexpression, sugar release

Introduction

Switchgrass (*Panicum virgatum*) is an outcrossing perennial C4 grass known for its vigorous growth and wide adaptability and, hence, is being developed as a candidate lignocellulosic biofuel feedstock (Yuan et al., 2008). The feasibility of commercial production of liquid transportation biofuel from switchgrass biomass is hampered by biomass recalcitrance (the resistance of cell wall to enzymatic breakdown into simple sugars). Lignin is considered to be a primary contributor to biomass recalcitrance as it hinders the accessibility of cell wall carbohydrates to hydrolytic enzymes. Substantial progress has been made in engineering the switchgrass lignin biosynthesis pathway to reduce lignin content and/or modify its composition (Fu et al., 2011a,b; Shen et al., 2012, 2013a,b; Baxter et al., 2014, 2015). The downregulation of individual genes in the lignin biosynthesis pathway has been effective to reduce lignin, but can result in the production of metabolites that can impede downstream fermentation processes (Tschaplinski et al., 2012). Alternatively, overexpression of transcription factors (TFs), such as switchgrass *MYB4*, has been shown to circumvent this inhibitory effect while leading to significantly reduced biomass recalcitrance and improved ethanol production (Shen et al., 2012, 2013b; Baxter et al., 2015).

The master regulators of gene cluster TFs with altered expression could, in turn, endow such traits as increased biomass yield, tiller number, improved germination/plant establishment, or root growth as well as tolerance to environmental stresses (Xu et al., 2011; Licausi et al., 2013; Ambavaram et al., 2014). Therefore, identification of TFs with such putative roles would provide a dynamic approach to developing better biofuel feedstocks that could thrive under adverse environmental conditions. The availability of switchgrass ESTs (Zhang et al., 2013) and draft genome sequences produced by Joint Genome Institute (JGI), Department of Energy, USA, provides a vital resource for the discovery of relevant target genes that could be utilized in the genetic improvement of perennial grasses, which could be used as dedicated bioenergy feedstocks. However, compared to dicots such as *Arabidopsis*, relatively little is known about the key regulatory mechanisms in monocots that control lignification and cell wall formation; this is especially true of switchgrass. Likewise, we also have depauperate knowledge about stress responses and defense against pests in these species.

APETALA2/ethylene responsive factor (AP2/ERF) is a large group of regulatory protein families in plants that are characterized by the presence of one or two conserved AP2 DNA binding domains. AP2/ERF TFs are involved in the transcriptional regulation of various growth and developmental processes and responses to environmental stressors. The AP2 domain is a stretch of 60–70 conserved amino acid sequences that is essential for the activity of AP2/ERF TFs (Jofuku et al., 2005). It has been demonstrated that the AP2 domain binds the *cis*-acting elements including the GCC box motif (Ohme-Takagi and Shinshi, 1995), the dehydration responsive element (DRE)/C-repeat element (CRT) (Sun et al., 2008), and/or TTG motif (Wang et al., 2015) present in the promoter regions of target genes thereby regulating their expression. The AP2/ERF superfamily can be divided into three major families, namely ERF, AP2, and RAV (related to API3/VP

(Licausi et al., 2013)). The ERF family is further subdivided into two subfamilies, ERF and dehydration responsive element binding proteins (DREB) based on similarities in amino acid residues in the AP2 domain. The DREB subfamily in *Arabidopsis* and rice has been further classified into 4 distinct groups while ERF subfamily was clustered into 8 groups in *Arabidopsis* and 11 groups in rice based on analysis of gene structure and conserved motifs (Nakano et al., 2006). The AP2 family comprises two groups of proteins differing in the number of AP2 domain in their amino acid sequences. The majority of proteins in this group are characterized by the presence of two AP2 domains, but a few members of this group have only a single AP2 domain that is more similar to the AP2 domains in the double domain groups. RAV proteins, on the other hand, are a small family TFs characterized by the presence of B3 DNA binding domain besides a single AP2 domain. Genome-wide analysis of AP2/ERF TFs has been extensively studied in many dicots including *Arabidopsis* (Nakano et al., 2006), *Populus* (Zhuang et al., 2008; Vahala et al., 2013), Chinese cabbage (Liu et al., 2013), grapevine (Licausi et al., 2010), peach (Zhang et al., 2012), and castor bean (Xu et al., 2013). However, with the exception of rice (Nakano et al., 2006; Rashid et al., 2012), and foxtail millet (Lata et al., 2014), little information is available on the AP2/ERF TF families in monocots such as switchgrass.

Numerous genes coding for AP2/ERF superfamily TFs have been identified and functionally characterized in various plant species (Xu et al., 2011; Licausi et al., 2013). The DREB subfamily proteins have been extensively studied with regard to tolerance to abiotic stress such as freezing (Jaglo-Ottosen et al., 1998; Ito et al., 2006; Fang et al., 2015), drought (Hong and Kim, 2005; Oh et al., 2009; Fang et al., 2015), heat (Qin et al., 2007), and salinity (Hong and Kim, 2005; Bouaziz et al., 2013). Moreover, it has been reported that DREB genes play roles in the regulation of ABA-mediated gene expression in response to osmotic stress during germination and early vegetative growth stage (Fujita et al., 2011). ERF TFs, on the other hand, have been shown to participate in the regulation of defense responses against various biotic stresses (Guo et al., 2004; Dong et al., 2015) and/or tolerance to environmental stressors, such as drought (Aharoni et al., 2004; Zhang et al., 2010b), osmotic stress (Zhang et al., 2010a), salinity (Guo et al., 2004), hypoxia (Hattori et al., 2009), and freezing (Zhang and Huang, 2010). Moreover, AP2/ERF TFs in aspen (*PtaERF1*) and *Arabidopsis* (*AtERF004* and *AtERF038*) have been suggested to be associated with the regulation of cell wall biosynthesis in some tissues (Van Raemdonck et al., 2005; Lasserre et al., 2008; Ambavaram et al., 2011). The functions of AP2 family TFs, on the other hand, have been associated with plant organ-specific regulation of growth and developmental programs (Elliott et al., 1996; Jofuku et al., 2005; Horstman et al., 2014). Genes in the RAV TF family have been shown to play a role in the regulation of gene expression in response to phytohormones such as ethylene and brassinosteroid as well as in response to biotic and abiotic stresses (Mittal et al., 2014). Therefore, AP2/ERF TF superfamily may hold tremendous potential for the improvement of bioenergy feedstocks, such as switchgrass, that is intended to be grown on marginal lands that could impose undue environmental stress.

In this study, we report the identification of 207 AP2/ERF TF genes in the switchgrass genome. Cluster analysis of the identified

proteins, distribution of conserved motifs, analysis of their gene structure, and expression profiling were presented. We highlight the potential application of these data to identify putative target genes that might be exploited to improve bioenergy feedstocks. To that end, we cloned one of the ERF subfamily genes, which was subsequently overexpressed in switchgrass to improve biomass productivity and sugar release efficiency.

Materials and Methods

Identification of AP2/ERF Gene Families in Switchgrass Genome

We used representative genes from appropriate rice gene families as the basis to search for orthologs in switchgrass. The amino acid sequences of AP2 domain-containing rice genes represented three families: AP2 (Os02g40070), ERF (Os06g40150), and RAV (Os01g04800). These proteins were used to query the derived amino acid sequences of all switchgrass AP2/ERF TFs using tblastn against the switchgrass EST database (Zhang et al., 2013) or blastp against the *P. virgatum* draft genome (Phytozome v1.1 DOE-JGI)¹. The sequences were retrieved and evaluated for the presence of AP2 domains by searching against the conserved domain database (CDD) at NCBI. The AP2-containing switchgrass sequences were further evaluated for any redundant and missing sequences by blastp searches using the previously identified homologous counterparts of the foxtail millet (Lata et al., 2014) and rice (Nakano et al., 2006; Rashid et al., 2012). The presence of multiple gene copies from the tetraploid switchgrass genome was addressed by the identification of only a single gene copy with the highest similarity to the corresponding homologs in foxtail millet or rice. Genes with additional domains besides the AP2 domain with no corresponding homologs in foxtail millet, rice, and *Arabidopsis* AP2/ERF TFs were excluded from our subsequent analysis.

Cluster and Protein Sequence Analysis of AP2/ERF TFs

The amino acid sequences of the AP2/ERF TFs were imported into the MEGA6 program and multiple sequence alignment analysis was conducted using MUSCLE with default parameters (Edgar, 2004). Construction of cluster trees was performed using the neighbor-joining (NJ) method by the MEGA6 program using a bootstrap value of 1000, Poisson correction and pairwise deletion (Tamura et al., 2013). Conserved motifs in switchgrass AP2/ERF TFs were identified with the online tool, MEME version 4.10.0² using the following parameters: optimum width, 6–200 amino acids; with any number of repetitions and maximum number of motifs set at 25 (Bailey and Elkan, 1994).

Analysis of Gene Structure and Gene Ontology Annotation

The genomic and coding DNA sequences of the identified AP2/ERF TFs were retrieved from the Phytozome (*P. virgatum* v1.1 DOE-JGI). The exon–intron organizations in these genes

were visualized by the gene structure display server³ (Guo et al., 2007). To evaluate the gene ontology (GO) annotation of the identified AP2/ERF TFs, their amino acid sequences were imported into the Blast2GO suite (Conesa and Gotz, 2008). Blastp search was performed against rice protein sequences at NCBI. The resulting hits were mapped to obtain the GO terms, which were annotated to assign functional terms to the query sequences. Plant GOSlim was used to filter the annotation to plant-related terms. The protein subcellular localization prediction tool WOLF PSORT⁴ was used to complement the results of the cellular localization predicted by blast2GO.

Analysis of Transcript Data from the Switchgrass Gene Expression Atlas

The transcript data for the AP2/ERF superfamily TFs were extracted from the publicly available switchgrass gene expression atlas (PviGEA)⁵ (Zhang et al., 2013), which was obtained by Affymetrix microarray analysis. The probe set IDs of 108 matching genes representing the switchgrass unitranscripts (PviUT) were identified by tblastn query search using the amino acid sequences of the AP2/ERF TFs. The transcript data for each tissues and stage of development were retrieved using the probe set IDs. The expression values of the genes were log2 transformed and a heatmap was created using an online graphing tool, Plotly⁶. Tissues used for the extraction of RNA to determine the level of expression included the following: whole seeds for seed germination at 24, 48, 72 and 96 h intervals post-imbibition, whole shoots and roots at vegetative stages, V1–V5, pooled leaf sheath (LSH), leaf blade (LB) and nodes, whole crown, the bottom, middle, and top portions of the fourth internode, vascular bundle tissues, and middle portion of the third internode all at E4 (stem elongation stage 4) developmental stage. For analysis of the expression level during reproductive developmental stages, inflorescence tissues and whole seeds along with floral tissues such as lemma and palea were used.

Vector Construction and Plant Transformation

Cloning and tissue culture was performed as previously described (Wuddineh et al., 2015). Briefly, the putative homolog of *Arabidopsis* AtSHN2 (At5g11190) and rice OsSHN (Os06g40150) was identified by tblastn or blastp against the switchgrass EST database or draft genome (Phytozome v1.1 DOE-JGI) followed by cluster and multiple sequence alignment analysis to discriminate the most closely related gene for cloning. For construction of over-expression cassette, the open reading frame (ORF) of *PvERF001* was isolated from cDNA obtained from ST1 clonal genotype of 'Alamo' switchgrass using gene-specific primers flanking the ORF of the gene and cloned into pANIC-10A expression vector by GATEWAY recombination (Mann et al., 2012). The primer pairs used for cloning are shown in Table S1 in Supplementary Material. Embryogenic callus derived from SA1 clonal genotype of 'Alamo' switchgrass (King et al., 2014) was transformed with the expression vector construct through *Agrobacterium*-mediated

¹<http://phytozome.jgi.doe.gov/pz/portal.html>

²<http://meme-suite.org/>

³<http://gsds.cbi.pku.edu.cn/>

⁴<http://www.genscript.com>

⁵<http://switchgrassgenomics.noble.org/>

⁶<https://plot.ly/plot>

transformation (Burris et al., 2009). Antibiotic selection was carried out for about 2 months on 30–50 mg/L hygromycin followed by regeneration of orange fluorescent protein reporter-positive callus sections on regeneration medium (Li and Qu, 2011) containing 400 mg/L timentin. Regenerated plants were rooted on MSO medium (Murashige and Skoog, 1962) with 250 mg/L cefotaxime to assure elimination of *Agrobacterium* from the tissues as well as promote shoot regeneration from transgenic callus (Grewal et al., 2006), and the transgenic lines were screened based on the presence of the insert and expression of the transgene. Simultaneously a non-transgenic control line was also generated from callus.

Plants and Growth Conditions

T0 transgenic and non-transgenic control plants were grown in growth chambers under standard conditions (16 h-day/8 h-night light at 24°C, 390 $\mu\text{E}\cdot\text{m}^{-2}\text{s}^{-1}$) and watered three times per week, including weekly nutrient supplements with 100 mg/L Peter's 20-20-20 fertilizer. Transgenic and non-transgenic control lines were propagated from a single tiller to produce three clonal replicates for measuring growth parameters (Hardin et al., 2013). The plants were grown in 12-L pots in Fafard 3B soil mix (Conrad Fafard, Inc., Agawam, MA, USA) and grown for 4 months to the R1 stage, in which shoot samples were collected to assay the transgene transcript abundance (Moore et al., 1991; Shen et al., 2009). Each sample was snap frozen in liquid nitrogen and macerated with mortar and pestle. The macerated samples were used for RNA extraction as described below.

RNA Extraction and Quantitative Reverse Transcription Polymerase Chain Reaction

RNA extraction and analysis of transgene transcripts were performed as previously described (Wuddineh et al., 2015). Briefly, total RNA was extracted from shoot tip samples of transgenic and non-transgenic control lines using Tri-Reagent (Molecular Research Center, Cincinnati, OH, USA), and 3 μg of the RNA was treated with DNase-I (Promega, Madison, WI, USA). High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA, USA) was used for the synthesis of first-strand cDNA. Power SYBR Green PCR master mix (Applied Biosystems) was utilized to conduct quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis according to the manufacturer's protocol. All the experiments were conducted in triplicates. The list of all primer pairs used for qRT-PCR is shown in Table S1 in Supplementary Material. Analysis of the relative expression was done as previously described (Wuddineh et al., 2015). There was no amplification products observed with all the primer pairs when using only the RNA samples or the water instead of cDNA.

Determination of Leaf Water Loss

The rate of water loss via leaf epidermal layer was determined as previously described (Zhou et al., 2014). The second fully expanded leaves of both transgenic and non-transgenic plants were excised and soaked in 50 mL distilled water for 2 h in the dark to saturate the leaves. Subsequently, the excess water was removed and initial leaf weight was measured and water loss determined by

weighing the leaves every 30 min for at least 3 h. Subsequently, the detached leaves were dried for 24 h at 80°C to determine the final dry weight. The rate of water loss was calculated as the weight of water lost divided by the initial leaf weight.

Analysis of Lignin Content and Composition

Both qualitative (phloroglucinol-HCl staining) and quantitative [pyrolysis molecular beam mass spectrometry (py-MBMS)] analysis of lignin content was performed as previously described (Wuddineh et al., 2015). Briefly, leaf samples collected at the R1 developmental stage and cleared in a 2:1 solution of ethanol and glacial acetic acid for 5 days were used for staining analysis. The cleared leaf samples were immersed in 1% phloroglucinol (in 2:1 ethanol/HCl) overnight for staining and the pictures were taken at 2 \times magnification. For the quantification of lignin content and S:G lignin monomer ratio by NREL high-throughput py-MBMS method, tillers were collected at R1 developmental stage, air-dried for 3 weeks at room temperature and milled to 1 mm (20 mesh) particle size. Lignin content and composition were determined on extractives- and starch-free samples (Sykes et al., 2009).

Determination of Sugar Release

For analysis of sugar release efficiency, tiller samples at R1 developmental stage were collected and air-dried for 3 weeks at room temperature. The dry samples were pulverized to 1 mm (20 mesh) particle size and sugar release efficiency was determined via NREL high-throughput sugar release assays on extractives- and starch-free samples (Decker et al., 2012). Glucose release and xylose release were measured by colorimetric assays and summed for total sugar release.

Statistical Analysis

To analyze the differences between treatment means, analysis of variance (ANOVA) with least significant difference (LSD) procedure was used while PROC TTEST procedure was used to examine the statistical difference between the expression of target genes in transgenic vs non-transgenic lines using SAS version 9.3 (SAS Institute Inc., Cary, NC, USA). Pearson's correlation coefficient to determine the relationship between relative transcript levels and growth parameters was calculated by SAS.

Results

Identification of AP2/ERF TFs in Switchgrass Genome

A total of 207 unique switchgrass genes containing one or two AP2 DNA binding domain were identified from the currently available switchgrass EST and genome databases. Amino acid sequence similarities within the conserved AP2 domain between these proteins and previously characterized AP2/ERF TFs from rice and *Arabidopsis* along with the presence of conserved B3 domain suggest that these proteins might be categorized as putative AP2/ERF TFs. The characteristic features of these genes are summarized in Table S2 in Supplementary Material. The amino acid sequences of AP2/ERF TFs showed wide variation in size (ranging from 119 to 666 amino acids) and sequence composition. Twenty-two of these TFs contained two AP2 DNA-binding domains and hence

TABLE 1 | Summary of the AP2/ERF superfamily gene members found in various plant species.

Family	Subfamily	Group	<i>Panicum virgatum</i>	<i>Oryza sativa</i> ^a	<i>Arabidopsis thaliana</i> ^a	<i>Populus trichocarpa</i> ^b
AP2			25	29	18	26
ERF	DREB	I	12	9	10	5
		II	11	15	15	20
		III	27	26	23	35
		IV	5	6	9	6
		Total	55	56	57	66
	ERF	V	10	8	5	10
		VI	9	6	8	11
		VI-L	7	3	4	4
		VII	17	15	5	6
		VIII	25	13	15	17
		IX	37	18	17	42
		X	12	13	8	9
		Xb-L	–	–	3	4
		XI	4	7	–	–
		Total	121	76	65	103
RAV			5	5	6	6
Singleton			1	1	1	1
Total			207	174	147	202

The switchgrass (*P. virgatum*) data are from this study. Note that switchgrass is the only polyploidy species listed above.

^aNakano et al. (2006).

^bZhuang et al. (2008).

were classified under AP2 family. Five of the AP2/ERF proteins had a B3 conserved domain at the C-terminus in addition to the common AP2 domain, and these genes were grouped into the RAV family. Three of the remaining 180 proteins, namely PvERF049, PvERF160, and PvERF177 with a single AP2 domain, which is more similar to the amino acid sequences of AP2 domains in the AP2 family TFs, were also grouped under the AP2 family. Moreover, one AP2/ERF protein showed a distinct AP2 domain different from all other switchgrass AP2/ERF proteins but with higher shared sequence similarity with the previously identified genes in rice and *Arabidopsis*. The remaining 176 proteins were grouped into ERF family, which was further subdivided into either one of two subfamilies (ERF and DREB) based on sequence similarity in the AP2 domain. The ERF subfamily members included 121 proteins while DREB had only 55 proteins (Table 1).

The distribution of the identified switchgrass AP2/ERF genes across the nine chromosomes was also evaluated. Thus far, only about half of the switchgrass genomic sequences are mapped into their chromosomal locations based on the draft genome assembly by JGI-DOE available at Phytozome. Accordingly, 166 of the 207 genes could be assigned a chromosomal location. The genes were non-evenly distributed across the nine switchgrass chromosomes wherein the highest number of genes was localized on chromosomes 9, 2, and 1, with the fewest number of genes being assigned to chromosome 8 (Table S3 in Supplementary Material).

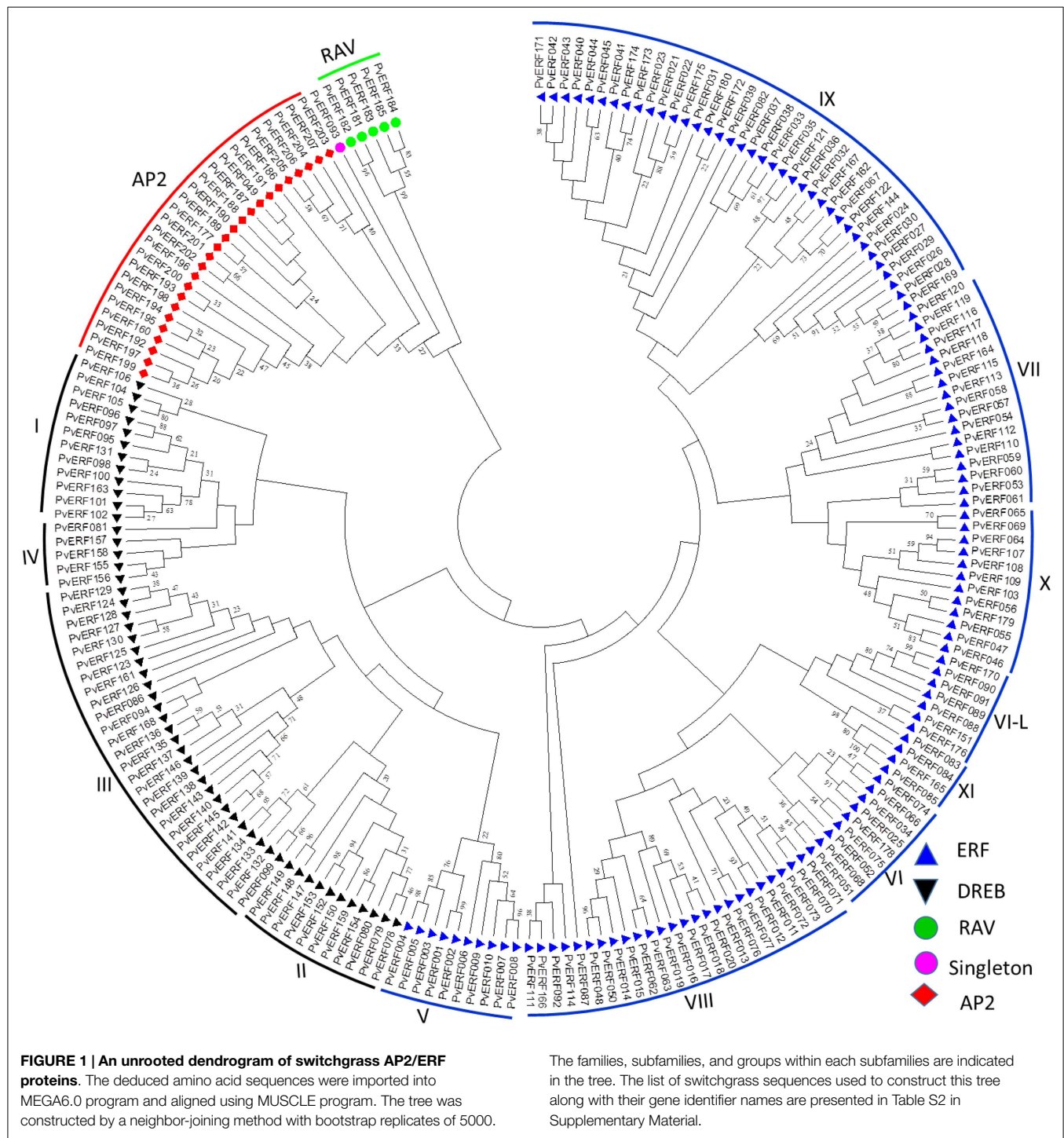
Cluster Analysis of Switchgrass AP2/ERF Proteins

To confirm the classification and evaluate the sequence similarities between the switchgrass AP2/ERF TFs, a dendrogram was constructed by NJ method using the whole amino acid sequences of the proteins. The analysis showed distinct clustering

of the proteins into specific groups and families as previously described in other species (Figure 1). Specifically, these clusters highlighted the distinction between the switchgrass AP2, ERF, and RAV families as well as between the ERF and DREB subfamilies. The ERF and DREB subfamilies were further subdivided into seven (groups V–XI) and four (I–IV) distinct groups, respectively. The cluster analysis also resolved the RAV protein family and the singleton into separate clusters, which was in accordance with the sequence similarities in the conserved domains as well as the presence of additional domains in the families/clusters.

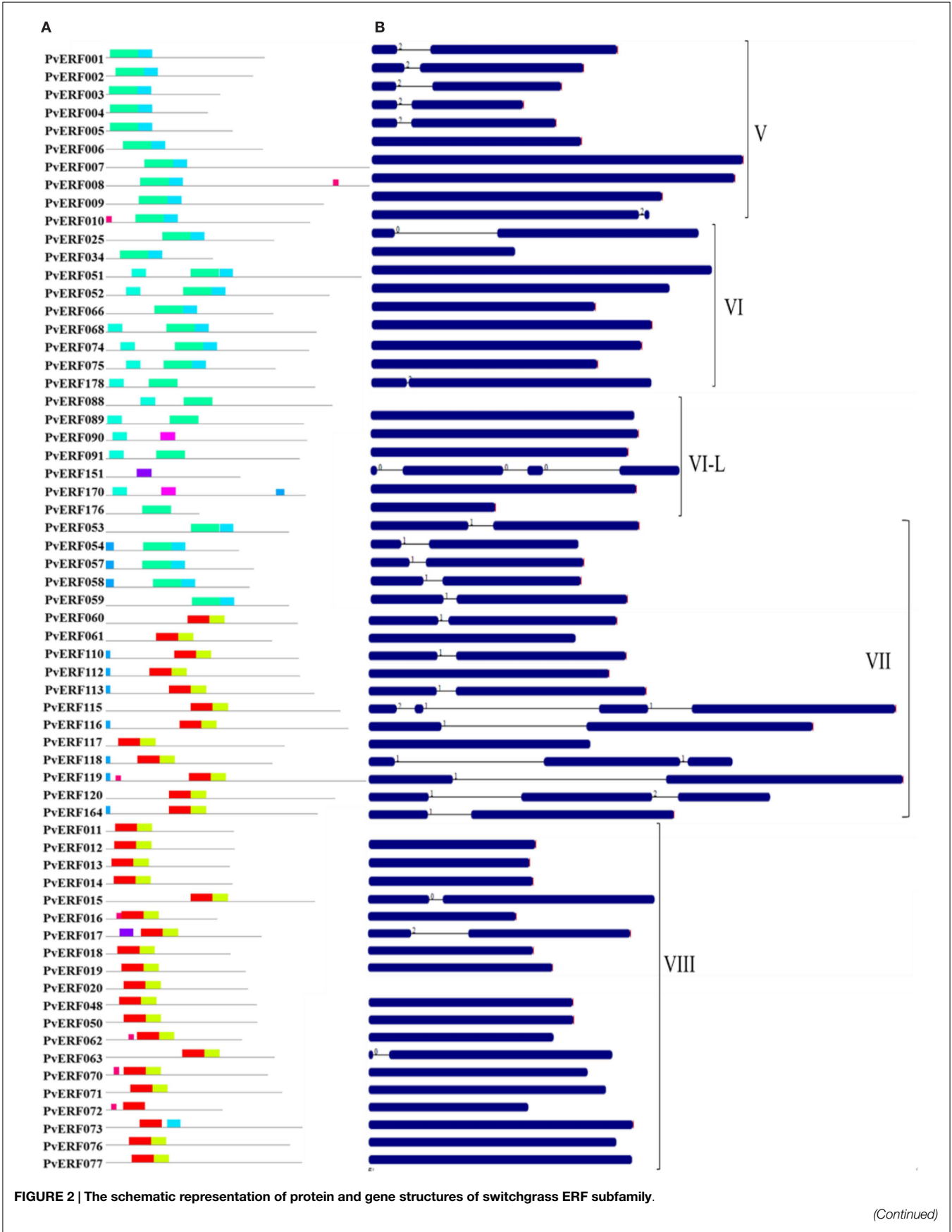
Characterization of AP2/ERF Gene Structures and Conserved Motifs

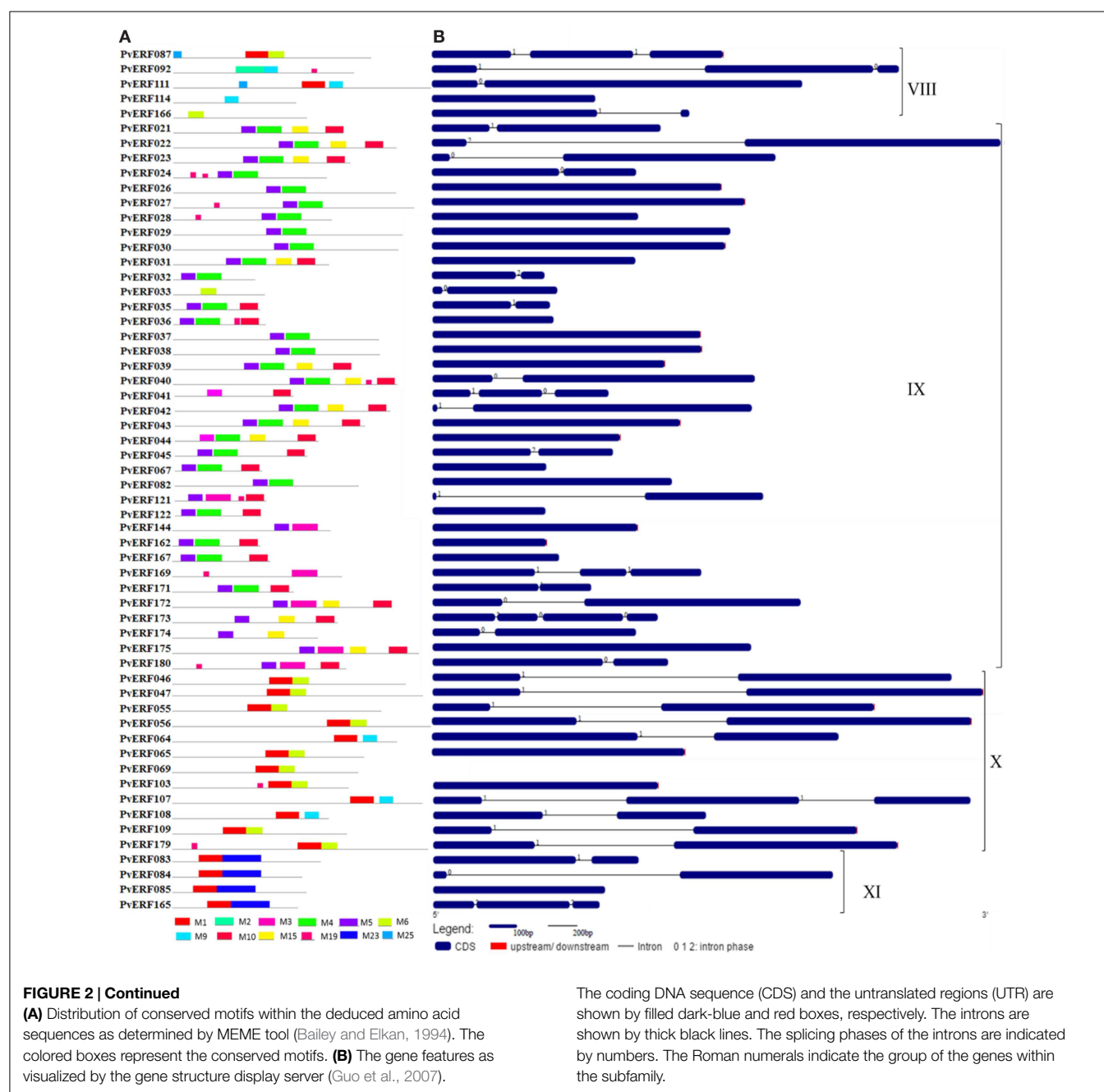
To complement the cluster analysis-based classification, the exon–intron structures of AP2/ERF genes were evaluated. The schematic representations of protein and gene structures of switchgrass AP2/ERF superfamily are presented in Figure 2 (ERF), Figure 3 (DREB), and Figure 4 (AP2, RAV, and Singleton). The ORF lengths of these genes vary from 394 bp for the shortest gene to 5409 bp for the longest gene. Analysis of their gene structure showed highly diverse distribution of intron regions within the ORF of the different gene groups or families. The majority of genes belonging to ERF and DREB subfamilies and all but one of the RAV genes appeared to be intronless. Only nine DREB genes (16%) belonging to group I, III, and VI had a single intron in their gene structures. Among ERF genes, 45 (37%) had a single intron in their ORF while eight genes had two and three of them with three introns in its ORF. On the other hand, genes in the AP2 family contained a higher number of introns; ranging from 1 to 10. Only one gene in the AP2 family had a single intron while majority of the genes had more than five introns. The position and state of the introns in the ORF



of ERF family genes belonging to groups V, VII, and X show high functional conservation. For instance, about half of the genes belonging to phylogenetic group V in the ERF family showed highly conserved intron positions with an intron phase of two, meaning the location of the intron is found between the second and third nucleotides in the codon. Similarly, the intron positions and splicing phases seems conserved in group VII of the ERF subfamily (Figures 2–4).

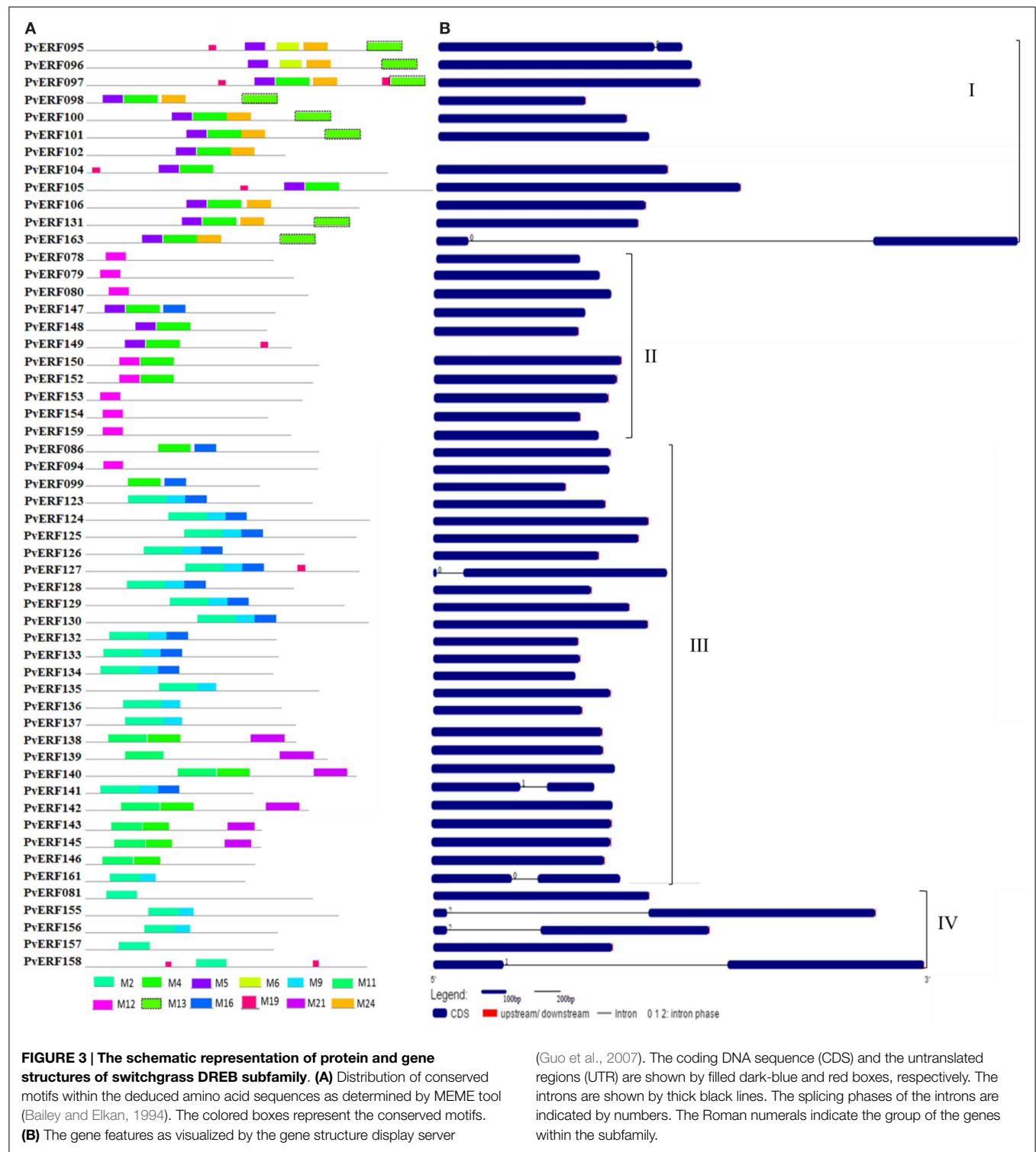
Analysis of amino acid sequence conservation in the whole proteins of AP2/ERF superfamily showed the presence of unique conserved motifs shared between proteins within families, subfamilies, or groups (Figures 2–4). Moreover, shared conserved motifs across families, subfamilies, or between groups within subfamilies were also detected, signifying the conservation of the proteins in the AP2/ERF superfamily. In general, a total of 25 conserved motifs (M1–M25) were identified in the superfamily





of which 14 motifs, M1–M7, M9, M11, M12, M16, M20, M22, and M23, were related to the AP2 domain (Table S4 in Supplementary Material). The conserved motifs from the non-AP2 domain region appear to specify individual groups within the subfamilies. Among the ERF subfamily, proteins in groups VII and IX have the most diverse set of motifs compared to others while proteins in group XI harbors merely two motifs, M1 and M23 with the last motif being unique to the group (Figure 2). Moreover, shared unique motifs were found in the ERF subfamily proteins belonging to group VII (M25), IX (M10 and M15), VI (M18), and VI-L (M18). Most of the DREB genes belonging to group II have only one specific motif (M12) while a few others

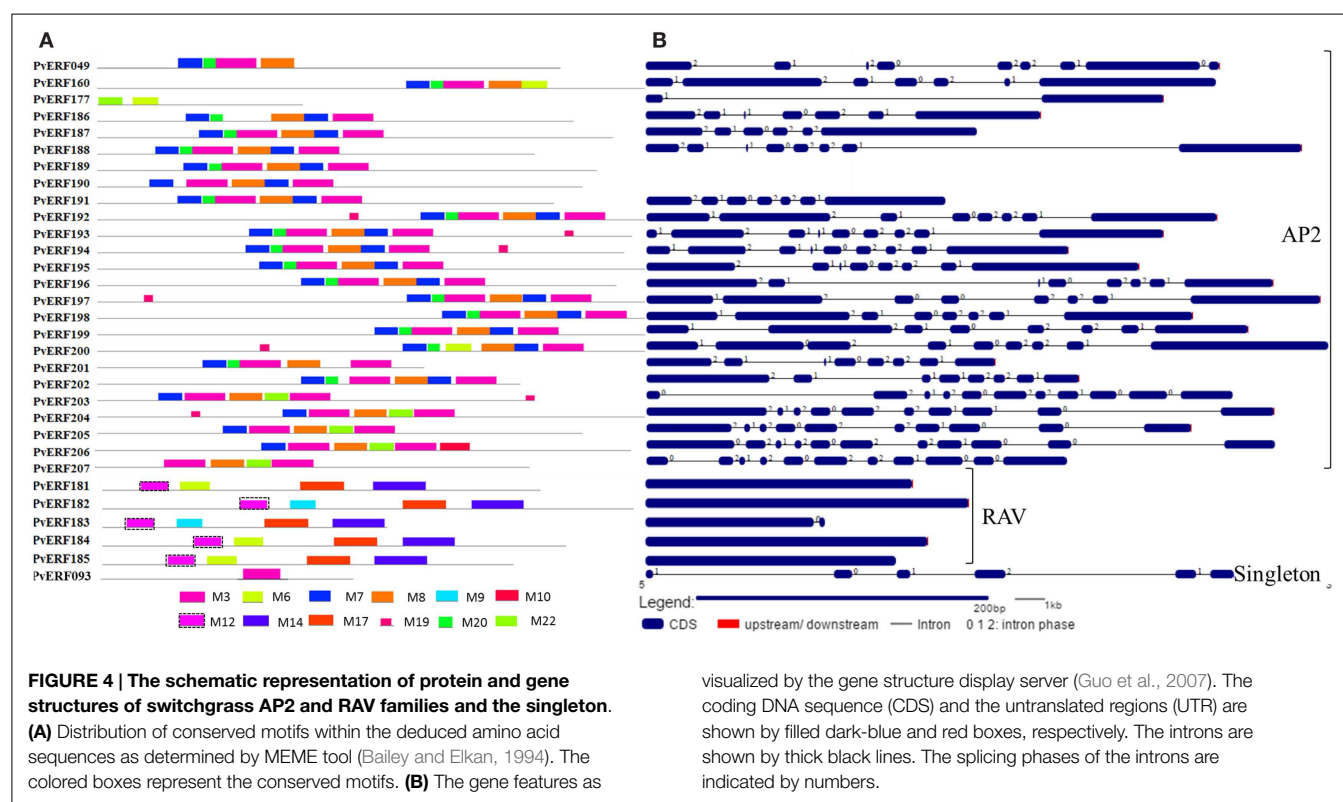
have additional motifs such as M5 (Figure 3). The pattern of conserved motif distribution within the largest group in the DREB subfamily (group III) showed the presence of two unique subgroups sharing a set of three conserved motifs, (M2, M9, and M16) and (M4, M11, and M21), respectively. Three of these motifs (11, 16, and 21) were specific to proteins in group III DREB subfamily. DREB subfamily proteins in group I were distinguished by conserved motif-M13 and motif-M24, while group IV DREB genes have unique motif-M2 (Figure 3). Proteins of AP2 family genes harbor four family-specific motifs, namely M7, M8, M20, and M22 (Figure 4). In addition, the majority of AP2 family proteins share M3 with ERF proteins belonging to group IX.



Similarly, RAV proteins also possess two unique motifs, M14 and M17 spanning the B3 DNA binding domain, in addition to M6 and M12 spanning the AP2 domain (Figure 4). M6 and M12 motifs are also present in most proteins in the ERF and DREB (group II) subfamilies (Figures 2 and 3; Table S4 in Supplementary Material).

Gene Ontology Annotation

Gene ontology analysis of switchgrass AP2/ERF TFs, based on rice reference sequences, predicted candidate genes' molecular functions, putative roles in the regulation of diverse biological processes, and their cellular localization (Figure 5; Table S5 in Supplementary Material). According to blast2GO outputs, over 95% of



the switchgrass genes in the AP2/ERF superfamily were predicted to have sequence-specific DNA binding activities (Figure 5A; Table S5 in Supplementary Material). Furthermore, these genes were anticipated to be involved in the regulation of various biosynthetic processes, which could include the biosynthesis of cuticle, waxes, hormones, and other organic compounds. Importantly, many of these genes were also predicted to participate in the regulation of responses to various environmental stresses caused either by biotic factors such as pathogens and insect pests or abiotic factors such as flooding, water deprivation, wounding, and osmotic stress (Figure 5B; Table S5 in Supplementary Material). Cellular localization of the AP2/ERF TFs was predicted by Blast2GO analysis complemented with subcellular localization prediction tool, WoLF PSORT for proteins with heretofore ambiguous results. The results showed that majority of switchgrass AP2/ERF proteins (>80%) were at least dual targeted, i.e., localized to nucleus, plastid, and/or mitochondrion (Figure 5C; Table S5 in Supplementary Material). Only 39 gene products (20%) were predicted to be localized solely to the nucleus (Table S5 in Supplementary Material).

Expression Pattern of Switchgrass AP2/ERF Genes

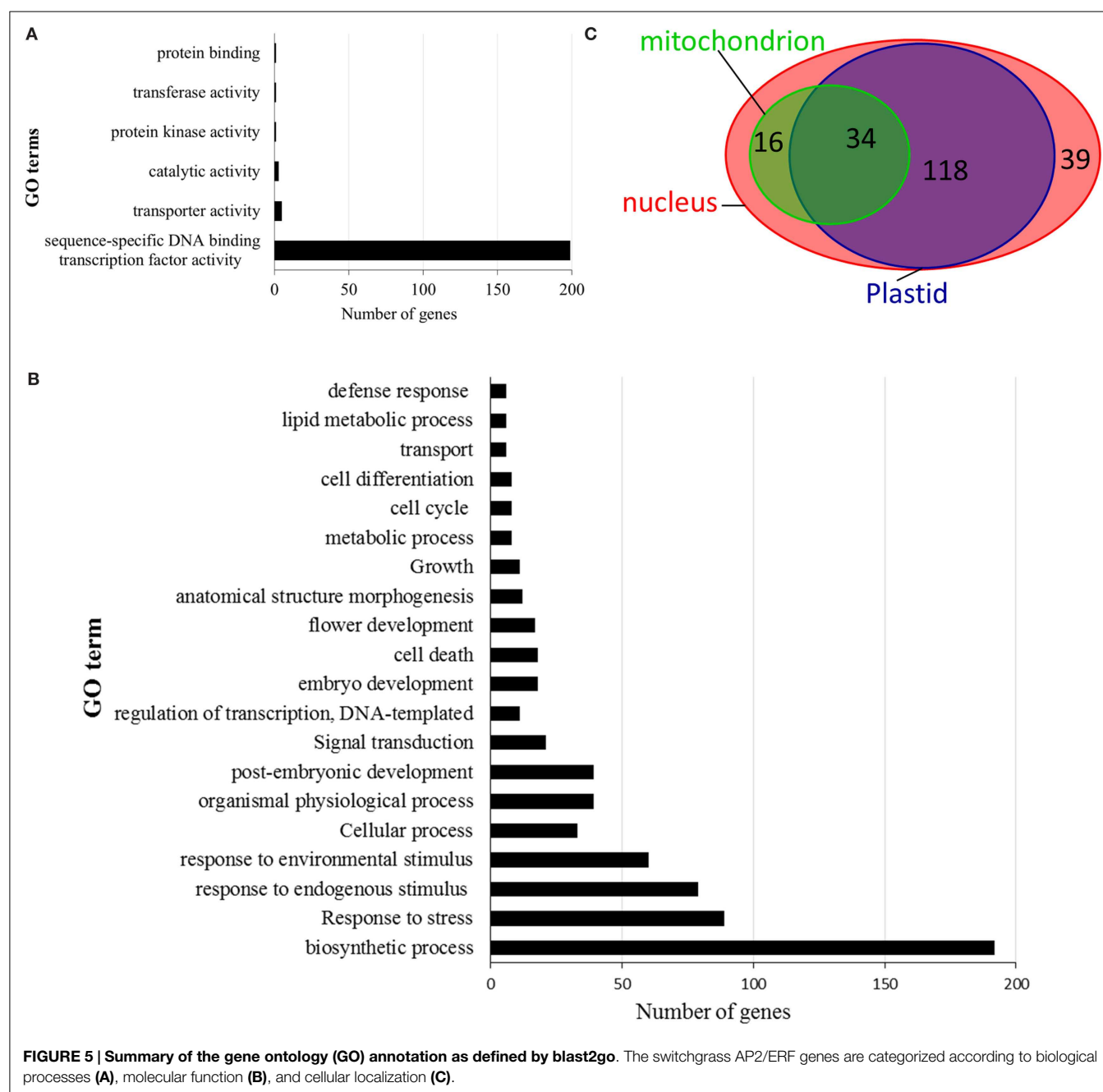
A switchgrass gene expression atlas (PviGEA) containing expression data for about 78,000 unique transcripts in various tissues was recently developed (Zhang et al., 2013) and is publicly available at web server⁷. To investigate whether the identified switchgrass AP2/ERF genes may have any association with various biological

processes that occur during seed germination, vegetative, and reproductive development as well as lignification or cell wall development, transcript data were pooled from the PviGEA web server to assess their expression profile.

During seed germination (Figure 6; Table S6 in Supplementary Material), some genes in the DREB subfamily showed high expression at early stages of germination (radicle emergence) (48 h after imbibition) while others showed increased expression at later stages of germination (mainly coleoptile emergence) (Figure 6; Table S6 in Supplementary Material). Similarly, the expression of many ERF genes showed dramatic increase during early germination stage while numerous others had peak expression at later stages (coleoptile emergence (72 h) and mesocotyl elongation (96 h) stages. Four of the AP2 family genes (*PvERF193*, *PvERF194*, *PvERF195*, and *PvERF201*) displayed increased expression level at radicle emergence whereas the other two (*PvERF049* and *PvERF203*) showed increased expression at coleoptile emergence. The expression of the RAV genes and the singleton gene were apparently relatively less variable throughout the seed germination process (Figure 6; Table S6 in Supplementary Material).

Comparison of the expression pattern of AP2/ERF genes in roots and shoots at three vegetative phases of development (first, third, or fifth fully collared leaf stages) revealed apparent differential expression pattern between the organs and different stages of vegetative development (Figure S1 and Table S6 in Supplementary Material). Moreover, the expression pattern of AP2/ERF genes during reproductive development also showed differential expression between the reproductive tissues from the initiation of inflorescence meristem to the maturation of the seeds (Figure S2 and Table S6 in Supplementary Material).

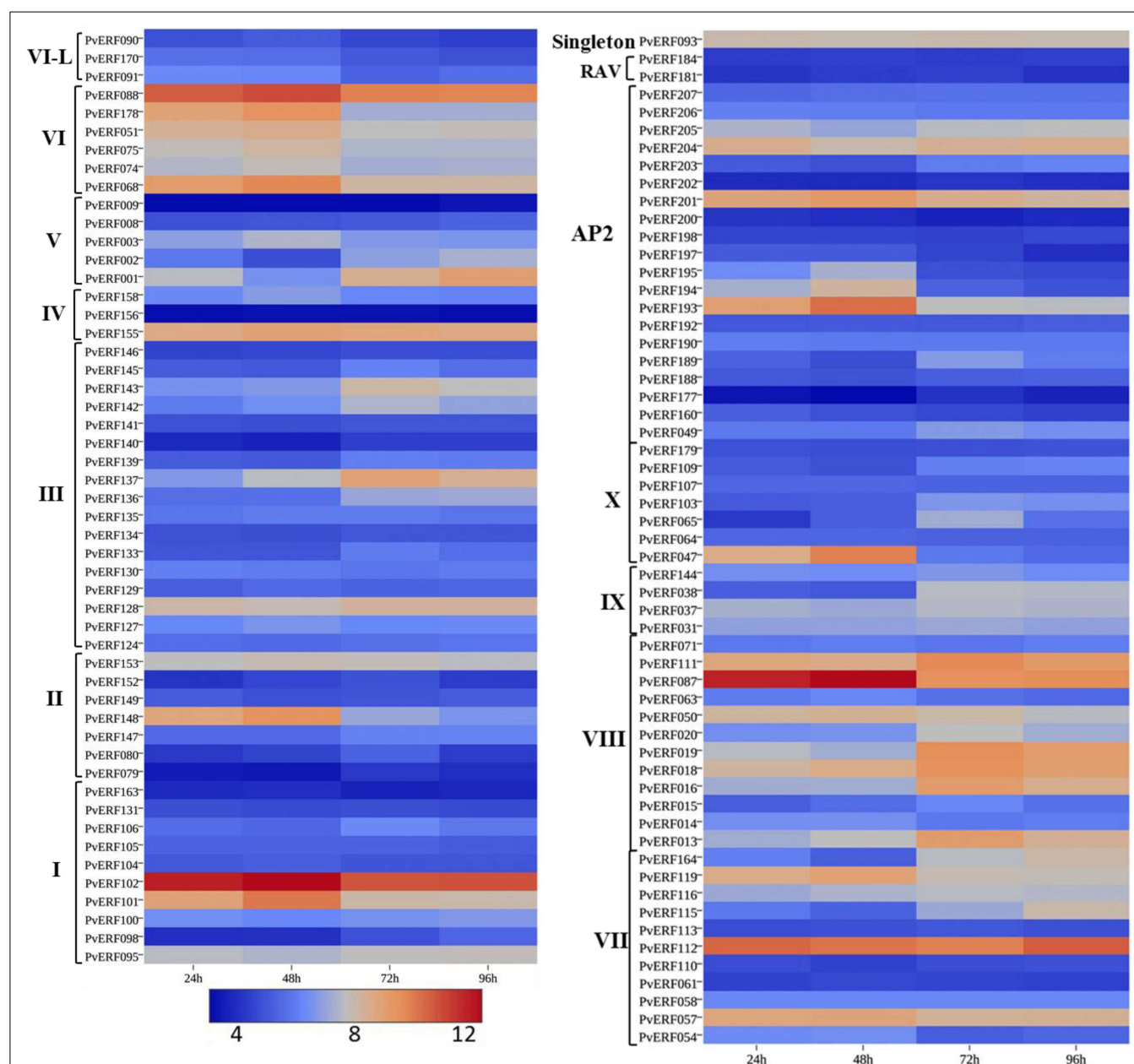
⁷<http://switchgrassgenomics.noble.org/>



Expression Profiles of Switchgrass AP2/ERF Genes in Lignified Tissues

To evaluate whether the identified switchgrass genes coding for AP2/ERF TFs are associated with the regulation of the cell wall biosynthetic genes during cell wall formation or lignification, the transcripts of the genes extracted from the PviGEA web server were used to compare the level of expression in the lignified tissues of vascular bundles and internode fragments against the expression level in less lignified plant tissues such as LBs and sheath. Four genes in group I (*PvERF95*, *PvERF98*, *PvERF101*, and *PvERF102*) and one gene in group II (*PvERF148*) of the DREB subfamily showed highest expression in vascular bundles

and internode tissues followed by internode portions where active lignification is expected (Figure 7; Table S6 in Supplementary Material). The majority of DREB genes belonging to group III were highly expressed mainly in the vascular bundles. Similarly, many genes in the ERF subfamily group VIII (*PvERF013*, *PvERF015*, *PvERF016*, *PvERF018*, *PvERF019*, and *PvERF020*) and X (*PvERF047*, *PvERF065*, and *PvERF103*) showed the highest expression in the vascular bundles followed by youngest internode sections (Figure 7; Table S6 in Supplementary Material). In comparison, only two genes in group IX (*PvERF037* and *PvERF038*), one gene in group VI-L (*PvERF088*), and three genes in group VII (*PvERF111*, *PvERF112*, and *PvERF116*) had high



represents the log₂ values of gene expression with blue color denoting low expression and red for high expression. The Roman numerals I–IV represent the groups of the genes in DREB subfamily while V–X showing the groups of genes in ERF the subfamily.

expression in vascular bundles. Contrastingly, some genes in the ERF subfamily belonging to group V (*PvERF001* and *PvERF002*) and VI (*PvERF068*) showed the highest expression in the basal fragments of the fourth internodes (E4) that is under less active lignification. Other genes including *PvERF178* (VI); *PvERF110* (VII), *PvERF115* (VII), and *PvERF164* (VII); and *PvERF038* (IX) had notably high relative expression in roots than in other tissues. Compared to the ERF family genes, the expression of AP2 genes was highly diverse with some genes having high specificity to roots and vascular bundles. The expression of the two RAV genes

analyzed was uniformly low throughout whereas the singleton gene was highly expressed in the LBs, LSH as well as the vascular bundles, and young internode sections (Figure 7; Table S6 in Supplementary Material).

Overexpression of *PvERF001* in Switchgrass Have Enhanced Plant Growth and Sugar Release Efficiency

Transgenic switchgrass is desired for less recalcitrance biomass for biofuels. To that end, we selected *PvERF001*, a putative

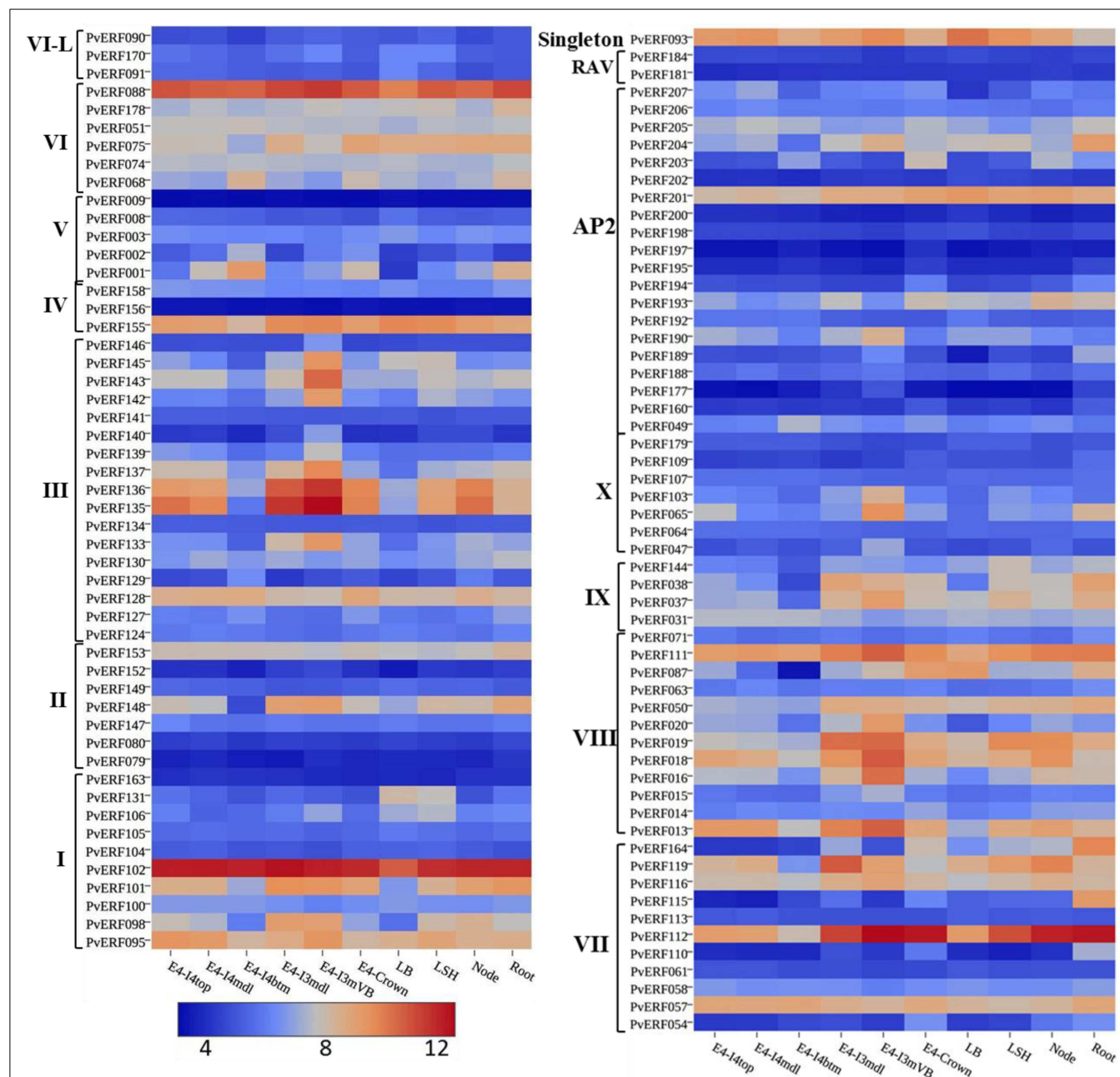
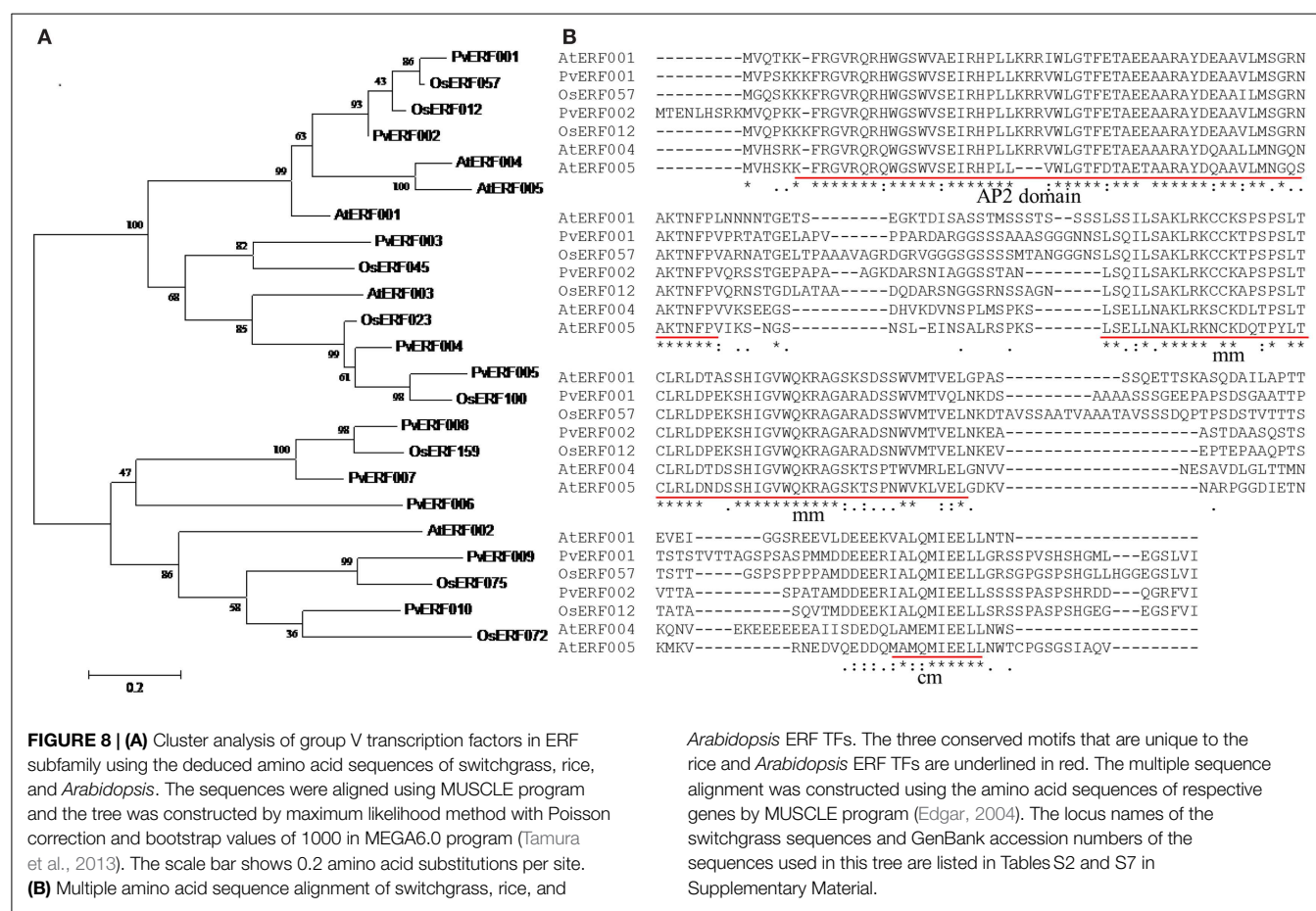


FIGURE 7 | The expression pattern of putative switchgrass AP2/ERF genes in roots and shoot parts including portions of developing internodes and vascular bundles at stem elongation stage 4 (E4). The heat-map depicting the log2-transformed values of the expression level of each gene was obtained from the switchgrass gene expression atlas (PviGEA). The color scale represents the log2 values of gene expression with blue color denoting low expression and red for high expression. The level of expression

was reported for roots, nodes, leaf sheath (LSH), leaf blade (LB), whole crown (E4-crown), vascular bundle isolated from fragments of the third internode (E4-I3mVB), middle fragments of the third internode (E4-I3mdl) and from the bottom (E4-I4btm), middle (E4-I4mdl), and top (E4-I4top) fragments of the fourth internode. The Roman numerals I–IV represent the groups of the genes in DREB subfamily while V–X showing the groups of genes in ERF the subfamily.

switchgrass homolog of *Arabidopsis* AtERF004 (AtSHN2) and rice OsERF057 (OsSHN) in ERF subfamily group V, for overexpression analysis in switchgrass. This gene was selected since the expression of its *Arabidopsis* homolog in transgenic rice resulted in modified cell wall composition (Ambavaram et al., 2011). Sequence grouping/cluster and sequence alignment analysis

suggested that PvERF001 is closely related with its rice and *Arabidopsis* homologs, sharing two highly conserved motifs: the middle motif (mm) and the C-terminal motif (cm) specific to the *Arabidopsis* SHINE clade of TFs (AtERF001, AtERF004, and AtERF005) and OsERF012 and OsERF057 (Figures 8A,B). Thus, the ORF of *PvERF001* was cloned and overexpressed in



switchgrass producing more than six independent transgenic lines, which were confirmed based on genomic PCR for the insertion of the transgene and the hygromycin-resistance gene, as well as visualization of OFP in transgenic plants compared to the non-transgenic control lines (Figure 9A; Figures S3A–C in Supplementary Material). Analysis of the transgene expression level by qRT-PCR showed 1–12-fold overexpression in transgenic lines (Figure 9B). The expression of the endogenous gene in transgenic lines was not affected compared to the non-transgenic control line (Figure 9C). All transgenic lines had equivalent or improved vegetative growth metrics relative to the non-transgenic control lines under greenhouse conditions, which was congruent with the relative transcript levels of the transgene [Pearson's correlation for biomass weights ($R = 0.77$ at $P < 0.05$) and tiller height ($R = 0.73$ at $P = 0.06$)] (Figure 9B; Table 2; Figure S4 in Supplementary Material). Three transgenic lines (3, 7, and 9) had increased biomass. Line 3 had statistically significant increases in four of the six growth traits and approximately twice the dry biomass of the control line (Table 2).

To investigate whether *PvERF001* overexpression could affect the leaf cuticular permeability, the water retention capacity in transgenic and non-transgenic control lines was analyzed in detached leaves measured in the dark to minimize transpirational water loss through stomata. Transgenic lines showed relative reduction in rate of water loss compared with the control lines (Figure S5 in Supplementary Material). However, no tangible

difference was observed in the rate of leaf chlorophyll leaching between transgenic and the control lines (data not shown). Subsequently, we analyzed whether the changes in leaf morphology might be accompanied by changes in the expression level of genes in the cutin and wax biosynthesis pathway, in which none were observed (Figure S6 in Supplementary Material). Moreover, overexpression of *PvERF001* in transgenic switchgrass showed relatively reduced expression of some lignin (*PvC4H* and *PvPAL*), hemicellulose (*PvCSLS2*), and cellulose (*PvCESA4*) biosynthetic genes, as well as some of the transcriptional regulators (*PvMYB48/59* and *PvNST1*) of cell wall biosynthesis (Figures S7A–C in Supplementary Material). The total lignin content in R1 tillers determined by Py-MBMS of cell wall residues and in leaves determined by phloroglucinol–HCl staining did not show sizeable difference between the transgenic and non-transgenic control lines (Figures S8 and S9A in Supplementary Material). Similarly, analysis of the S/G lignin monomer ratio in transgenic lines did not significantly change as compared to that of the non-transgenic control line (Figure S9B in Supplementary Material). However, significant improvement in glucose release efficiency was observed in lines 7 (10%) and 8 (16%) relative to the non-transgenic control line (Table 3). In contrast, none of the transgenic lines released significantly more xylose than the control. The total sugar release, however, was significantly increased in transgenic line 8 by 11% relative to the non-transgenic control (Table 3).

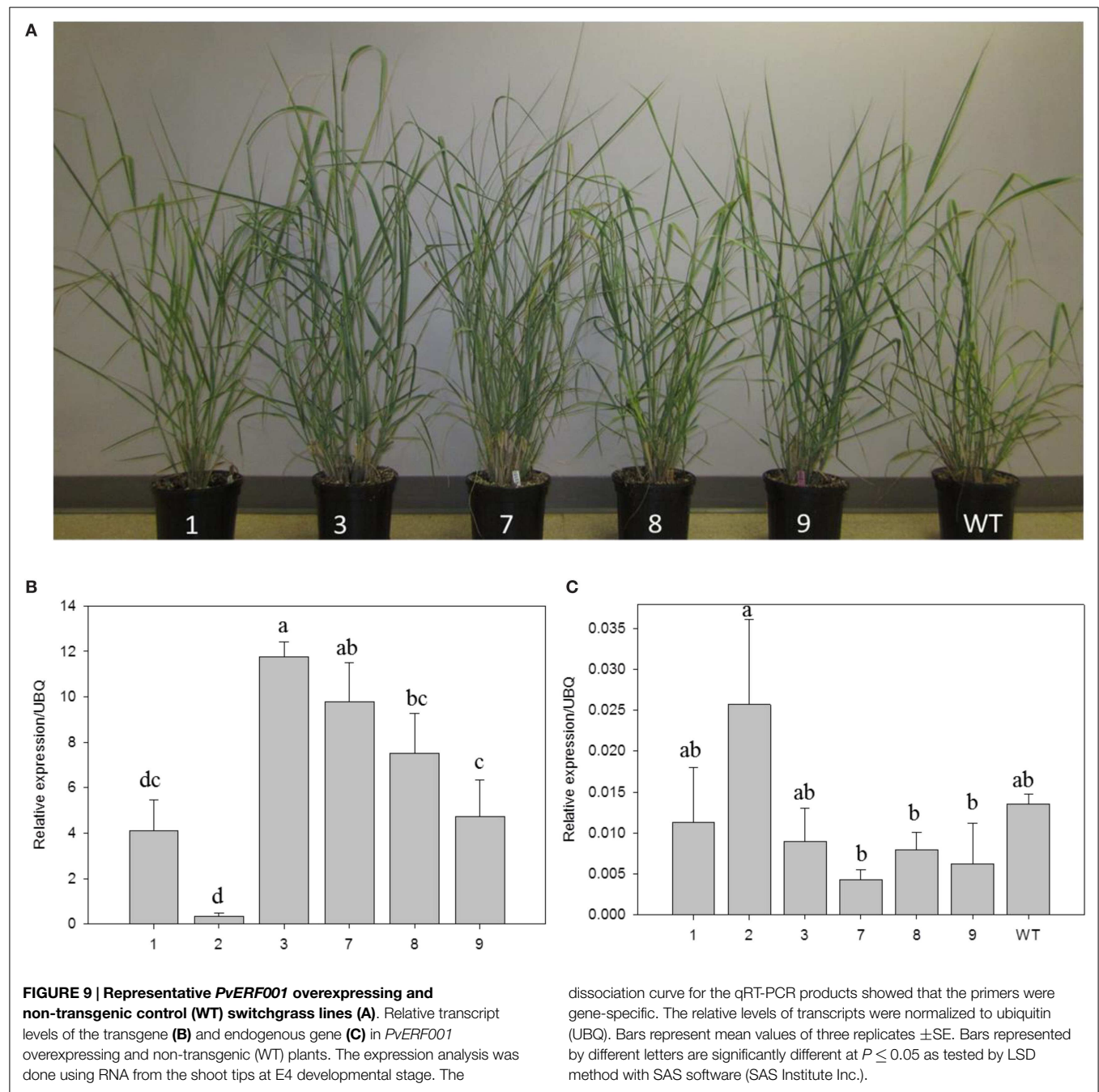


TABLE 2 | Morphology and biomass yields of transgenic switchgrass lines overexpressing *PvERF001* and non-transgenic control (WT) plants.

Lines	Tiller height (cm)	Tiller number	Fresh weight (g)	Dry weight (g)	Plant diameter (cm)	Fresh/dry weight ratio
1	98.9 \pm 2.0 ^b	13.3 \pm 1.5 ^{bc}	40.5 \pm 1.7 ^{cd}	12.8 \pm 0.3 ^c	1.38 \pm 0.06 ^b	3.15 \pm 0.07 ^a
2	105.8 \pm 2.9 ^{ab}	12.3 \pm 2.6 ^c	45.8 \pm 11.3 ^{bcd}	15.1 \pm 3.9 ^{bc}	1.36 \pm 0.06 ^{bc}	3.05 \pm 0.04 ^a
3	115.3 \pm 1.2 ^a	17.7 \pm 1.9 ^{ab}	70.2 \pm 9.9 ^a	21.9 \pm 3.3 ^a	1.54 \pm 0.03 ^a	3.21 \pm 0.08 ^a
7	115.7 \pm 3.6 ^a	21.0 \pm 1.6 ^a	66.7 \pm 3.6 ^{ab}	17.7 \pm 5.5 ^{ab}	1.23 \pm 0.06 ^{cd}	3.03 \pm 0.17 ^a
8	101.3 \pm 2.1 ^b	15.3 \pm 0.9 ^{abc}	42.9 \pm 1.8 ^{cd}	13.3 \pm 0.9 ^c	1.20 \pm 0.02 ^d	3.23 \pm 0.14 ^a
9	109.7 \pm 3.1 ^{ab}	17.0 \pm 0.8 ^{abc}	61.4 \pm 3.6 ^{abc}	16.1 \pm 5.2 ^{ab}	1.44 \pm 0.05 ^{ab}	3.00 \pm 0.11 ^a
WT	85.8 \pm 2.8 ^c	15.3 \pm 1.5 ^{abc}	34.6 \pm 4.7 ^d	10.7 \pm 1.6 ^c	1.05 \pm 0.03 ^e	3.27 \pm 0.08 ^a

Tiller height estimates were determined for each plant by taking the mean of the five tallest tillers within each biological replicate. The fresh and dry biomass measurements were obtained from aboveground plant biomass harvested at similar growth stages. Values are means of three biological replicates \pm SEs ($n = 3$). Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

TABLE 3 | Sugar release by enzymatic hydrolysis in transgenic and non-transgenic control (WT) lines.

Lines	Glucose release (g/g CWR)	Xylose release (g/g CWR)	Total sugar release (g/g CWR)
1	0.214 ± 0.009 ^d	0.175 ± 0.003 ^c	0.389 ± 0.011 ^c
2	0.238 ± 0.008 ^{bc}	0.182 ± 0.014 ^{abc}	0.420 ± 0.018 ^b
3	0.234 ± 0.003 ^{bcd}	0.192 ± 0.008 ^a	0.427 ± 0.006 ^{ab}
7	0.247 ± 0.003 ^{ab}	0.176 ± 0.009 ^{bc}	0.423 ± 0.007 ^b
8	0.261 ± 0.003 ^a	0.188 ± 0.005 ^{ab}	0.448 ± 0.012 ^a
9	0.227 ± 0.020 ^{bcd}	0.188 ± 0.007 ^{ab}	0.415 ± 0.024 ^b
WT	0.225 ± 0.007 ^{cd}	0.181 ± 0.007 ^{abc}	0.405 ± 0.003 ^{bc}

All data are means ± SE (n = 3). CWR, cell wall residues. Values represented by different letters are significantly different at $P \leq 0.05$ as tested by LSD method with SAS software (SAS Institute Inc.).

Discussion

Significance of AP2/ERF TFs for Improvement of Bioenergy Crops

AP2/ERF TFs constitute one of the largest protein superfamilies in plants. These TFs play a role in regulating a wide array of developmental and growth processes. Thus, they are interesting targets for crop genetic engineering and breeding (Licausi et al., 2013; Bhatia and Bosch, 2014). Numerous TFs belonging to this superfamily have been characterized in various plant species and their potential biotechnological applications in crop improvement has focused primarily on biotic and abiotic stress tolerance (Xu et al., 2011; Licausi et al., 2013; Hoang et al., 2014). However, less effort has been made to utilize this potential for genetic improvement of bioenergy feedstocks such as switchgrass (Bhatia and Bosch, 2014). We found this lack of development to be somewhat anachronistic since these TFs are variably associated with plant growth and cell wall biosynthesis, which are directly related to two most important traits to a bioenergy crops, such as switchgrass: biomass and cell wall recalcitrance.

Sequence-Based Classification of Putative AP2/ERF TFs in Switchgrass

With this in mind, we conducted a whole genome search for putative switchgrass AP2/ERF superfamily of TFs and found 207 members (Figure 1; Table 1; Table S2 in Supplementary Material). Based on comparative genome analysis with the published results in rice, foxtail millet, and *Arabidopsis*, the identified proteins were classified into three families, namely AP2, RAV, and ERF with the later further divided into two subfamilies (ERF and DREB) (Nakano et al., 2006; Lata et al., 2014). The number of genes in the DREB subfamily found in switchgrass (55) was comparable with that of rice (56), *Arabidopsis* (57), and *Populus* (66). All three species along with switchgrass have a singleton in their genome. Consistent with the previous report in rice (Nakano et al., 2006), the switchgrass DREB and ERF subfamilies comprise four and seven groups, respectively. Moreover, based on comparative analysis of the AP2/ERF TFs between different plant species, it seems that group XI of ERF subfamily is specific to monocots as the Xb-L was reported only in dicots (Nakano et al., 2006; Liu et al., 2013). In general, the relative distribution of genes within the different groups in each subfamily appears to be conserved between the three plant species (Table 1). Classification

of the switchgrass AP2/ERF TFs into distinct groups was clearly supported by the amino acid sequence-based dendrogram of the identified proteins suggesting robust evolutionary conservation between the superfamily among plant species.

In Silico Predicted Gene Functions and Subcellular Localization of AP2/ERF TFs in Switchgrass

Consistent with the purported role of AP2/ERF proteins as transcriptional regulators of target genes (Magnani et al., 2004), GO analysis predicted that the majority of the switchgrass AP2/ERF genes appear to have DNA-binding activity consistent with the previous observation in foxtail millet (Lata et al., 2014). Therefore, these genes might be associated with the regulation of various biosynthetic processes as well as responses to environmental stimuli as previously demonstrated for numerous genes in other plant species (Xu et al., 2011; Mizoi et al., 2012; Licausi et al., 2013) (Figures 5A,B). The predicted subcellular localization pattern of AP2/ERF superfamily genes in switchgrass, which was mainly to the nucleus as would be expected for transcriptional regulators but also to the plastids and/or mitochondria in addition to the nucleus, was comparable to that reported in foxtail millet (Figure 5C) (Lata et al., 2014). Such multi-localization of the proteins could be attributed to post-translational modifications, protein folding, or interactions with other proteins (Karniely and Pines, 2005), and might serve to facilitate the coordinated regulation of the expression of nuclear and organellar genomes (Duchene and Giege, 2012).

Gene and Protein Sequence Diversity of Switchgrass AP2/ERF TFs

The exon/intron structures of switchgrass AP2/ERF genes were analogous with that of foxtail millet (Lata et al., 2014), castor bean (Xu et al., 2013), rice, and *Arabidopsis* (Nakano et al., 2006). Consistent with these species, we observed a high diversity in the distribution of the intron regions of AP2 genes versus a single or no intron in most genes in the ERF and RAV families (Figures 2 and 4). The pattern of intron distribution within the ORF and their splicing phases was highly conserved in genes within specific groups as reported in castor bean (Xu et al., 2013). Consistent with the observation in rice, the majority of proteins in the groups or subfamilies of switchgrass AP2/ERF superfamily could be distinguished by the presence of one or more diagnostic motifs located outside the AP2 domain region (Table S4 in Supplementary Material) (Rashid et al., 2012). These groups or subfamily-specific conservation in gene structures and protein motifs supported the accuracy of the predicted cluster relationships between the switchgrass AP2/ERF TFs.

AP2/ERF TFs that function as repressors or activators of specific target genes are distinguished by the presence of conserved motifs called repression domains (RD) that are highly conserved, or by the presence of activation domains which are generally less conserved (Licausi et al., 2013). One of the characteristic motif in AP2/ERF transcriptional activators is the activation domain, EDLL motif (Tiwari et al., 2012), while repressors have unique RD namely the ERF-associated amphiphilic repression (EAR) motif (LxLxL or DLNxxP)

(Kagale and Rozwadowski, 2011) and B3 repression domain (BRD: R/KLFGV) motif (Ikeda and Ohme-Takagi, 2009). Analysis of the switchgrass AP2/ERF TF sequences also indicated the presence of these motifs in many proteins (Table S4 in Supplementary Material). For instance, many genes in group IX of ERF subfamily appear to be transcriptional activators due to the presence of motif M10, which is an EDLL-like motif. Moreover, this motif is rich in acidic amino residues which has been suggested as the characteristics of transcriptional activators (Licausi et al., 2013). Majority of the ERF family TFs in group VIII and DREB family TFs in group I displayed a DLNxxP-like motifs. Four TFs belonging to the AP2 family (PvERF204, PvERF205, PvERF206, and PvERF207) also displayed similar EAR motif while PvERF203 and PvERF207 harbors DLELSL and NLDLS-like RD, respectively. Similarly, switchgrass TFs in RAV family also displayed unique repression domain, RLFGV (Ikeda and Ohme-Takagi, 2009). ERF subfamily TFs in groups VI and VI-L share a characteristic motif at the N-terminus (M18), also known as the cytokinin responsive factor (CRF) domain in *Arabidopsis* that is also shared by rice ERF genes belonging to same group in rice ERF subfamily (Nakano et al., 2006). Genes containing the CRF domain (VI and VI-L) were shown to be responsive to cytokinin (Rashotte et al., 2006). The distinguishing N-terminal motif in group VII ERF subfamily proteins, M25 was conserved in both *Arabidopsis* and rice as described previously (Nakano et al., 2006). This motif was shown to dictate the stability of proteins based on the level of oxygen via N-end rule pathway (Dubouzet et al., 2003; Licausi et al., 2011). DREB genes in rice with characteristic LWSY motif have been shown to function in regulation of drought, cold, and salinity responsive gene expression (Dubouzet et al., 2003). Switchgrass genes belonging to group III in DREB subfamily (PvERF133, PvERF134, PvERF135, PvERF136, PvERF137, PvERF139, PvERF140, PvERF141, PvERF142, PvERF143, PvERF145, and PvERF146) displayed LWSY conserved motif (M21) at the C-terminal and thus may play similar roles. No information is available in the literature on some of the conserved motifs identified here including M8, M13, M14, M15, M17, and M24 (Table S4 in Supplementary Material), which might potentially be specific to switchgrass.

Diverse Expression Profiles of Switchgrass AP2/ERF TFs and Functional Implications

Differential expression of genes according to developmental stages and tissue or organ types may provide an insight into the specialized biological processes that are taking place in the specific plant parts (Cassan-Wang et al., 2013; Zhang et al., 2013). The observed pattern of expression for the majority of switchgrass AP2/ERF genes at different stages of plant development as well as in different tissues/organ types highlight the significance of these genes in the regulation of various plant growth and developmental processes at the specific stages (Figures 6 and 7; Figures S1 and S2 in Supplementary Material). One of the engrossing observations in this study is the association of the expression of numerous genes with tissues/organs undergoing lignification or secondary cell wall development/modification, suggesting that these genes may have intrinsic association with the regulatory

machinery of cell wall formation/lignification, which is not as well characterized compared to their roles in stress response (Licausi et al., 2013; Bhatia and Bosch, 2014). Activation of genes responsible for cell wall modification has already been reported to be key during the initiation of seed germination in barley (Sreenivasulu et al., 2008; An and Lin, 2011). In agreement with this, we reported here the transcriptional upregulation of ERF (PvERF057, PvERF068, PvERF088, and PvERF119), DREB (PvERF101, PvERF102, and PvERF148), and AP2 genes (PvERF193, PvERF201, and PvERF204) during the initiation of seed germination as well as in vascular bundles and internode sections. Moreover, the observed robust expression of 14 DREB, 17 ERF, and 3 AP2 genes in tissues or organs undergoing active lignification (vascular bundles, top or middle internode sections as well as roots) but less robust expression in less lignified tissues (leaves) also supports this assertion (Figure 7). It should also be noted that the transcript levels of several of these genes showed a relative increase with the developmental stage of the plants (Figure 7; Figure S1 in Supplementary Material) while exhibiting only marginal expression in less lignified tissues such as inflorescence meristem and germinating seedlings (Figure 6; Figure S2 in Supplementary Material). Differential gene expression profiling between elongating and non-elongating internodes in maize was used to identify a total of seven AP2/ERF TFs that are highly expressed in non-elongating internodes undergoing secondary wall development suggesting that these genes may involve in the regulation of secondary cell wall formation (Bosch et al., 2011). Moreover, recent study in *Arabidopsis* and rice identified several putative secondary cell wall-related AP2/ERF TFs based on preferential expression in secondary cell wall-related tissues and coexpression analysis (Cassan-Wang et al., 2013; Hirano et al., 2013a; Bhatia and Bosch, 2014). Some of the switchgrass genes identified in this study (PvERF037, PvERF115, PvERF116, PvERF143, PvERF148, and PvERF164) appear to be putative homologs of maize, rice, and *Arabidopsis* genes identified in the aforementioned studies. Overexpression of *Populus* ERF genes in wood-forming tissues of hybrid aspen was recently shown to result in modified stem growth (including increased stem diameter following the overexpression of five different ERF genes), reduced lignification, and enhanced carbohydrate content (cellulose) in the wood of transgenic lines hinting that these TFs may indeed interact with the transcriptional machinery regulating cell wall biosynthesis (Vahala et al., 2013). Another evidence supporting this is a recent study suggesting that an ERF TF from loquat fruit (*Eriobotrya japonica*) (EjAP2-1) is an indirect transcriptional repressor of lignin biosynthesis via interaction with EjMYB1 TFs (Zeng et al., 2015).

Overexpression of PvERF001 Improved Biomass Productivity and Sugar Release Efficiency in Switchgrass

Based on global gene coexpression analysis, the rice homolog of AtSHN2, OsSHN (OsERF057) was proposed to have a native association with cell wall regulatory and biosynthetic pathways, yet this was not experimentally verified (Ambavaram et al., 2011). In this study, we investigated whether PvERF001, the closest putative switchgrass homolog of these genes based on clustering,

sequence alignment analysis, and the sharing of conserved motifs (mm and cm) specific to *Arabidopsis* SHN clade of TFs and the rice SHN, may participate in the regulation of cell wall biosynthesis (Figure 8). Our results suggest that PvERF001 may not be directly involved in the regulation cell wall biosynthesis though its transgenic overexpression resulted in increased sugar release efficiency (Figure S7 in Supplementary Material; Table 3). Despite the observed reduction in relative expression of some lignin biosynthetic genes and their transcriptional regulators in switchgrass that seem to relate with the results in rice overexpressing *AtSHN2*, no significant changes in the lignin content and composition was detected in transgenic switchgrass in contrast to the reduced lignin content observed in rice overexpressing *AtSHN2* (Ambavaram et al., 2011) (Figures S7, S8, and S9A in Supplementary Material). The increased sugar release might be attributed to altered storage carbohydrates such as starches as recently reported in *Arabidopsis* where ectopic expression of rice ERF TF (SUB1A-1) gene resulted in improved enzymatic saccharification efficiency via increased level of starch (Nunez-Lopez et al., 2015). Similar results were obtained from overexpression of maize *corngrass1* microRNA in switchgrass (Chuck et al., 2011). However, whether PvERF001 is associated with starch biosynthesis remains to be determined. Moreover, in contrast to the previous reports where heterologous expression of *AtSHN2* in rice did not significantly affect the growth characteristics of transgenic lines (Ambavaram et al., 2011), overexpression of *PvERF001* resulted in increased plant growth including plant height, stem diameter and aboveground biomass weight in transgenic lines (Table 2). The discrepancy in lignin content and biomass productivity traits between the *AtSHN2* and *PvERF001* may indicate the differences in functional specialization between the two genes in monocots and dicots even though sequence analysis seems to suggest that they might be homologs. The fact that overexpression of *AtSHN* genes in *Arabidopsis* rather showed association with the regulation of wax, cutin, and pectin biosynthesis supports this assertion (Aharoni et al., 2004; Shi et al., 2011). Moreover, recent study showed that the homolog of *Arabidopsis* SHN genes in tomato (*SlERF52*) was expressed mainly in the abscission zone and functionally associated with the regulation of the pedicel abscission zone-specific transcription of genes including cell wall-hydrolytic enzymes (polygalacturonase and Cellulase) required for abscission (Nakano et al., 2014). These differences in the expression pattern and function may suggest functional divergence between *SlERF52* and its *Arabidopsis* homologs. Functional divergence between homologous TFs in monocots and dicots has also been reported in previous studies involving the homologs of *AtMYB58/63*, which is a known activator of lignin biosynthesis that did not appear to play similar roles in rice (Hirano et al., 2013b).

A recent study involving overexpression of rice homolog of *AtSHN2*, *OsSHN*, in rice showed enhanced tolerance of transgenic plants to water deprivation and association of the gene with the regulation of wax and cutin biosynthesis and hence named rice wax synthesis regulatory gene (*OsWR2*) (Zhou et al., 2014). The closest homolog of this gene, *OsERF012* (*OsWR1*), was also shown to be induced by drought stress and involved in the regulation of wax synthesis (Wang et al., 2012). Therefore, we examined

whether PvERF001 might be involved in the regulation of wax and cutin biosynthesis. Consistent with previous studies in rice, relative increase in leaf water retention capacity was detected in transgenic plants though the effect on the expression of wax and cutin biosynthetic genes was minimal (Figures S5 and S6 in Supplementary Material). Possible explanation for the observed differences between overexpression of rice and switchgrass homologs might be an indication of the functional divergence in the switchgrass genes due to gene duplication. This may explain the discrepancy between transgenic rice overexpressing rice *SHN* (*OsWR2*) exhibiting reduction in plant height but increase in the number of tillers (Zhou et al., 2014) and transgenic switchgrass overexpressing *PvERF001* showing increased plant height but no difference in number of tillers. This suggests that ERF genes might functionally be highly diversified and PvERF001 may be part of a different pathway than we anticipated such as regulation of responses to biotic stress or other abiotic stress or regulation of cell elongation or division in coordination with the cytokinin pathway, with the latter perhaps explaining the observed increase in biomass and vegetative growth in transgenic lines.

In summary, the expression profiling of the switchgrass AP2/ERF genes provides baseline information as to the putative roles of these genes and thus a useful resource for future reverse genetic studies to characterize genes for economically important bioenergy crops. With the current advancements in switchgrass research and establishment of efficient transformation system, this inventory of genes along with the information provided here could facilitate our understanding regarding the functional roles of AP2/ERF TFs in plant growth and development. Furthermore, it would aid in the identification of potential target genes that may be used to improve stress adaptation, plant productivity, and sugar release efficiency in bioenergy feedstocks such as switchgrass. The increased biomass yield and sugar release efficiency from overexpressing *PvERF001* highlight the potential of these TFs for improvement of bioenergy feedstocks.

Author Contributions

WW designed and performed the experiments, analyzed the data, and wrote the manuscript. MM participated in experimental design and data analysis, assisted with revisions to the manuscript and coordination of the study. GT, RS, SD, and MD assisted with performing lignin and sugar release assays and contributed in revision of the manuscript. CS conceived the study and its design and coordination, and assisted with revisions to the manuscript. All authors read and consented to the final version of the manuscript.

Acknowledgments

We thank Angela Ziebell, Erica Gjersing, Crissa Doeppke, and Melvin Tucker for their assistance with the cell wall characterization and Susan Holladay for her assistance with data entry into LIMS. We thank Wayne Parrot for providing the switchgrass SA1 clone. This work was supported by funding from the BioEnergy Science Center (DE-PS02-06ER64304). The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research

Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. We also thank Tennessee Agricultural Experiment Station for providing partial financial support for WW.

References

- Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., Van Arkel, G., and Pereira, A. (2004). The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis*. *Plant Cell* 16, 2463–2480. doi:10.1105/tpc.104.022897
- Ambavaram, M. M., Krishnan, A., Trijatmiko, K. R., and Pereira, A. (2011). Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. *Plant Physiol.* 155, 916–931. doi:10.1104/pp.110.168641
- Ambavaram, M. M. R., Basu, S., Krishnan, A., Ramegowda, V., Batlang, U., Rahman, L., et al. (2014). Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nat. Commun.* 5, 5302. doi:10.1038/ncomms6302
- An, Y.-Q., and Lin, L. (2011). Transcriptional regulatory programs underlying barley germination and regulatory functions of gibberellin and abscisic acid. *BMC Plant Biol.* 11:105. doi:10.1186/1471-2229-11-105
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Baxter, H., Poovaiah, C., Yee, K., Mazarei, M., Rodriguez, M. Jr., Thompson, O., et al. (2015). Field evaluation of transgenic switchgrass plants overexpressing *PvMYB4* for reduced biomass recalcitrance. *Bioenerg. Res.* 1–12. doi:10.1007/s12155-014-9570-1
- Baxter, H. L., Mazarei, M., Labbe, N., Kline, L. M., Cheng, Q., Windham, M. T., et al. (2014). Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnol. J.* 12, 914–924. doi:10.1111/pbi.12195
- Bhatia, R., and Bosch, M. (2014). Transcriptional regulators of *Arabidopsis* secondary cell wall formation: tools to re-program and improve cell wall traits. *Front. Plant Sci.* 5:192. doi:10.3389/fpls.2014.00192
- Bosch, M., Mayer, C.-D., Cookson, A., and Donnison, I. S. (2011). Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *J. Exp. Bot.* 62, 3545–3561. doi:10.1093/jxb/err045
- Bouaziz, D., Pirrello, J., Charfeddine, M., Hammami, A., Jbir, R., Dhieb, A., et al. (2013). Overexpression of *StDREB1* transcription factor increases tolerance to salt in transgenic potato plants. *Mol. Biotechnol.* 54, 803–817. doi:10.1007/s12033-012-9628-2
- Burris, J., Mann, D. J., Joyce, B., and Stewart, C. N. Jr. (2009). An improved tissue culture system for embryogenic callus production and plant regeneration in switchgrass (*Panicum virgatum* L.). *Bioenerg. Res.* 2, 267–274. doi:10.1007/s12155-009-9048-8
- Cassan-Wang, H., Goué, N., Saidi, M. N., Legay, S., Sivadon, P., Goffner, D., et al. (2013). Identification of novel transcription factors regulating secondary cell wall formation in *Arabidopsis*. *Front. Plant Sci.* 4:189. doi:10.3389/fpls.2013.00189
- Chuck, G. S., Tobias, C., Sun, L., Kraemer, F., Li, C., Dibble, D., et al. (2011). Overexpression of the maize *Corngrass1* microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17550–17555. doi:10.1073/pnas.1113971108
- Conesa, A., and Gotz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 619832. doi:10.1155/2008/619832
- Decker, S. R., Carlile, M., Selig, M. J., Doepfke, C., Davis, M., Sykes, R., et al. (2012). Reducing the effect of variable starch levels in biomass recalcitrance screening. *Methods Mol. Biol.* 908, 181–195. doi:10.1007/978-1-61779-956-3_17
- Dong, L., Cheng, Y., Wu, J., Cheng, Q., Li, W., Fan, S., et al. (2015). Overexpression of *GmERF5*, a new member of the soybean EAR motif-containing ERF transcription factor, enhances resistance to *Phytophthora sojae* in soybean. *J. Exp. Bot.* 66, 2635–2647. doi:10.1093/jxb/erv078
- Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–763. doi:10.1046/j.1365-3113X.2003.01661.x
- Duchene, A. M., and Giege, P. (2012). Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? *Front. Plant Sci.* 3:221. doi:10.3389/fpls.2012.00221
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340
- Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., et al. (1996). *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8, 155–168. doi:10.1105/tpc.8.2.155
- Fang, Z., Zhang, X., Gao, J., Wang, P., Xu, X., Liu, Z., et al. (2015). A buckwheat (*Fagopyrum esculentum*) DRE-Binding transcription factor gene, *FeDREB1*, enhances freezing and drought tolerance of transgenic *Arabidopsis*. *Plant Mol. Biol. Rep.* 1–16. doi:10.1007/s11105-015-0851-4
- Fu, C., Mielenz, J. R., Xiao, X., Ge, Y., Hamilton, C. Y., Rodriguez, M., et al. (2011a). Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3803–3808. doi:10.1073/pnas.1100310108
- Fu, C., Xiao, X., Xi, Y., Ge, Y., Chen, F., Bouton, J., et al. (2011b). Downregulation of *Cinnamyl Alcohol Dehydrogenase (CAD)* leads to improved saccharification efficiency in switchgrass. *Bioenerg. Res.* 4, 153–164. doi:10.1007/s12155-010-9109-z
- Fujita, Y., Fujita, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). ABA-mediated transcriptional regulation in response to osmotic stress in plants. *J. Plant Res.* 124, 509–525. doi:10.1007/s10265-011-0412-3
- Grewal, D., Gill, R., and Gosal, S. S. (2006). Influence of antibiotic cefotaxime on somatic embryogenesis and plant regeneration in indica rice. *Biotechnol. J.* 1, 1158–1162. doi:10.1002/biot.200600139
- Guo, A. Y., Zhu, Q. H., Chen, X., and Luo, J. C. (2007). [GSDS: a gene structure display server]. *Yi Chuan* 29, 1023–1026. doi:10.1360/yc-007-1023
- Guo, Z. J., Chen, X. J., Wu, X. L., Ling, J. Q., and Xu, P. (2004). Overexpression of the AP2/EREBP transcription factor *OPBP1* enhances disease resistance and salt tolerance in tobacco. *Plant Mol. Biol.* 55, 607–618. doi:10.1007/s11103-004-1521-3
- Hardin, C. F., Fu, C., Hisano, H., Xiao, X., Shen, H., Stewart, C. N. Jr., et al. (2013). Standardization of switchgrass sample collection for cell wall and biomass trait analysis. *Bioenerg. Res.* 6, 755–762. doi:10.1007/s12155-012-9292-1
- Hattori, Y., Nagai, K., Furukawa, S., Song, X. J., Kawano, R., Sakakibara, H., et al. (2009). The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* 460, 1026–1030. doi:10.1038/nature08258
- Hirano, K., Aya, K., Morinaka, Y., Nagamatsu, S., Sato, Y., Antonio, B. A., et al. (2013a). Survey of genes involved in rice secondary cell wall formation through a co-expression network. *Plant Cell Physiol.* 54, 1803–1821. doi:10.1093/pcp/pct121
- Hirano, K., Kondo, M., Aya, K., Miyao, A., Sato, Y., Antonio, B. A., et al. (2013b). Identification of transcription factors involved in rice secondary cell wall formation. *Plant Cell Physiol.* 54, 1791–1802. doi:10.1093/pcp/pct122
- Hoang, X. L. T., Thu, N. B. A., Thao, N. P., and Tran, L.-S. P. (2014). “Transcription factors in abiotic Stress responses: their potentials in crop improvement,” in *Improvement of Crops in the Era of Climatic Changes*, eds P. Ahmad, M. R. Wani, M. M. Azooz and L.-S. Phan Tran (New York, NY: Springer), 337–366.
- Hong, J. P., and Kim, W. T. (2005). Isolation and functional characterization of the *Ca-DREBLP1* gene encoding a dehydration-responsive element binding-factor-like protein 1 in hot pepper (*Capsicum annuum* L. cv. Pukang). *Planta* 220, 875–888. doi:10.1007/s00425-004-1412-5
- Horstman, A., Willemsen, V., Boutilier, K., and Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends Plant Sci.* 19, 146–157. doi:10.1016/j.tplants.2013.10.010
- Ikeda, M., and Ohme-Takagi, M. (2009). A novel group of transcriptional repressors in *Arabidopsis*. *Plant Cell Physiol.* 50, 970–975. doi:10.1093/pcp/pcp048
- Ito, Y., Katsura, K., Maruyama, K., Tajiri, T., Kobayashi, M., Seki, M., et al. (2006). Functional analysis of rice DREB1/CBF-type transcription factors involved

Supplementary Material

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00101>

- in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol.* 47, 141–153. doi:10.1093/pcp/pci230
- Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., and Thomashow, M. F. (1998). *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. *Science* 280, 104–106. doi:10.1126/science.280.5360.104
- Jofuku, K. D., Omidyar, P. K., Gee, Z., and Okamoto, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene *APETALA2*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3117–3122. doi:10.1073/pnas.0409893102
- Kagale, S., and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* 6, 141–146. doi:10.4161/epi.6.2.13627
- Karnieli, S., and Pines, O. (2005). Single translation – dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Rep.* 6, 420–425. doi:10.1038/sj.embor.7400394
- King, Z. R., Bray, A. L., Lafayette, P. R., and Parrott, W. A. (2014). Biolistic transformation of elite genotypes of switchgrass (*Panicum virgatum* L.). *Plant Cell Rep.* 33, 313–322. doi:10.1007/s00299-013-1531-1
- Lasserre, E., Jobet, E., Llauro, C., and Delseny, M. (2008). AtERF38 (At2g35700), an AP2/ERF family transcription factor gene from *Arabidopsis thaliana*, is expressed in specific cell types of roots, stems and seeds that undergo suberization. *Plant Physiol. Biochem.* 46, 1051–1061. doi:10.1016/j.plaphy.2008.07.003
- Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., and Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e113092. doi:10.1371/journal.pone.0113092
- Li, R., and Qu, R. (2011). High throughput *Agrobacterium*-mediated switchgrass transformation. *Biomass Bioenergy* 35, 1046–1054. doi:10.1016/j.biombioe.2010.11.025
- Licausi, F., Giorgi, F. M., Zenoni, S., Osti, F., Pezzotti, M., and Perata, P. (2010). Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics* 11:719. doi:10.1186/1471-2164-11-719
- Licausi, F., Kosmacz, M., Weits, D. A., Giuntoli, B., Giorgi, F. M., Voesenek, L. A. C. J., et al. (2011). Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature* 479, 419–422. doi:10.1038/nature10536
- Licausi, F., Ohme-Takagi, M., and Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* 199, 639–649. doi:10.1111/nph.12291
- Liu, Z., Kong, L., Zhang, M., Lv, Y., Liu, Y., Zou, M., et al. (2013). Genome-wide identification, phylogeny, evolution and expression patterns of AP2/ERF genes and cytokinin response factors in *Brassica rapa* ssp. *pekinensis*. *PLoS ONE* 8:e83444. doi:10.1371/journal.pone.0083444
- Magnani, E., Sjolander, K., and Hake, S. (2004). From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell* 16, 2265–2277. doi:10.1105/tpc.104.023135
- Mann, D. G., Lafayette, P. R., Abercrombie, L. L., King, Z. R., Mazarei, M., Halter, M. C., et al. (2012). Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnol. J.* 10, 226–236. doi:10.1111/j.1467-7652.2011.00658.x
- Mittal, A., Gampala, S. S., Ritchie, G. L., Payton, P., Burke, J. J., and Rock, C. D. (2014). Related to ABA-Insensitive3 (ABI3)/Viviparous1 and AtABI5 transcription factor coexpression in cotton enhances drought stress adaptation. *Plant Biotechnol. J.* 12, 578–589. doi:10.1111/pbi.12162
- Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta* 1819, 86–96. doi:10.1016/j.bbagr.2011.08.004
- Moore, K. J., Moser, L. E., Vogel, K. P., Waller, S. S., Johnson, B. E., and Pedersen, J. F. (1991). Describing and quantifying growth stages of perennial forage grasses. *Agron. J.* 83, 1073–1077. doi:10.2134/agronj1991.00021962008300060027x
- Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi:10.1111/j.1399-3054.1962.tb08052.x
- Nakano, T., Fujisawa, M., Shima, Y., and Ito, Y. (2014). The AP2/ERF transcription factor SlERF52 functions in flower pedicel abscission in tomato. *J. Exp. Bot.* 65, 3111–3119. doi:10.1093/jxb/eru154
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 140, 411–432. doi:10.1104/pp.105.073783
- Nunez-Lopez, L., Aguirre-Cruz, A., Barrera-Figueroa, B. E., and Pena-Castro, J. M. (2015). Improvement of enzymatic saccharification yield in *Arabidopsis thaliana* by ectopic expression of the rice *SUB1A-1* transcription factor. *PeerJ* 3, e817. doi:10.7717/peerj.817
- Oh, S. J., Kim, Y. S., Kwon, C. W., Park, H. K., Jeong, J. S., and Kim, J. K. (2009). Overexpression of the transcription factor AP37 in rice improves grain yield under drought conditions. *Plant Physiol.* 150, 1368–1379. doi:10.1104/pp.109.137554
- Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7, 173–182. doi:10.1105/tpc.7.2.173
- Qin, F., Kakimoto, M., Sakuma, Y., Maruyama, K., Osakabe, Y., Tran, L. S., et al. (2007). Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in *Zea mays* L. *Plant J.* 50, 54–69. doi:10.1111/j.1365-313X.2007.03034.x
- Rashid, M., Guangyuan, H., Guangxiao, Y., Hussain, J., and Xu, Y. (2012). AP2/ERF transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. *Evol. Bioinform. Online* 8, 321–355. doi:10.4137/EBO.S9369
- Rashotte, A. M., Mason, M. G., Hutchison, C. E., Ferreira, F. J., Schaller, G. E., and Kieber, J. J. (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11081–11085. doi:10.1073/pnas.0602038103
- Shen, H., Fu, C., Xiao, X., Ray, T., Tang, Y., Wang, Z., et al. (2009). Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. *Bioenerg. Res.* 2, 233–245. doi:10.1007/s12155-009-9058-6
- Shen, H., He, X., Poovaiah, C. R., Wuddineh, W. A., Ma, J., Mann, D. G., et al. (2012). Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytol.* 193, 121–136. doi:10.1111/j.1469-8137.2011.03922.x
- Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., et al. (2013a). A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25, 4342–4361. doi:10.1105/tpc.113.118828
- Shen, H., Poovaiah, C. R., Ziebell, A., Tschaplinski, T. J., Pattathil, S., Gjersing, E., et al. (2013b). Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol. Biofuels* 6, 71. doi:10.1186/1754-6834-6-71
- Shi, J. X., Malitsky, S., De Oliveira, S., Branigan, C., Franke, R. B., Schreiber, L., et al. (2011). SHINE transcription factors act redundantly to pattern the archetypal surface of *Arabidopsis* flower organs. *PLoS Genet.* 7:e1001388. doi:10.1371/journal.pgen.1001388
- Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., et al. (2008). Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiol.* 146, 1738–1758. doi:10.1104/pp.107.111781
- Sun, S., Yu, J. P., Chen, F., Zhao, T. J., Fang, X. H., Li, Y. Q., et al. (2008). TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*. *J. Biol. Chem.* 283, 6261–6271. doi:10.1074/jbc.M706800200
- Sykes, R., Yung, M., Novaes, E., Kirst, M., Peter, G., and Davis, M. (2009). High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. *Methods Mol. Biol.* 581, 169–183. doi:10.1007/978-1-60761-214-8_12
- Tamura, K., Stecher, G., Peterson, D., Filipinski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi:10.1093/molbev/mst197
- Tiwari, S. B., Belachew, A., Ma, S. F., Young, M., Ade, J., Shen, Y., et al. (2012). The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* 70, 855–865. doi:10.1111/j.1365-313X.2012.04935.x
- Tschaplinski, T. J., Standaert, R. F., Engle, N. L., Martin, M. Z., Sangha, A. K., Parks, J. M., et al. (2012). Down-regulation of the *caffeic acid O-methyltransferase* gene in switchgrass reveals a novel monolignol analog. *Biotechnol. Biofuels* 5, 71. doi:10.1186/1754-6834-5-71
- Vahala, J., Felten, J., Love, J., Gorzsas, A., Gerber, L., Lamminmaki, A., et al. (2013). A genome-wide screen for ethylene-induced ethylene response factors (ERFs)

- in hybrid aspen stem identifies ERF genes that modify stem growth and wood properties. *New Phytol.* 200, 511–522. doi:10.1111/nph.12386
- Van Raemdonck, D., Pesquet, E., Cloquet, S., Beeckman, H., Boerjan, W., Goffner, D., et al. (2005). Molecular changes associated with the setting up of secondary growth in aspen. *J. Exp. Bot.* 56, 2211–2227. doi:10.1093/jxb/eri221
- Wang, L., Wang, C., Qin, L., Liu, W., and Wang, Y. (2015). TheERF1 regulates its target genes via binding to a novel cis-acting element in response to salt stress. *J. Integr. Plant Biol.* doi:10.1111/jipb.12335
- Wang, Y., Wan, L., Zhang, L., Zhang, Z., Zhang, H., Quan, R., et al. (2012). An ethylene response factor OsWR1 responsive to drought stress transcriptionally activates wax synthesis related genes and increases wax production in rice. *Plant Mol. Biol.* 78, 275–288. doi:10.1007/s11103-011-9861-2
- Wuddineh, W. A., Mazarei, M., Zhang, J., Poovaiah, C. R., Mann, D. G., Ziebell, A., et al. (2015). Identification and overexpression of *gibberellin 2-oxidase (GA2ox)* in switchgrass (*Panicum virgatum* L.) for improved plant architecture and reduced biomass recalcitrance. *Plant Biotechnol. J.* 13, 636–647. doi:10.1111/pbi.12287
- Xu, W., Li, F., Ling, L., and Liu, A. (2013). Genome-wide survey and expression profiles of the AP2/ERF family in castor bean (*Ricinus communis* L.). *BMC Genomics* 14:785. doi:10.1186/1471-2164-14-785
- Xu, Z. S., Chen, M., Li, L. C., and Ma, Y. Z. (2011). Functions and application of the AP2/ERF transcription factor family in crop improvement. *J. Integr. Plant Biol.* 53, 570–585. doi:10.1111/j.1744-7909.2011.01062.x
- Yuan, J. S., Tiller, K. H., Al-Ahmad, H., Stewart, N. R., and Stewart, C. N. Jr. (2008). Plants to power: bioenergy to fuel the future. *Trends Plant Sci.* 13, 421–429. doi:10.1016/j.tplants.2008.06.001
- Zeng, J. K., Li, X., Xu, Q., Chen, J. Y., Yin, X. R., Ferguson, I. B., et al. (2015). *EjAP2-1*, an AP2/ERF gene, is a novel regulator of fruit lignification induced by chilling injury, via interaction with *EjMYB* transcription factors. *Plant Biotechnol. J.* doi:10.1111/pbi.12351
- Zhang, C. H., Shangguan, L. F., Ma, R. J., Sun, X., Tao, R., Guo, L., et al. (2012). Genome-wide analysis of the AP2/ERF superfamily in peach (*Prunus persica*). *Genet. Mol. Res.* 11, 4789–4809. doi:10.4238/2012.October.17.6
- Zhang, H., Liu, W., Wan, L., Li, F., Dai, L., Li, D., et al. (2010a). Functional analyses of ethylene response factor JERF3 with the aim of improving tolerance to drought and osmotic stress in transgenic rice. *Transgenic Res.* 19, 809–818. doi:10.1007/s11248-009-9357-x
- Zhang, Z., Li, F., Li, D., Zhang, H., and Huang, R. (2010b). Expression of ethylene response factor JERF1 in rice improves tolerance to drought. *Planta* 232, 765–774. doi:10.1007/s00425-010-1208-8
- Zhang, J. Y., Lee, Y. C., Torres-Jerez, I., Wang, M., Yin, Y., Chou, W. C., et al. (2013). Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). *Plant J.* 74, 160–173. doi:10.1111/tpj.12104
- Zhang, Z., and Huang, R. (2010). Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor TERF2/LeERF2 is modulated by ethylene biosynthesis. *Plant Mol. Biol.* 73, 241–249. doi:10.1007/s11103-010-9609-4
- Zhou, X., Jenks, M., Liu, J., Liu, A., Zhang, X., Xiang, J., et al. (2014). Overexpression of transcription factor OsWR2 regulates wax and cutin biosynthesis in rice and enhances its tolerance to water deficit. *Plant Mol. Biol. Rep.* 32, 719–731. doi:10.1007/s11105-013-0687-8
- Zhuang, J., Cai, B., Peng, R. H., Zhu, B., Jin, X. F., Xue, Y., et al. (2008). Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochem. Biophys. Res. Commun.* 371, 468–474. doi:10.1016/j.bbrc.2008.04.087

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Wuddineh, Mazarei, Turner, Sykes, Decker, Davis and Stewart. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phenotypic Changes in Transgenic Tobacco Plants Overexpressing Vacuole-Targeted *Thermotoga maritima* BglB Related to Elevated Levels of Liberated Hormones

Quynh Anh Nguyen¹, Dae-Seok Lee², Jakyun Jung¹ and Hyeun-Jong Bae^{1,2*}

¹ Department of Bioenergy Science and Technology, Chonnam National University, Gwangju, South Korea, ² Bio-Energy Research Center, Chonnam National University, Gwangju, South Korea

OPEN ACCESS

Edited by:

Robert Henry,
The University of Queensland,
Australia

Reviewed by:

Tianju Chen,
Chinese Academy of Sciences, China
Chiranjeevi Thulluri,
Jawaharlal Nehru Technological
University Hyderabad, India

*Correspondence:

Hyeun-Jong Bae
baehj@chonnam.ac.kr

Specialty section:

This article was submitted to
Bioenergy and Biofuels, a section of
the journal Frontiers in Bioengineering
and Biotechnology

Received: 18 August 2015

Accepted: 26 October 2015

Published: 09 November 2015

Citation:

Nguyen QA, Lee D-S, Jung J and
Bae H-J (2015) Phenotypic Changes
in Transgenic Tobacco Plants
Overexpressing Vacuole-Targeted
Thermotoga maritima BglB Related
to Elevated Levels of
Liberated Hormones.
Front. Bioeng. Biotechnol. 3:181.
doi: 10.3389/fbioe.2015.00181

The hyperthermostable β -glucosidase BglB of *Thermotoga maritima* was modified by adding a short C-terminal tetrapeptide (AFVY, which transports phaseolin to the vacuole, to its C-terminal sequence). The modified β -glucosidase BglB was transformed into tobacco (*Nicotiana tabacum* L.) plants. We observed a range of significant phenotypic changes in the transgenic plants compared to the wild-type (WT) plants. The transgenic plants had faster stem growth, earlier flowering, enhanced root systems development, an increased biomass biosynthesis rate, and higher salt stress tolerance in young plants compared to WT. In addition, programmed cell death was enhanced in mature plants. Furthermore, the C-terminal AFVY tetrapeptide efficiently sorted *T. maritima* BglB into the vacuole, which was maintained in an active form and could perform its glycoside hydrolysis function on hormone conjugates, leading to elevated hormone [abscisic acid (ABA), indole 3-acetic acid (IAA), and cytokinin] levels that likely contributed to the phenotypic changes in the transgenic plants. The elevation of cytokinin led to upregulation of the transcription factor *WUSCHEL*, a homeodomain factor that regulates the development, division, and reproduction of stem cells in the shoot apical meristems. Elevation of IAA led to enhanced root development, and the elevation of ABA contributed to enhanced tolerance to salt stress and programmed cell death. These results suggest that overexpressing vacuole-targeted *T. maritima* BglB may have several advantages for molecular farming technology to improve multiple targets, including enhanced production of the β -glucosidase BglB, increased biomass, and shortened developmental stages, that could play pivotal roles in bioenergy and biofuel production.

Keywords: *Thermotoga maritima*, hyperthermostable β -glucosidase BglB, C-terminal AFVY tetrapeptide, vacuole-targeted, hormone conjugates, shoot apical meristem

INTRODUCTION

β -glucosidase is critical for many developmental processes in plants, and the hydrolysis of phyto-hormone conjugates is one of its most important roles (Schliemann, 1984; Sembdner et al., 1994; Kleczkowski et al., 1995). The rolC gene of the bacterial pathogen *Agrobacterium rhizogenes* encodes β -glucosidase, and results in abnormal development when transformed into plants. In particular,

heterologous β -glucosidase can release active forms of phytohormones from their inactive conjugates that consist of glycoside links (Spena et al., 1992; Brzobohaty et al., 1993). Inactive conjugates of each phytohormone can be found abundantly in plant tissues. Their active forms are liberated via β -glucosidase-mediated hydrolysis. Furthermore, many studies have revealed that the inactive forms of phytohormone conjugates act as reversible deactivated storage molecules, and are important for the regulation of physiologically active hormone levels; however, their normal biological functions remain unknown (Staswick, 2009; Piotrowska and Bajguz, 2011).

The vacuole is considered a storage organelle and is an important component of the secretory pathway in plants. Detailed knowledge of the sorting mechanisms, out of and into the vacuole, is lacking (Hall, 2000; Vitale and Hinz, 2005). Previous studies of several lytic enzymes that are specifically targeted to the vacuole (e.g., phaseolin) have revealed some sorting signals (N-terminal or C-terminal polypeptides or internal sequences) that can sort proteins into the vacuole (Frigerio et al., 2001; De Marcos Lousa et al., 2012). Unfortunately, because the internal environment of the vacuole leads to the rapid degradation and hydrolysis of proteins, and other compounds, it has been difficult to determine whether heterologous expressed proteins maintain their functions and features inside the vacuole.

We previously expressed the hyperthermostable β -glucosidase BglB of *T. maritima* in tobacco plants to obtain transgenic plants for application in bioconversion. The optimal temperature and pH of the plant-expressed BglB were 80°C and 4.5, respectively (Jung et al., 2010). Moreover, we also observed some phenotypic modifications, such as longer stems, larger leaves, and shortened developmental stages (Jung et al., 2010, 2013), which we hypothesized may have been due to changes in hormone homeostasis. Therefore, in the present study, we overexpressed heterologous BglB of *T. maritima* in tobacco plants. We targeted the vacuole by insertion of the AFVY tetrapeptide to examine whether BglB maintain its functions of hydrolyzing glycoside bonds to release free hormones from its conjugates, and to determine how such changes in hormones levels may affect the growth and development of transgenic plants. All of the changes in the above-ground or belowground organs in plants can be explained via the development, division, and reproduction of stem cells harbored in the shoot and root apical meristems, which are regulated by the expression of homeodomain genes and hormone levels. For example, in the shoot apical meristem, the transcription factor *WUSCHEL* (*WUS*) can be upregulated via cytokinin, and a group of dividing cells called the quiescent center (QC) is upregulated by indole 3-acetic acid (IAA) in the root apical meristem (Kerk et al., 2000; Overvoorde et al., 2010; Yadav et al., 2010; Zhao et al., 2010).

MATERIALS AND METHODS

Vector Constructions, Plant Transformation, and Molecular Analysis

For cytosol expression, the full-length sequence of the *T. maritima* β -glucosidase *BglB* gene (Jung et al., 2010) was constructed under control of the 35S promoter, and named Cyt-BglB (CB). For vacuole targeting, *BglB* was modified by replacing its stop codon

with nucleotide sequences encoding the AFVY signal tetrapeptide from the vacuolar storage glycoprotein phaseolin (Frigerio et al., 2001), with a stop codon inserted at the end, and named Vac-BglB (VB). According to previous studies, AFVY tetrapeptide signals are sufficient to target a heterologous protein to the vacuole (Frigerio et al., 2001; Lau et al., 2010). The 35S promoter was also used for vacuole targeting of the recombinant variants. These expression cassettes were then sub-cloned into the modified multiple cloning sites of the binary vector pCambia 2300 (Kim et al., 2010), as shown in **Figure 1A**. *Agrobacterium tumefaciens* strain GV3013 was used for transformation of tobacco (*Nicotiana tabacum* L.) via the leaf-disk method (Helmer et al., 1984). Transformed shoots were selected on solid Murashige-Skoog (MS) medium (Murashige and Skoog, 1962) containing 100 μ g/ml kanamycin and 500 μ g/ml cefotaxime. Transgenic tobacco plants were grown in a growth chamber under a 16-/8-h light/dark cycle at 25 \pm 3°C. After the presence of the transgene was confirmed by genomic DNA polymerase chain reaction (PCR), reverse transcription (RT)-PCR, and Western blotting, the T₀ generation of transgenic and wild-type (WT) plants was moved to a greenhouse for development.

Total genomic DNA was isolated from the T₀ generation transgenic plant leaves using genomic DNA extraction buffer [200 mM Tris-HCl, 250 mM NaCl, 25 mM Na₂-EDTA, 0.5% sodium dodecyl sulfate (SDS)]. The concentration of genomic DNA was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA). To confirm the presence of *BglB*, PCR of genomic DNA was performed using two sets of flanking primers: first FP 5'-GTC GCT CAT CAC GAA ACC GT-3' and RP 5'-ACT ACA GAG GAA AAG GTG AA-3' for checking the presence of a 0.7-kb sequence within *BglB* in the CB and VB constructs, and second FP 5'-TAT GCA GGC TCC CAC CCC TT-3' and RP 5'-GTA TAC GAA TGC ACT ACA GA-3' for checking the presence of a 0.4-kb sequence within *BglB* and nucleotides sequences of AFVY tetrapeptide in VB constructs. For RT-PCR, total RNA was extracted from the leaf tissues and cDNA was synthesized using avian myeloblastosis virus (AMV) reverse transcriptase (Promega, USA) with random hexamers, and the RT-PCR of *BglB* was performed using first primers set mentioned above. RT-PCR was also performed to determine the expression level of *N. tabacum* *WUS* (JQ686923.1) after RNA was extracted from the stems of seedlings and cDNA was synthesized, with the specific primers: FP 5'-ATG CAC ATG AGA GGT GTT TG-3' and RP 5'-TTA AGG GGA ATT AGG AGA TC-3'.

BglB proteins were extracted from the T₀ to T₃ generation transgenic tobacco leaves by grinding the leaf material to a powder in liquid nitrogen and then suspending the powder in protein extraction buffer at pH 8.0 (50 mM Tris-HCl, 5 mM Na₂-EDTA, 20 mM Na₂S₂O₅ \times 5H₂O, 100 mM KCl, 5% glycerol, 1% β -mercaptoethanol). Leaf debris was removed by centrifugation at 13,000 \times g for 20 min at 4°C. Total soluble protein (TSP) in the supernatants was measured using the Bradford method (Bradford, 1976). Using transfer buffer (39 mM glycine, 48 mM Tris, 10% SDS, 20% methanol), 10 μ g of protein was electrophoresed on 12% polyacrylamide gels and transferred to polyvinylidene fluoride membranes (Immobilon-P; Millipore). The membrane was blocked by incubation with 5% skimmed milk (Difco, USA)

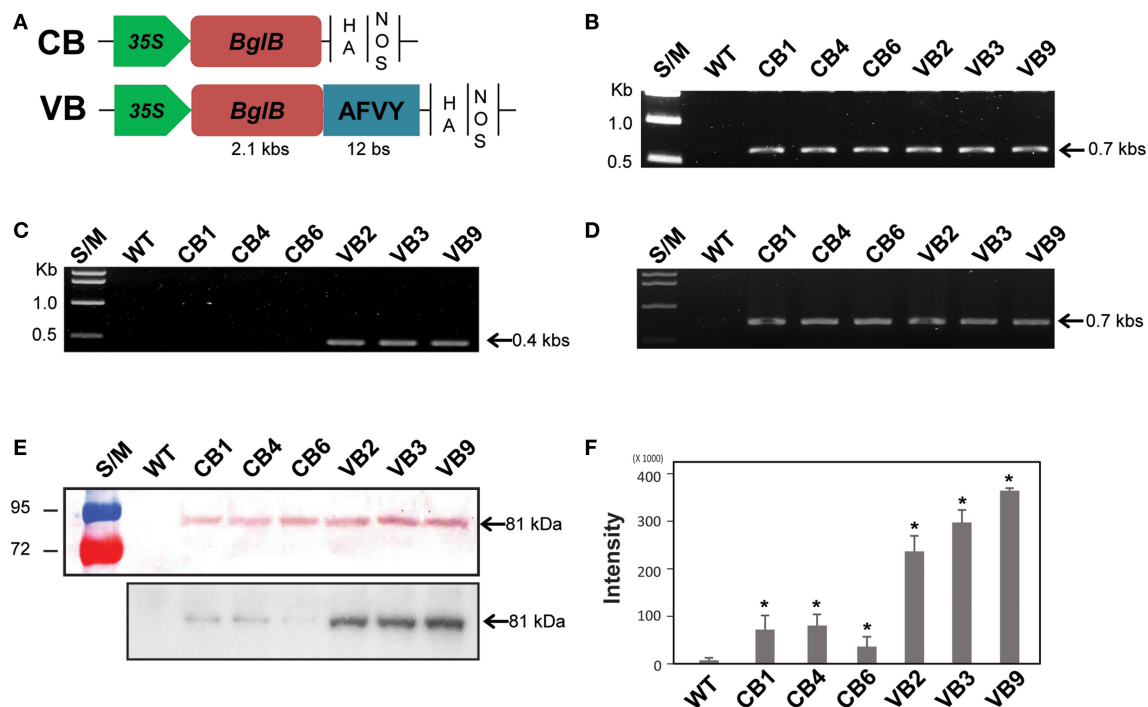


FIGURE 1 | Schematic representation of the BglB expression cassettes and confirmation of heterologously expressed BglB in the transgenic plants. (A) Vector constructions of CB (cytosol-targeted BglB) and VB (vacuole-targeted BglB, with the 12 bp represent AFVY tetrapeptide). (B) Confirmation of the presence of BglB in CB and VB constructs in tobacco genome by PCR genomic DNA. (C) Confirmation of the presence of the sequence encoding AFVY tetrapeptide in the C-terminal of BglB of VB transformants by PCR genomic DNA. (D) Confirmation of the transcript sequence encoding BglB in CB and VB constructs after mRNA extraction, cDNA synthesis, and RT-PCR. (E) Confirmation of the presence of heterologously expressed BglB in the transgenic plants, with BglB present in total soluble protein (TSP) extracted from leaf of CB and VB transgenic plants but no from WT plant (above panel, with 10 µg of TSP was used); and the presence of the heterologous BglB in TSP extracted from isolated vacuole of CB and VB plants but no from those of WT plant (below panel, with 10 µg of TSP from isolated vacuole was used), after concentrated by UFC 710008/Centricon Plus-70 Centrifugal Filter (EM Millipore, USA) by using fluorophore-conjugated secondary antibody anti-rabbit IgG, incubate with substrate within 5 min before transferring to exposure film and keeping within 10 min. (F) The intensity of bands exposed in below panel from (E), measured by Adobe Photoshop CS6. *indicates significant differences from the control (WT) ($P < 0.05$).

in phosphate-buffered saline at pH 7.0 (1 mM KH_2PO_4 , 10 mM $\text{Na}_2\text{HPO}_4 \times 12\text{H}_2\text{O}$, 137 mM NaCl, 2.7 mM KCl), and then incubated with a polyclonal anti- β -glucosidase antibody as the primary antibody. Alkaline phosphatase-conjugated goat anti-rabbit IgG antibody (Promega, USA) was used at a 1:2500 dilution as the secondary antibody. For detecting β -glucosidase BglB targeting vacuole, total protein from isolated vacuole was used to conduct western blot with the same as above, except incubating with a fluorophore-conjugated secondary antibody anti-rabbit IgG (H + L) (DyLight TM 680 Conjugate) (red) in a 1:2500 dilution.

Growth Conditions, Sampling, and Phenotypic Observation

After the presence of the transgene was confirmed, T_0 generation transgenic and WT plants were moved to the greenhouse for development. Seeds from the T_0 generation transgenic plants were sprayed in MS medium containing kanamycin and grown in a growth chamber under a 16-/8-h light/dark cycle at $25 \pm 3^\circ\text{C}$ to produce the T_1 generation, and the germination day was recorded. After 2 weeks, seedlings from ten lines of the CB and VB transgenic plants were used to determine β -glucosidase activity. A

100 mg sample of grinded powder from the leaf was used to extract TSP (Jung et al., 2010), and after checking TSP by the Bradford method, an amount of extracted protein equivalent to 10 µg of TSP were used to examine β -glucosidase enzymatic activity by using p-nitrophenyl b-D-glucopyranoside (pNPG) as the substrate. One unit of β -glucosidase is defined as the amount of enzyme that released 1 mmol of p-nitrophenol from the pNPG substrate under the assay conditions described below. The assay mixture containing 10 mM pNPG in citrate-phosphate buffer (pH 4.5) was incubated with the enzyme for 30 min at 70°C in a total volume of 1 ml. The reaction was stopped by adding 1M Na_2CO_3 , and absorbance was measured at 405 nm. Based on these results, we selected three transgenic lines from each of the CB and VB transgenic plants that showed the highest crude extract β -glucosidase enzymatic activity. Thirty plants from each of the chosen lines were grown in a growth chamber and then moved to 10-l pots in soil-perlite mixtures at $25 \pm 3^\circ\text{C}$ under a 16-/8-h light/dark photoperiod and a light intensity of $100 \text{ mmol m}^{-2} \text{ s}^{-1}$ in a greenhouse for further analysis.

The fifth leaf from the tops of three plants of each of the chosen transgenic lines and W plants was harvested at the same time, and after each 20 days, from 30 to 90 day after germination (DAG),

stored at -70°C and ground in liquid N_2 to analyze β -glucosidase enzymatic activity.

The phenotypic characteristic of the transgenic and WT plants were also recorded from the T_1 to T_3 generation of transgenic plants, including: stem height, number of leaves, root lengths, number of lateral roots, time from germination to initial flowering, and dry weight (the leaves, stems, and roots of non-sample plants were separately harvested and freeze-dried after harvesting the seeds). The carbohydrate content of each part of the plant was also determined using gas chromatography (Coleman et al., 2009).

To conduct a salt stress tolerance experiment in the growth chamber, after germination, the seedlings were transferred to new MS media (without sucrose) containing 200 mM NaCl, and phenotypic characteristics were measured after 15 DAG. Mature (80 DAG) transgenic and WT plants grown in the greenhouse were used for the salt stress experiment, which the plants were watered with the same amount of 200 mM NaCl within 10 days. The weight of spots (including spot, soil, and plants) was measured before and after 10 days NaCl treatment. Simultaneously, the same position on the leaf (10th leaf from the ground to above) was harvested, ground with liquid N_2 , chlorophyll was extracted 90% ethanol, boiled for 5 min and absorbance was measured at an optical density of 620 nm to calculate the chlorophyll concentration (Lichtenthaler and Wellburn, 1983).

Phytohormone Extraction and Measurement

The phytohormones [include abscisic acid (ABA), IAA, and cytokinin] from young leaves or seedlings of CB, VB, and WT plants were extracted with 80% methanol (Oliver et al., 2007) and measured using the Phytodetek competitive enzyme-linked immunosorbent assay (ELISA) kit (Agdia; Elkhardt, IN, USA) the Phytodetek competitive ELISA kits (Agdia). Briefly, young leaves or seedlings of transgenic and WT plants were harvested and ground in N_2 liquid and stored at -70°C for further analysis. One gram of ground powder was mixed with 1 ml 80% methanol and incubated overnight at -4°C . One milliliter (ml) of the supernatant was collected after centrifuging at 13,000 rpm for 10 min to remove debris, and then freeze-dried. The freeze-dried powder was used to measure the levels of each hormone, according to the kit protocols. Each measurement was conducted in triplicate.

Vacuoles Isolation

For purification of vacuole-targeted BglB, transformed protoplasts from young plants (30 DAG) were isolated by hydrolysis with cell-wall hydrolysis enzymes and fractionated by ultracentrifugation according to Mettler et al. (Mettler and Leonard, 1979) and Raikhel et al. (Robert et al., 2007), with some modifications. Due to the requirement of a highly purified of vacuole, the transformed protoplasts were loaded on top of step gradients consisting of 4, 7, 12, and 15% Ficoll, and centrifuged at $97,000 \times g$ for 4 h. The vacuole was isolated in the top layer of the fraction, and then disrupted by sonication before measuring β -glucosidase enzymatic activity with 10 μg TSP.

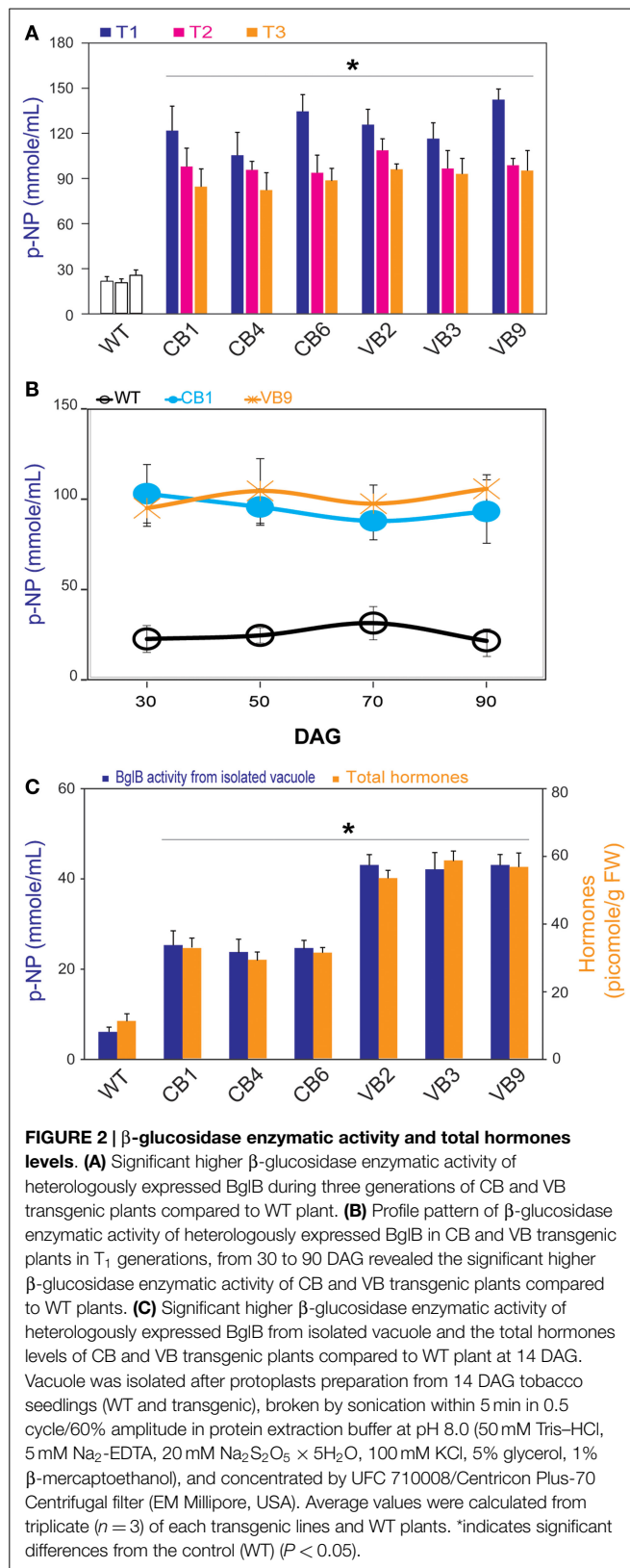
RESULTS

β -glucosidase Enzymatic Activity from Isolated Vacuoles and Total Hormone Levels were Significantly Higher in Transgenic than in WT Plants

Hormone conjugates, which are found in each class of plant hormones, are mainly localized in the vacuoles of plant. The mechanism controlling their transport across membranes and between plant organs remain unknown (Bajguz and Piotrowska, 2009). To analyze the effects of thermostable *T. maritima* BglB on changes in phytohormone metabolism and the consequences for plant development, we built two constructs of BglB. The CB construct was for ectopic expression of BglB in the cytosol, and the VB construct was for expression of vacuole-targeted BglB, under the control of the 35S promoter (Figure 1A). In total, 10 and 12 lines of CB and VB transgenic plants, respectively, were confirmed, and three of the lines were used for further analysis after confirmation of the transgenes by genomic DNA PCR, reverse transcription (RT)-PCR, and Western blotting (Figures 1B–E). The presence of the nucleotide sequences encoding the AFVY tetrapeptide in the VB construct was confirmed using PCR with a reverse primer specific to the VB construct (Figure 1C), and the transcript of the heterologous BglB in the transgenic plants were confirmed by RT-PCR (Figure 1D). The existence of BglB in TSP from the CB and VB plants had molecular weights similar to BglB, as mentioned in the previous study (Jung et al., 2010), were detected by western blot (Figure 1E, above panel), presented no different between CB and VB plants, but showed significantly higher level of BglB in the isolated vacuoles of VB plant compared to CB plant (Figure 1E, below panel), indicated by the higher intensity of the BglB band exposed by the present of BglB in the isolated vacuoles from the VB transgenic than from the CB transgenic and WT plants (Figure 1F). These results obviously indicate that VB plants were the highest vacuole-targeted heterologous BglB.

The three best-performing T_0 generation transgenic lines were selected according to their β -glucosidase enzymatic activity, and then self-pollinated. The β -glucosidase enzymatic activity was significantly higher in the transgenic plants, compared to WT plants over three generations (T_1 to T_3 ; Figure 2A). A slight reduction of β -glucosidase enzymatic activity after a few generations was observed, possibly due to factors such as epigenetic silencing mechanisms (Iyer et al., 2000; Matzke et al., 2000). Heterologous BglB was also stably expressed in the transgenic plants, as indicated by the pattern of β -glucosidase enzymatic activity during plant development from 30 to 90 days after germination (DAG) of the CB1 and VB9 transgenic plants (Figure 2B).

To examine the efficiency of vacuole targeting, we isolated the vacuoles of the WT plants, and the CB and VB transgenic lines, from the T_1 generation. The results showed that the isolated vacuoles of the VB transgenic lines had the highest β -glucosidase enzymatic activity, compared to Cyt-BglB and WT plants (Figure 2C). In particular, compared to WT plants, increased β -glucosidase enzymatic activity of 452 and 759% were recorded in the vacuoles of CB1 and VB2 transformants, respectively. These results accompany to above identification (presented



in **Figure 1E**), indicate that the VB constructs which imposed AFVY tetrapeptide effectively targeted β -glucosidase to the vacuole, and that BglB was still active in the vacuole.

Moreover, significantly higher total hormone (including IAA, ABA, and cytokinin) levels were recorded in the transgenic plants compared to the WT plants, based on the ELISA results, with the highest hormone levels obtained from the VB transformants (**Figure 2C**). In particular, maximum increases of 268 and 463%, when comparing total extracted hormone levels in CB1 and VB3 to WT plants, respectively, were attributed to higher levels of each hormone in the transgenic plants (Figure S1 in Supplementary Material).

Pronounced Phenotypic Changes in the Transgenic Plants

The transgenic CB and VB tobacco plants displayed pronounced phenotypic changes compared to WT plants. Phenotypic characteristics such as stem height, time from germination to initial flowering, and dry weight were proportional to the levels of β -glucosidase enzymatic activity of transgenic and WT plants, suggesting a correlation between the enhancement of β -glucosidase enzymatic activity and these phenotypic changes. In particular, faster development was observed in the transgenic plants than in the WT plants, as indicated by increased stem height, earlier flowering, increased biomass accumulation, and enhanced root system development (**Figure 3A**; Figure S2A in Supplementary Material). Moreover, a shorter time from germination to initial flowering was recorded in the T₁ generation of transgenic plants compared to WT plants. We reported an average of 103.6 and 94.1 DAG in the CB and VB transgenic plants, respectively, compared to 141.7 DAG in WT plants; **Figure 3B**), and similar results were observed for the T₂ and T₃ generations (Figure S2B in Supplementary Material). Higher β -glucosidase enzymatic activity and total hormone levels were recorded at flowering time, with maximum increases in total hormone levels of 222 and 387% for CB1 and VB3 compared to WT plants, respectively (**Figure 3C**; Figure S2C in Supplementary Material), while no significant differences in β -glucosidase enzymatic activity between the CB and VB transgenic plants was observed (**Figure 3C**). These results indicate that more liberated hormones were released in the VB than the CB transgenic plants. After the seeds were harvested, the stem height and dry weight of total biomass accumulation were significantly higher in the mature transgenic plants than in the WT plants, with maximum increases of 133% for stem height (CB1 compared to WT plants) and 124% for total dry weight (VB9 compared to WT plants; **Figures 3D,E**). Similar results were obtained for the T₂ and T₃ generations (Figures S2D,E in Supplementary Material). These results clearly indicated that the increase in liberated hormone levels (particularly IAA and cytokinin) contributed to increased biomass accumulation, despite the shortened growth cycle (earlier flowering after germination) in the transgenic plants. The same phenotypic characteristics were observed in previous studies that targeted β -glucosidase to either general or particular cellular compartments (Jung et al., 2010; Jin et al., 2011). However, despite the significant changes in biomass accumulation and shortened growth cycle, no significant differences in carbohydrate content were observed between the transgenic and WT plants (**Table 1**), indicating that only total biomass accumulation was influenced in the transgenic plants.

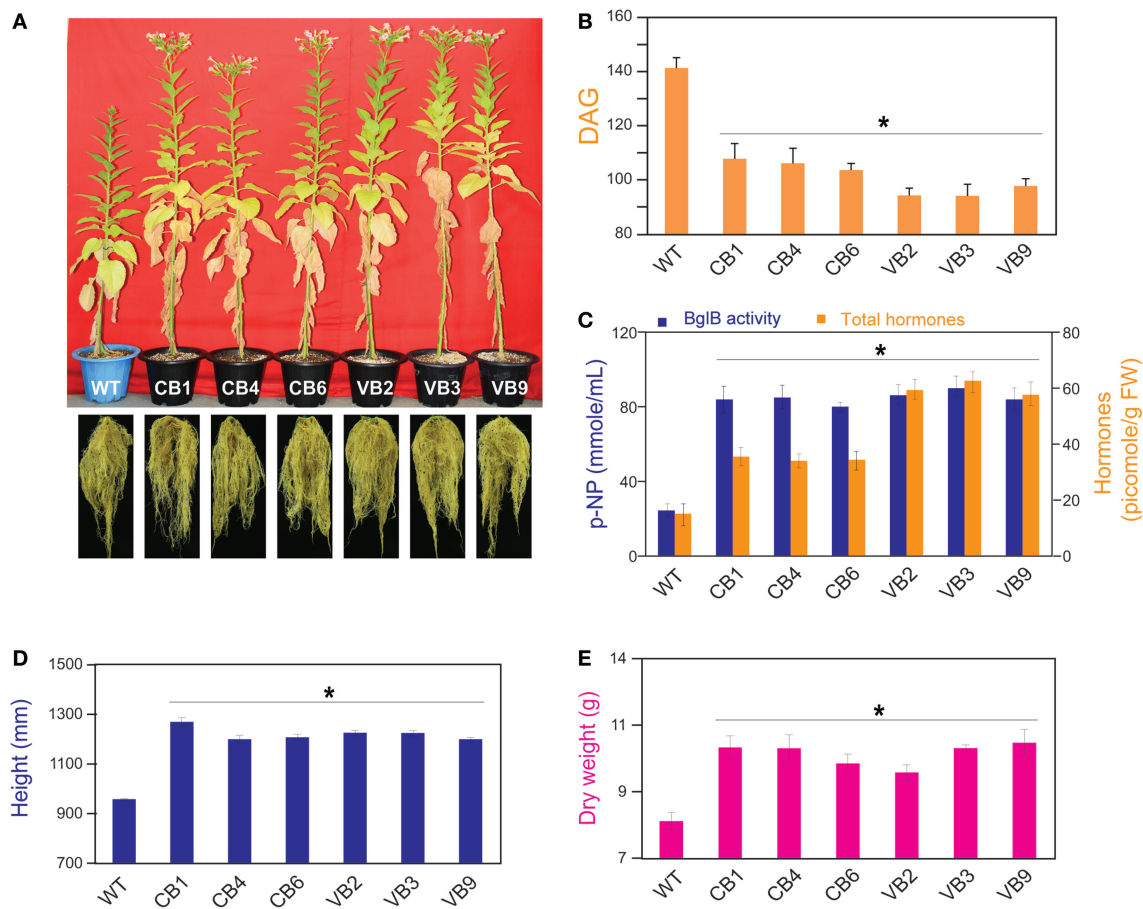


FIGURE 3 | Pronounced phenotypic changes occurred in heterologously expressed BglB transgenic tobacco plants. (A) Faster growth in stem height, shortened growth cycle with earlier flowering, and enhanced root development in the transgenic plants compared to WT plants. **(B)** Time required from germination to flowering of the transgenic vs. WT plants, showed the shortest time required belong to VB plants. **(C)** Combination of β -glucosidase enzymatic activity and total hormones levels of the transgenic and WT plants, measured by harvesting the fifth leaf from the top of plant immediately after flowering. **(D)** Stem height and **(E)** dry weight of the transgenic and WT plants were calculated after harvesting seeds and freeze-dried. Average values were calculated from data recorded from twenty individuals ($n = 20$) of each transgenic lines and WT plants for **(B,D,E)**, and from triplicate ($n = 3$) of each transgenic lines and WT plants for **(C)**. *indicates significant differences from the control (WT) ($P < 0.05$).

Transgenic Plants Showed Faster Development of the Stem and Roots, Elevation of IAA and Cytokinin Levels, and Upregulation of WUS

Based on the increase in stem height and enhanced roots system development, which appeared to be correlated with increased hormone levels of the transgenic plants, we asked whether the increased levels of cytokinin and auxin would affect the development of stems and roots of transgenic seedlings. Seeds from the T₁ to T₃ generations, and the WT plants, were sprayed with MS media containing kanamycin. Immediately after the seeds germinated, tiny seedlings were transferred to new MS media in a line to compare stem development. Faster development of the transgenic plants was clearly observed, as presented by the larger size of the transgenic plants compared to WT plants (Figure 4A). At 15 DAG, along with the increase in β -glucosidase enzymatic activity, IAA and cytokinin levels were significantly higher in

the transgenic plants compared to WT plants, with the highest hormone level obtained in VB transgenic plants (increase in 585% in VB3 compared to WT plants), whereas there was no difference in β -glucosidase enzymatic activity for the CB and VB transgenic plants (Figure 4B). These results indicate that more liberated hormones were released from the vacuole in VB than in CB transgenic plants.

Correlation analysis between the development of the stem (height) and root system (root lengths) to cytokinin and auxin levels showed a maximum increase in 200% for stem height corresponded to an increase in 458% in cytokinin level in VB9 compared to WT plants (Figure 4C). The maximum increases in roots lengths and IAA level were 186 and 725%, respectively, in VB3 compared to WT plants (Figure 4D). Moreover, these results indicate that, despite slight differences in stem height and root lengths between the CB and VB transgenic plants, the increase in liberated IAA and cytokinin levels promoted faster development in the transgenic lines compared to WT plants. We observed

TABLE 1 | Comparison of carbohydrate content in leaves, stem, and roots in WT and vacuole-targeted *T. maritima* BglB transgenic plants.

	(%)	WT	CB1	CB4	CB6	VB2	VB3	VB9
Leaves	Rhamnose	2.3 ± 0.2	2.2 ± 0.3	2.3 ± 0.5	2.6 ± 0.6	2.4 ± 0.3	2.6 ± 0.4	2.6 ± 0.5
	Arabinose	1.4 ± 0.1	1.4 ± 0.2	1.5 ± 0.1	1.5 ± 0.3	1.1 ± 0.1	1.1 ± 0.1	1.1 ± 0.1
	Xylose	1.6 ± 0.2	1.7 ± 0.3	1.6 ± 0.2	1.7 ± 0.0	1.5 ± 0.1	1.6 ± 0.1	1.3 ± 0.0
	Mannose	0.7 ± 0.1	0.9 ± 0.1	1.0 ± 0.2	0.7 ± 0.1	0.7 ± 0.2	0.7 ± 0.3	1.1 ± 0.0
	Galactose	2.1 ± 0.2	2.2 ± 0.4	2.2 ± 0.4	2.2 ± 0.4	2.0 ± 0.3	2.2 ± 0.5	1.9 ± 0.1
	Glucose	45.3 ± 3.5	46.2 ± 4.1	44.6 ± 3.3	44.9 ± 0.8	46.1 ± 3.8	43.9 ± 2.1	45.1 ± 1.9
	Total	53.4 ± 2.8	54.6 ± 3.3	53.2 ± 4.2	53.6 ± 2.6	53.8 ± 4.1	52.1 ± 3.2	53.1 ± 2.5
Stems	Rhamnose	1.8 ± 0.4	1.2 ± 0.1	0.9 ± 0.4	1.2 ± 0.2	1.0 ± 0.2	0.8 ± 0.1	1.2 ± 0.2
	Arabinose	0.7 ± 0.2	0.8 ± 0.0	1.0 ± 0.0	1.1 ± 0.0	1.3 ± 0.0	0.9 ± 0.1	1.1 ± 0.1
	Xylose	5.6 ± 0.7	5.9 ± 1.5	6.2 ± 0.6	7.1 ± 1.1	6.2 ± 0.1	5.8 ± 0.7	5.9 ± 0.9
	Mannose	1.0 ± 0.2	1.3 ± 0.3	1.1 ± 0.3	1.2 ± 0.0	1.1 ± 0.1	1.2 ± 0.1	1.2 ± 0.0
	Galactose	1.2 ± 0.1	1.4 ± 0.2	1.3 ± 0.2	1.2 ± 0.1	1.0 ± 0.1	1.1 ± 0.1	1.0 ± 0.2
	Glucose	44.7 ± 1.2	45.7 ± 6.1	46.2 ± 5.6	47.1 ± 5.4	46.1 ± 2.9	45.5 ± 3.6	43.9 ± 3.2
	Total	55.0 ± 4.2	56.3 ± 6.5	56.7 ± 7.2	58.9 ± 6.5	56.7 ± 3.1	55.3 ± 4.7	54.3 ± 4.8
Roots	Rhamnose	1.3 ± 0.5	1.1 ± 0.0	1.3 ± 0.3	1.0 ± 0.3	1.3 ± 0.1	1.4 ± 0.2	1.3 ± 0.3
	Arabinose	0.8 ± 0.1	0.9 ± 0.1	1.1 ± 0.0	0.8 ± 0.1	1.1 ± 0.1	0.8 ± 0.2	0.9 ± 0.2
	Xylose	6.4 ± 0.5	6.9 ± 1.1	7.1 ± 1.1	6.4 ± 0.3	6.8 ± 0.5	6.6 ± 0.4	6.7 ± 1.9
	Mannose	1.1 ± 0.5	1.3 ± 0.1	1.2 ± 0.0	1.3 ± 0.0	1.2 ± 0.1	1.2 ± 0.0	1.0 ± 0.2
	Galactose	1.2 ± 0.1	1.1 ± 0.3	1.0 ± 0.1	1.1 ± 0.0	1.2 ± 0.2	1.1 ± 0.1	0.8 ± 0.2
	Glucose	46.1 ± 3.8	45.2 ± 2.8	44.8 ± 2.9	44.3 ± 1.5	45.6 ± 3.4	44.8 ± 1.9	46.1 ± 2.8
	Total	56.9 ± 5.1	56.5 ± 3.7	56.5 ± 3.9	54.9 ± 2.4	57.2 ± 4.9	55.9 ± 3.1	56.8 ± 4.8

Average values were calculated from triplicate (n = 3) of each CB and VB transgenic lines and WT plants.

maximum increases in stem height and root lengths of 164 vs. 200%, and 183 vs. 194%, for CB and VB vs. WT plants, respectively. Faster development of the stem and roots was also observed in the T₂ and T₃ generations (Figures S3A,B in Supplementary Material). Furthermore, a higher number of leaves and lateral roots, and greater average fresh weight of 20 young plants, were also observed in the transgenic plants compared to WT plants (Figures S3C–E in Supplementary Material), indicating that the faster development of the transgenic plants, compared to WT plants, was stable after three generations.

Plant stem cells are harbored inside the meristem, which is located in the growing tips of the shoots and roots. The faster development observed in the transgenic plants suggests stronger stimulation of stem cell reproduction, which could then induce changes in plant growth and organogenesis (Murray et al., 2012). The population of stem cells in shoot apical meristems is regulated by expression of the homeodomain gene *WUS*, a transcription factor that can be upregulated by cytokinin level. In the root apical meristem, a group of dividing cells, called the quiescent center (QC) in the root apical meristem, is upregulated by IAA level (Yadav et al., 2010; Zhao et al., 2010). To determine the expression levels of *WUS* for transgenic lines and WT plants, RNA was extracted from the stems of young plants (15 DAG) for cDNA synthesis and RT-PCR. The results showed higher *WUS* expression levels in the transgenic plants compared to WT plants (Figure 4E), providing evidence that superior development of the transgenic plants compared to WT plants was due to elevated hormone levels.

Enhanced Resistance to NaCl Stress and Elevation of ABA in Transgenic Plants

Next, we asked whether increased ABA levels in the transgenic plants led to increased tolerance of salt stress, as mentioned in previous studies (Lee et al., 2006; Wang et al., 2011; Han et al.,

2012; Xu et al., 2012). Seeds from the T₁ to T₃ generations were used, and after germination, tiny seedlings were transferred to new MS media containing 200 mM NaCl to examine the response to high NaCl stress. Observations at 15 DAG showed that the transgenic plants were more resistant to high NaCl, as indicated by enhanced development in the transgenic compared to WT seedlings in term of increased root length, number of leaves, number of lateral roots, and fresh weight (Figure 5A). As shown in Figure 5B, the higher ABA level was clearly related to increased β-glucosidase enzymatic activity in the transgenic seedlings, with the highest ABA levels recorded in the VB transformants (maximum increase in 504% in VB3 compared to WT seedlings). Increased tolerance to high NaCl stress was also displayed by the obviously longer root lengths and greater fresh weight of 100 transgenic seedlings compared to WT seedlings (maximum increase in 271% in root lengths and 256% in fresh weight in the VB2 compared to WT seedlings; Figures 5C,D).

Next, to examine the resistance of mature plants to salt stress, transgenic and WT plants grown in the greenhouse were subjected to a salt stress experiment at 80 DAG. As shown in Figure 5E, more senescent leaves appeared in the transgenic than WT plants, which may explain the higher rate of weight reduction (8.3% in VB2 compared to 3.4% in WT plants; Figure 5F). The appearance of leaves senescence indicated that a faster programmed cell death process occurred in the transgenic than WT plants, which was also represented by the higher rate of chlorophyll degradation in transgenic plants (43.2% in VB3 compared to 14.1% in WT plants; Figure 5G).

DISCUSSION

Because of its thermostability and transglycosylation properties, the *T. maritima* BglB enzyme is considered to be a useful

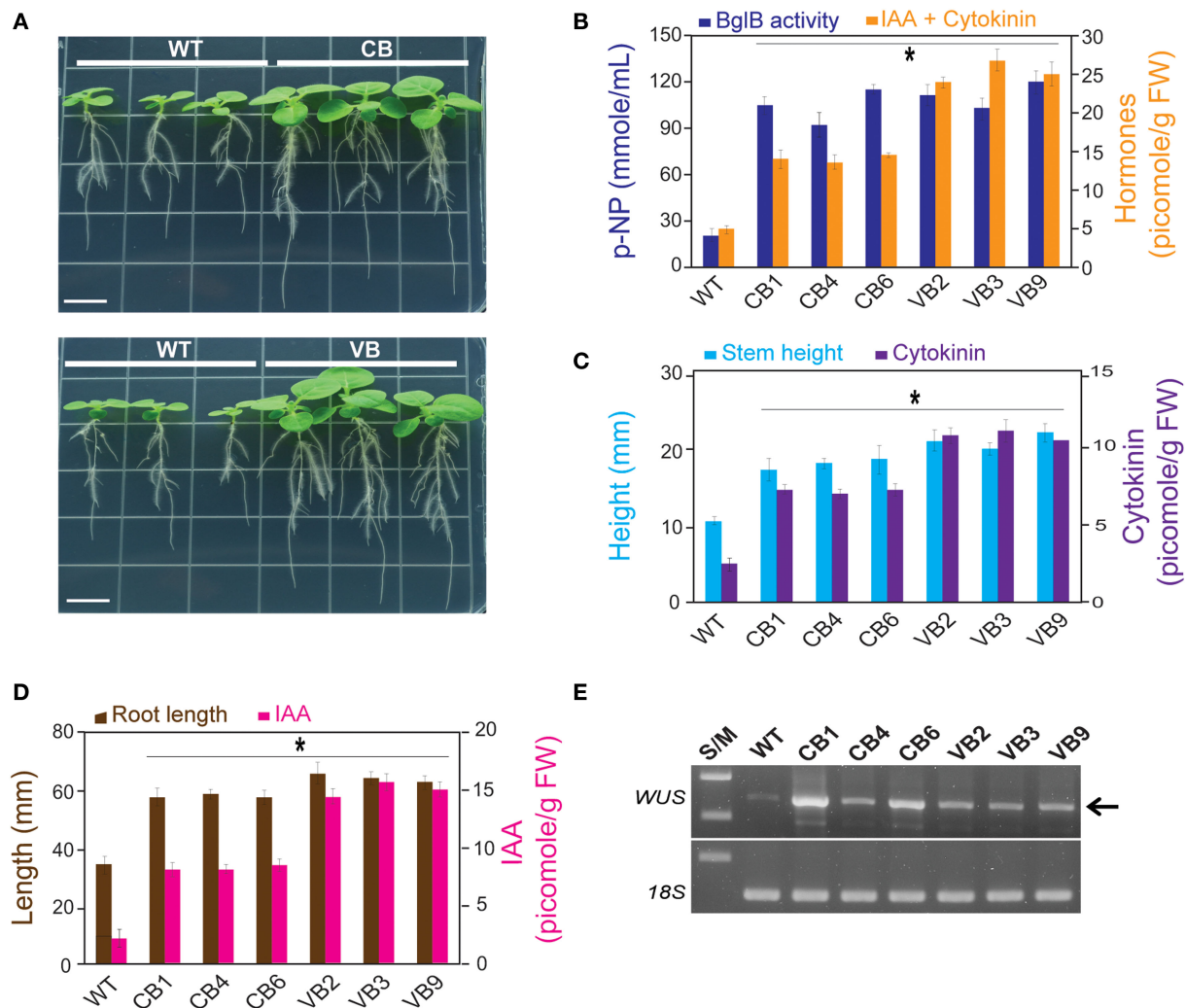
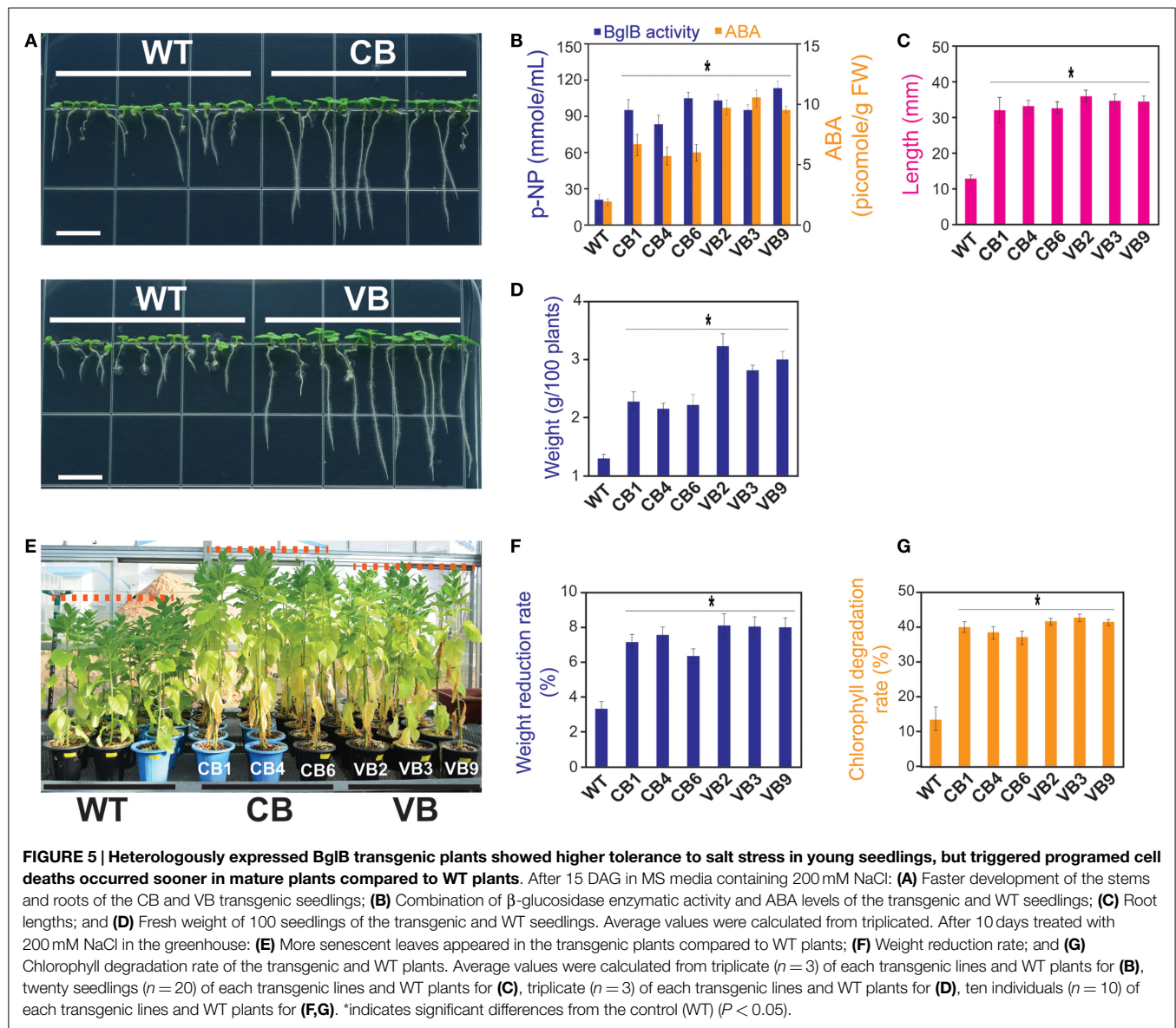


FIGURE 4 | IAA and cytokinin are involved in faster development of the stem and root of the transgenic plants. After 15 DAG: **(A)** faster development of the stem and roots of young CB and VB transgenic young plants compared to WT plants; **(B)** Combination of β -glucosidase enzymatic activity and total levels of IAA and cytokinin of the CB and VB transgenic and WT plants; **(C)** Combination of the stem height and cytokinin levels of the transgenic and WT plants; **(D)** Combination of the root lengths and IAA levels of the transgenic and WT plants; **(E)** Transcript expression levels of *N. tabacum* *WUS* (JQ686923.1) in the transgenic and WT plants. Average values were calculated from triplicate ($n = 3$) of each transgenic lines and WT plants for **(B)**, and twenty individuals ($n = 20$) of each transgenic lines and WT plants for **(C,D)**. *indicates significant differences from the control (WT) ($P < 0.05$).

catalyst for biotechnological applications (Goyal et al., 2001). According to Jung et al. (2010), transgenic tobacco plants can not only be utilized for the mass production of BglB, but also, the overexpression of heterologous BglB in tobacco has led to changes in phenotypic characteristics (such as larger leaves and taller plants) (Jung et al., 2013). Plants contain their own β -glucosidase genes, and previous studies have demonstrated that the expression of β -glucosidase, including heterologous expression, affects the hydrolysis of hormone conjugates and homeostasis in plants, which in turn control plant development (Schliepmann, 1984; Brzobohaty et al., 1993; Dietz et al., 2000; Kiran et al., 2006). In the present study, by observing pronounced phenotypic changes in the *T. maritima* BglB transgenic tobacco compared to WT plants. The transgenic tobacco remained stable

over three offspring generations (Figures 3–5). We were encouraged to evaluate the relationship between β -glucosidase enzymatic activity of *T. maritima* BglB and changes in plant hormone levels.

For vacuole targeting, among the three different types of vacuolar sorting signals (N- or C-terminal polypeptides or internal sequences) that have been identified (Jiang and Rogers, 1998; Matsuoka and Neuhaus, 1999), C-terminal polypeptides, such as the C-terminal amino acids AFVY tetrapeptide from phaseolin, are considered be the most efficient (Frigerio et al., 2001; Nausch et al., 2012a,b). However, due to the presence of numerous hydrolytic enzymes in the vacuole of plant cells, it is generally difficult for proteins to maintain their activity inside the vacuole (Boller and Kende, 1979; Marty, 1999). Here, we showed that



the β -glucosidase enzymatic activity of heterologously expressed BglB was significantly higher in the transgenic (both the CB and VB transformants) plant compared to WT plants (Figure 2A). The transgenic plants remained stable during the life cycle and durable after three offspring generation (Figures 2A,B). These results indicate that the *T. maritima* BglB was effectively expressed in the transgenic tobacco plants.

For the first time, the present of the heterologous BglB and β -glucosidase enzymatic activity assays were conducted after vacuole isolation, which showed that BglB expression was dramatically higher in the VB than CB transgenic plants (Figures 1E and 2C). This result clearly indicated that AFVY tetrapeptide were effective for sorting *T. maritima* BglB into the vacuole, and that its β -glucosidase enzymatic activity was maintained and could tolerate the protein-degrading conditions of the vacuole environment. Therefore, vacuole-targeted *T. maritima* BglB transgenic

plants should be considered candidates for plant molecular farming, where plants are used as bioreactors to produce degrading enzymes for hydrolysis of lignocellulosic material, which is similar to chloroplast-targeted *T. maritima* BglB transgenic plants (Jung et al., 2010, 2013).

Hormone glucoside conjugates, which are mainly stored in the plant vacuole, are considered inactive forms in hormone metabolism, and can be liberated by β -glucosidases, a large group of enzymes that can hydrolyze glucoside ester linkages (Sembdner et al., 1994; Bajguz and Piotrowska, 2009). A wide variety of β -glucosidase enzymes from plants have been proven to be hormone conjugates with hydrolysis capability (Schliemann, 1984; Brzobohaty et al., 1993; Dietz et al., 2000; Kiran et al., 2006; Lee et al., 2006; Yao et al., 2007; Jin et al., 2011). We demonstrated a novel approach in which transformation of BglB, encoding a thermostable β -glucosidase from the bacterium *T. maritima* (Goyal

et al., 2001), affected plant hormone levels through hydrolyzation of glucoside ester links in hormone conjugates in the transgenic plants, which seemed to be the result of non-specific activity. For example, previous studies demonstrated that each kind of β -glucosidase likely performs its functions in specific hormone conjugates (Brzobohaty et al., 1993; Dietz et al., 2000). Kiran et al. (2012) reported that Zm-p60.1 is capable of releasing active cytokinin from O- and N-glucosides, and confirmed that the liberated hormones are still in the active state. Knowledge of the transportation mechanism from inside to outside of the vacuole is still lacking (Vitale and Hinz, 2005; De Marcos Lousa et al., 2012). In the present study, significantly higher enzymatic activity, particularly in isolated vacuoles, was accompanied by dramatically higher levels of hormones (IAA, ABA, and cytokinin) in the VB plants compared to CB transgenic plants, with WT plants showing the lowest levels (Figure 2C). These results clearly demonstrated that, when greater amounts of BglB were targeted to the vacuole, more liberated hormones were released.

In contrast to the results obtained by Kiran et al. (2012), who found no significant phenotypic changes in vacuole-targeted Zm-p60.VAL transgenic plants, our results showed pronounced phenotypic changes in *T. maritima* BglB transgenic plants compared to WT plants. Mature transgenic plants exhibited enhanced development, in terms of faster growth in stem height and a shortened growth cycle, with earlier flowering (Figure 3). Young seedlings had increased stem height and longer roots (Figure 4), which were accompanied by significantly higher hormones levels that were maintained over three offspring generations of the transgenic plants. These results provide clear evidence that heterologously expressed BglB increases the plant hormones levels, which then influence their phenotypes.

Due to the elevated levels of IAA, ABA, and cytokinin, it is difficult to determine the specific factor that directly contributes to the phenotypic changes in the transgenic plants. Fortunately, previous works can provide clues to trace the cause of such changes. For example, IAA is known to regulate root development (Overvoorde et al., 2010), cytokinin plays pivotal roles in the formation and activity of shoot meristems (Werner et al., 2003; Werner and Schmülling, 2009), and ABA functions in the plant response to dehydrating/salinity stresses and programmed cell death (Finkelstein, 2006; Yang et al., 2014). Previous studies have also shown that the reproduction and differentiation of stem cells harbored in the shoot and root apical meristem contribute to development and organogenesis in plants (Williams and Fletcher, 2005; Powell and Lenhard, 2012). Therefore, the taller stem height, longer roots, and earlier flowering observed in the transgenic plants could indicate enhancement of the shoot and root apical meristem in the transgenic plants compared to WT plants. Specifically, the expression level of WUS, a transcription factor that regulates the development and division of stem cells in the shoot apical meristem, is upregulated by cytokinin (Kurakawa et al., 2007; Werner and Schmülling, 2009; Zhao et al., 2010), shedding light on the mechanism contributing to the role of cytokinin, which was increased in our transgenic plants, in enhancing the development of the stems and aboveground organs.

Our result showed enhanced development of the root systems (represented by increased roots dry weight, number of lateral roots, and root length), confirming the effect of a larger amount of IAA on the development of root systems in the transgenic plants (Figures 3 and 4). ABA mainly functions in the plant's response to dehydration by inducing stomatal opening/closing, and also plays a role in limiting cell division and expansion, decreasing shoot growth and lateral root initiation, and promoting developmental phase changes such as vegetative-to-reproductive transitions (Finkelstein, 2006). In the present study, the increased ABA levels were related to increased salt stress tolerance in young seedlings. The faster chlorophyll degradation and higher rates of weight reduction after treatment with NaCl solution in mature plants revealed that programmed cell death was promptly triggered in the transgenic plants for both the VB and CB transformants, compared to WT plants (Figure 5). Notably, no significant difference in β -glucosidase enzymatic activity, but significantly higher hormones levels, in the VB transgenic plants compared to CB transgenic plants, were observed, confirming that the hormone conjugates are mainly stored in the vacuole, and more liberated hormones were released from the conjugates in the VB transgenic plants, which contributed to the greater effect on plant development in the VB transgenic plants.

CONCLUSION

After *T. maritima* BglB was first overexpressed and effectively targeted into the vacuole by the addition of AFVY C-terminal tetrapeptides, BglB was still active and functional. The main results emerging from this study are that the hormone (ABA, IAA, and cytokinin) conjugates are mainly stored in the vacuole, and perhaps more importantly, higher levels of hormones liberated from their conjugates via BglB-mediated hydrolysis enhance the growth and development in VB transgenic plants to a greater extent than in CB transgenic plants. Therefore, the use of heterologously overexpressed vacuole-targeted *T. maritima* BglB may be an approach to develop molecular farming technology to achieve multiple targets: increased production of the β -glucosidase BglB, increased biomass accumulation, and shortened of developmental stages. Also this BglB vacuole-targeted plant farming system influences of total biomass accumulation and as such may be useful in increasing biomass production for bioenergy and biofuel production.

FUNDING

This work was supported by Priority Centers Program (2010-0020141) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and technology, and by a grant (S211314L010120) from Forest Science & Technology Projects, Forest Service, Republic of Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00181>

REFERENCES

- Bajguz, A., and Piotrowska, A. (2009). Conjugates of auxin and cytokinin. *Phytochemistry* 70, 957–969. doi:10.1016/j.phytochem.2009.05.006
- Boller, T., and Kende, H. (1979). Hydrolytic enzymes in the central vacuole of plant cells. *Plant Physiol.* 63, 1123–1132. doi:10.1104/pp.63.6.1123
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi:10.1016/0003-2697(76)90527-3
- Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J., et al. (1993). Release of active cytokinin by a beta-glucosidase localized to the maize root meristem. *Science* 262, 1051–1054. doi:10.1126/science.8235622
- Coleman, H. D., Yan, J., and Mansfield, S. D. (2009). Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultra-structure. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13118–13123. doi:10.1073/pnas.0900188106
- De Marcos Lousa, C., Gershlick, D. C., and Denecke, J. (2012). Mechanisms and concepts paving the way towards a complete transport cycle of plant vacuolar sorting receptors. *Plant Cell* 24, 1714–1732. doi:10.1105/tpc.112.095679
- Dietz, K. J., Sauter, A., Wichert, K., Messdaghi, D., and Hartung, W. (2000). Extracellular β -glucosidase activity in barley involved in the hydrolysis of ABA glucose conjugate in leaves. *J. Exp. Bot.* 51, 937–944. doi:10.1093/jxb/51.346.937
- Finkelstein, R. R. (2006). Studies of abscisic acid perception finally flower. *Plant Cell* 18, 786–791. doi:10.1105/tpc.106.041129
- Frigerio, L., Foresti, O., Felipe, D. H., Neuhaus, J.-M., and Vitale, A. (2001). The C-terminal tetrapeptide of phaseolin is sufficient to target green fluorescent protein to the vacuole. *J. Plant Physiol.* 158, 499–503. doi:10.1078/0176-1617-00362
- Goyal, K., Selvakumar, P., and Hayashi, K. (2001). Characterization of a thermostable β -glucosidase (BglB) from *Thermotoga maritima* showing transglycosylation activity. *J. Mol. Catal. B Enzym.* 15, 45–53. doi:10.1016/S1381-1177(01)00003-0
- Hall, J. L. (2000). Deepesh N. De. Plant cell vacuoles: an introduction. CSIRO Publishing, Collingwood, 2000. Pp. 288. Price US\$ 60.00. ISBN 0643 062548. *J. Exp. Bot.* 51, 2127. doi:10.1093/jxb/51.353.2127
- Han, Y.-J., Cho, K.-C., Hwang, O.-J., Choi, Y.-S., Shin, A.-Y., Hwang, I., et al. (2012). Overexpression of an *Arabidopsis* β -glucosidase gene enhances drought resistance with dwarf phenotype in creeping bentgrass. *Plant Cell Rep.* 31, 1677–1686. doi:10.1007/s00299-012-1280-6
- Helmer, G., Casadaban, M., Bevan, M., Kayes, L., and Chilton, M.-D. (1984). A new chimeric gene as a marker for plant transformation: the expression of *Escherichia coli* [beta]-galactosidase in sunflower and tobacco cells. *Nat. Biotech.* 2, 520–527. doi:10.1038/nbt0684-520
- Iyer, L., Kumpatla, S., Chandrasekharan, M., and Hall, T. (2000). Transgene silencing in monocots. *Plant Mol. Biol.* 43, 323–346. doi:10.1023/A:1006412318311
- Jiang, L., and Rogers, J. C. (1998). Integral membrane protein sorting to vacuoles in plant cells: evidence for two pathways. *J. Cell Biol.* 143, 1183–1199. doi:10.1083/jcb.143.5.1183
- Jin, S., Kanagaraj, A., Verma, D., Lange, T., and Daniell, H. (2011). Release of hormones from conjugates: chloroplast expression of beta-glucosidase results in elevated phytohormone levels associated with significant increase in biomass and protection from aphids or whiteflies conferred by sucrose esters. *Plant Physiol.* 155, 222–235. doi:10.1104/pp.110.160754
- Jung, S., Kim, S., Bae, H., Lim, H.-S., and Bae, H.-J. (2010). Expression of thermostable bacterial β -glucosidase (BglB) in transgenic tobacco plants. *Bioresour. Technol.* 101, 7144–7150. doi:10.1016/j.biortech.2010.03.140
- Jung, S., Lee, D.-S., Kim, Y.-O., Joshi, C., and Bae, H.-J. (2013). Improved recombinant cellulase expression in chloroplast of tobacco through promoter engineering and 5' amplification promoting sequence. *Plant Mol. Biol.* 83, 317–328. doi:10.1007/s11103-013-0088-2
- Kerk, N. M., Jiang, K., and Feldman, L. J. (2000). Auxin metabolism in the root apical meristem. *Plant Physiol.* 122, 925–932. doi:10.1104/pp.122.3.925
- Kim, S., Lee, D.-S., Choi, I., Ahn, S.-J., Kim, Y.-H., and Bae, H.-J. (2010). *Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. *Transgenic Res.* 19, 489–497. doi:10.1007/s11248-009-9330-8
- Kiran, N. S., Benkova, E., Reková, A., Dubová, J., Malbeck, J., Palme, K., et al. (2012). Retargeting a maize beta-glucosidase to the vacuole – evidence from intact plants that zeatin-O-glucoside is stored in the vacuole. *Phytochemistry* 79, 67–77. doi:10.1016/j.phytochem.2012.03.012
- Kiran, N. S., Polanska, L., Fohlerova, R., Mazura, P., Valkova, M., Smeral, M., et al. (2006). Ectopic over-expression of the maize beta-glucosidase Zm-p60.1 perturbs cytokinin homeostasis in transgenic tobacco. *J. Exp. Bot.* 57, 985–996. doi:10.1093/jxb/erj084
- Kleczkowski, K., Schell, J., and Bandur, R. (1995). Phytohormone conjugates: nature and function. *CRC Crit. Rev. Plant Sci.* 14, 283–298. doi:10.1080/07352689509382361
- Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., et al. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445, 652–655. doi:10.1038/nature05504
- Lau, O. S., Ng, D. W., Chan, W. W., Chang, S. P., and Sun, S. S. (2010). Production of the 42-kDa fragment of *Plasmodium falciparum* merozoite surface protein 1, a leading malaria vaccine antigen, in *Arabidopsis thaliana* seeds. *Plant Biotechnol. J.* 8, 994–1004. doi:10.1111/j.1467-7652.2010.00526.x
- Lee, K. H., Piao, H. L., Kim, H.-Y., Choi, S. M., Jiang, F., Hartung, W., et al. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell* 126, 1109–1120. doi:10.1016/j.cell.2006.07.034
- Lichtenthaler, H., and Wellburn, A. (1983). Determination of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. *Biochem. Soc. Trans.* 11, 591–592. doi:10.1042/bst0110591
- Marty, F. (1999). Plant vacuoles. *Plant Cell* 11, 587–599. doi:10.1105/tpc.11.4.587
- Matsuoka, K., and Neuhaus, J.-M. (1999). Cis-elements of protein transport to the plant vacuoles. *J. Exp. Bot.* 50, 165–174. doi:10.1093/jxb/50.331.165
- Matzke, M. A., Mette, M. F., and Matzke, A. J. M. (2000). Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* 43, 401–415. doi:10.1023/A:1006484806925
- Mettler, I. J., and Leonard, R. T. (1979). Isolation and partial characterization of vacuoles from tobacco protoplasts. *Plant Physiol.* 64, 1114–1120. doi:10.1104/pp.64.6.1114
- Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi:10.1111/j.1399-3054.1962.tb08052.x
- Murray, J. A. H., Jones, A., Godin, C., and Traas, J. (2012). Systems analysis of shoot apical meristem growth and development: integrating hormonal and mechanical signaling. *Plant Cell* 24, 3907–3919. doi:10.1105/tpc.112.102194
- Nausch, H., Mikschofsky, H., Koslowski, R., Meyer, U., Broer, I., and Huckauf, J. (2012a). High-level transient expression of ER-targeted human interleukin 6 in *Nicotiana benthamiana*. *PLoS ONE* 7:e48938. doi:10.1371/journal.pone.0048938
- Nausch, H., Mikschofsky, H., Koslowski, R., Meyer, U., Broer, I., and Huckauf, J. (2012b). Expression and subcellular targeting of human complement factor C5a in *Nicotiana species*. *PLoS ONE* 7:e53023. doi:10.1371/journal.pone.0053023
- Oliver, S. N., Dennis, E. S., and Dolferus, R. (2007). ABA regulates apoplastic sugar transport and is a potential signal for cold-induced pollen sterility in rice. *Plant Cell Physiol.* 48, 1319–1330. doi:10.1093/pcp/pcm100
- Overvoorde, P., Fukaki, H., and Beeckman, T. (2010). Auxin control of root development. *Cold Spring Harb. Perspect. Biol.* 2, a001537. doi:10.1101/cshperspect.a001537
- Piotrowska, A., and Bajguz, A. (2011). Conjugates of abscisic acid, brassinosteroids, ethylene, gibberellins, and jasmonates. *Phytochemistry* 72, 2097–2112. doi:10.1016/j.phytochem.2011.08.012
- Powell, A. E., and Lenhard, M. (2012). Control of organ size in plants. *Curr. Biol.* 22, R360–R367. doi:10.1016/j.cub.2012.02.010
- Robert, S., Zouhar, J., Carter, C., and Raikhel, N. (2007). Isolation of intact vacuoles from *Arabidopsis rosette* leaf-derived protoplasts. *Nat. Protoc.* 2, 259–262. doi:10.1038/nprot.2007.26
- Schliemann, W. (1984). Hydrolysis of conjugated gibberellins by β -glucosidases from dwarf rice (*Oryza sativa* L. cv. ‘Tan-ginbozu’). *J. Plant Physiol.* 116, 123–132. doi:10.1016/S0176-1617(84)80069-3
- Sembdner, G., Atzorn, R., and Schneider, G. (1994). Plant hormone conjugation. *Plant Mol. Biol.* 26, 1459–1481. doi:10.1007/BF00016485
- Spena, A., Estruch, J. J., Prinsen, E., Nacken, W., Van Onckelen, H., and Sommer, H. (1992). Anther-specific expression of the rolB gene of *Agrobacterium rhizogenes* increases IAA content in anthers and alters anther development and whole flower growth. *Theor. Appl. Genet.* 84, 520–527. doi:10.1007/BF00224147
- Staswick, P. (2009). Plant hormone conjugation: a signal decision. *Plant Signal. Behav.* 4, 757–759. doi:10.1104/pp.109.138529
- Vitale, A., and Hinz, G. (2005). Sorting of proteins to storage vacuoles: how many mechanisms? *Trends Plant Sci.* 10, 316–323. doi:10.1016/j.tplants.2005.05.001

- Wang, P., Liu, H., Hua, H., Wang, L., and Song, C.-P. (2011). A vacuole localized β -glucosidase contributes to drought tolerance in *Arabidopsis*. *Chin. Sci. Bull.* 56, 3538–3546. doi:10.1007/s11434-011-4802-7
- Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H., and Schmülling, T. (2003). Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* 15, 2532–2550. doi: 10.1105/tpc.014928
- Werner, T., and Schmülling, T. (2009). Cytokinin action in plant development. *Curr. Opin. Plant Biol.* 12, 527–538. doi:10.1016/j.pbi.2009.07.002
- Williams, L., and Fletcher, J. C. (2005). Stem cell regulation in the *Arabidopsis* shoot apical meristem. *Curr. Opin. Plant Biol.* 8, 582–586. doi:10.1016/j.pbi.2005.09.010
- Xu, Z. Y., Lee, K. H., Dong, T., Jeong, J. C., Jin, J. B., Kanno, Y., et al. (2012). A vacuolar beta-glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in *Arabidopsis*. *Plant Cell* 24, 2184–2199. doi:10.1105/tpc.112.095935
- Yadav, R. K., Tavakkoli, M., and Reddy, G. V. (2010). WUSCHEL mediates stem cell homeostasis by regulating stem cell number and patterns of cell division and differentiation of stem cell progenitors. *Development* 137, 3581–3589. doi: 10.1242/dev.054973
- Yang, J., Worley, E., and Udvardi, M. (2014). A NAP-AAO3 regulatory module promotes chlorophyll degradation via ABA biosynthesis in *Arabidopsis* leaves. *Plant Cell* 26, 4862–4874. doi:10.1105/tpc.114.133769
- Yao, J., Huot, B., Foune, C., Doddapaneni, H., and Enyedi, A. (2007). Expression of a β -glucosidase gene results in increased accumulation of salicylic acid in transgenic *Nicotiana tabacum* cv. Xanthi-nc NN genotype. *Plant Cell Rep.* 26, 291–301. doi:10.1007/s00299-006-0212-8
- Zhao, Z., Andersen, S. U., Ljung, K., Dolezal, K., Miotk, A., Schultheiss, S. J., et al. (2010). Hormonal control of the shoot stem-cell niche. *Nature* 465, 1089–1092. doi:10.1038/nature09126

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Nguyen, Lee, Jung and Bae. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Book review: Socio-economic impacts of bioenergy production

Sheikh Adil Edrisi and P. C. Abhilash*

Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India

Keywords: biomass and biofuel, socio-economic factors, sustainable development, waste land development, job opportunities, income effect

A book review on Socio-economic impacts of bioenergy production

by Rutz, D., and Janssen, R. (eds). (2014). Dordrecht: Springer, 310 pp. ISBN 978-3-319-03828-5

During the last few decades, the bioenergy production has been radically promoted worldwide (EIA, 2013) as a clean and eco-friendly source of energy mainly due to the global concern of climate change, primarily attributed by the increasing emission of CO₂ by fossil fuel consumptions. Although vegetable oils was first used as liquid fuels (bioenergy) in internal combustion engine by Rudolf Diesel in early 1900 (Pousa et al., 2007), the low cost and easy availability have made the fossil fuels, such as petroleum, as a primary fuel for vehicular transports. Nevertheless, the limited reserves of petroleum and its derivative had necessitated the search for cleaner and alternative source of energy (Parente, 2003; Pousa et al., 2007). Although bioenergy is considered as a versatile type of renewable energy, the large scale production of bioenergy from biomass is always under criticism as it requires large tracks of arable lands for bioenergy plantation. Hence, there is a direct conflict of interest between the food and biofuel production in the concerned societies. As a result, the present global bioenergy production is not promising and (55 EJ) (World Bioenergy Association, 2014) is still not enough to satisfy the energy demands of burgeoning global population. Hence, it is the need of the hour to maximize the bioenergy production without having any social and economic conflicts. However, the socio-economic considerations are mainly influenced by the local and regional frameworks (societal and economic) of the concerned nations, including the educational level, cultural aspects, and policies, including environmental and social targets. In this context, the book “*Socio-economic impacts of bioenergy production*” is a relevant and in-depth analysis of the socio-economic issues associated with bioenergy production and provides suitable indicators for assessing the sustainability of such bioenergy production programs and also providing strategies for overcoming all such negative impacts in an amicable way. As editors rightly pointed out, the book “*illustrates the complexity of interrelated topics in the bioenergy value chain, ranging from agriculture to conversion processes, as well as from social implications to environmental effects. It furthermore gives an outlook on future challenges associated with the expected boom of a global bio-based economy, which contributes to the paradigm shift from a fossil based to a biomass and renewable energy based economy.*” Therefore, the book is targeted to a wider readers ranging from “policy makers, scientists, and NGOs in the fields of agriculture, forestry, biotechnology, and energy.”

Though there are few books in the same domain, such as “*Sustainable Bioenergy Production: An Integrated Approach*” (Ruppert et al., 2013) and “*Bioenergy for Sustainable Development in Africa*” (Janssen and Rutz, 2012), addressing one or another aspects of socio-economic implications of bioenergy production, the current book provides a complete deliberations on the above issues along with solutions in a doable way. Therefore, we could appropriately say that the current book has its own perspectives, ideas, and deliberations that handle the cross cutting edges of socio-economic

OPEN ACCESS

Edited by:

Robert Henry,
The University of Queensland,
Australia

Reviewed by:

Abu Yousuf,
Universiti Malaysia Pahang, Malaysia

*Correspondence:

P. C. Abhilash
pca.iesd@bhu.ac.in,
pcabhilash@hotmail.com

Specialty section:

This article was submitted to
Bioenergy and Biofuels, a section of
the journal *Frontiers in Bioengineering
and Biotechnology*

Received: 13 September 2015

Accepted: 12 October 2015

Published: 26 October 2015

Citation:

Edrisi SA and Abhilash PC (2015)
Book review: Socio-economic
impacts of bioenergy production.
Front. Bioeng. Biotechnol. 3:174.
doi: 10.3389/fbioe.2015.00174

impacts of bioenergy production. Importantly, “*this publication builds upon the results of the Global-Bio-Pact project on ‘Global Assessment of Biomass and Bio-product Impacts on Socio-economics and Sustainability’ which was supported by the European Commission in the 7th Framework Programme for Research and Technological Development from February 2012 to January 2013.*” Moreover, the “*contributions to this book are based on the experience of selected authors from Europe, Africa, Asia, and Latin America, including researchers, investors, policy makers and other stakeholders such as representatives from NGOs.*”

The book basically focuses on the following aspects, such as (i) various tools for socio-economic impact assessments, (ii) indicators for assessing socio-economic sustainability, (iii) test auditing of the indicators, (iv) linkages between socio-economic and environmental impacts of bioenergy, (v) socio-economic impact of biofuels on land use change, (vi) effects on food security, (vii) socio-economic impacts of sweet sorghum value chains in temperate and tropical regions, (viii) the use of soybean by-products as a biofuel in Argentina, (ix) socio-economic impacts of palm oil and biodiesel products in Indonesia, (x) socio-economic experiences of Jatropha production in Africa and Mali, (xi) socio-economic impacts of bioethanol from sugarcane in Brazil and Costa Rica, (xii) socio-economic analysis of lignocelluloses ethanol refinery in Canada, (xiii) biogas production from organic waste in Africa, (xiv) socio-economic indicators on different bioenergy case studies, and (xv) the contribution of bioenergy to energy access and energy security.

The editors presented the intended themes in an organized and progressive manner via its arrayed, interrelated, and well-structured chapters. Furthermore, it reveals that the socio-economic impact assessment basically consists of scoping and determination of issues, social and economic baselines, its impacts, significance, mitigation, management, and monitoring. Moreover, it also describes that these assessments can also be “*used as an add-on to environmental impact assessment and/or to support biomass certification schemes.*” The introductory chapter delivers a conceptual framework regarding the different tools used for the assessment of socio-economic impacts with glorious glimpses of sustainability concepts.

The book also highlights various issues related to the impacts of bioenergy production, its assessment and screening strategies, including key indicators for test auditing and also provides linkages between the socio-economic and environmental impacts of bioenergy production and utilization. Editors have made a special attempt to address the further implications of bioenergy production on land use change and food security, but fails to explore the issues under the current global land use scenarios as the concerned chapter is only having the basic definitions of several land use pattern. However, there are some interesting sections, such as “*land use rights, land tenure, and ownership,*” which are useful for

owing the ownership right to various stakeholders. Moreover, the impact analysis on food security is a well-handled issue corroborated with connections and controversies of food and fuel production, envisioning methodology for an economy-wide assessment of food security and biofuels, and also quantified the different biofuel policies’ impact on food security. The chapter on the socio-economic impacts of sweet sorghum is well-supported through value-chain analysis ranging from its cultivation to conversion scenarios in tropical and temperate production systems. The book also deals with the bioenergy production from soybean, palm, and sugarcane in Argentina, Indonesia, and Brazil, respectively, in different individual chapters and also emphasized Jatropha individually with its different business models in tropical areas specifically in Africa and Mali that will certainly attracts scientists, policy makers, entrepreneurs, stakeholders, and NGOs concerned with agroforestry for sustainable bioenergy production.

Apart from the issues related with the socio-economic perspectives of biomass conversion for bioenergy production, the book also addresses various issues related to the value-chain analysis of biomass conversion technologies with special reference to sugarcane to ethanol production in Costa Rica and also separately illustrates the impacts of a refinery in Canada targeting lignocellulosic ethanol production. It clearly elucidates the socio-economics of lignocellulosic biomass supply chain in a national context with special emphasis on the forestry sector and land ownership. The section also having regional and local case studies of British Columbia that deals with the lignol technology, supply chain, products of the lignol process, pyrolysis of biomass technology group (BTG), products of the BTG process, macroeconomics in the lignocellulosic biomass chain in Canada and British Columbia, employment generation, working conditions, relevance of impacts, threshold determination, mitigation options and biomass certifications, etc.

Although the use of genetically modified products of sweet sorghum and other plant products were discussed superficially, it would be much better if separate sections or chapters with specific case studies at national, regional, or local levels. The volume also lacks the future challenges and recommendations in the lignocellulosic biomass conversion processes. Moreover, the qualities of the display items (figures and illustrations) are not good and difficult to understand. However, our overall impression is that the book is a well written “*guideline to assess the socio-economic implications of bioenergy production for scientists, practitioners, and decision makers who are interested in a biomass supply, cost-value, macro and micro-economics and value-chain perspective of bioenergy production.*”

AUTHOR CONTRIBUTIONS

SE and PA wrote the review.

REFERENCES

- EIA. (2013). *Energy Information Administration; International Energy Outlook*. Available at: [www.eia.gov/forecasts/ieo/pdf/0484\(2013\).pdf](http://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf)
- Janssen, R., and Rutz, D. (2012). *Bioenergy for Sustainable Development in Africa*. Heidelberg: Springer.
- Parente, E. J. S. (2003). *Biodiesel: Uma Aventura Tecnológica num País Engraçado*, first Edn. Fortaleza: Unigráfica.
- Pousa, G. P. A. G., Santos, A. L. F., and Suarez, P. A. Z. (2007). History and policy of biodiesel in Brazil. *Energy Policy* 35, 5393–5398. doi:10.1016/j.enpol.2007.05.010
- Ruppert, H., Kappas, M., and Ibendorf, J. (2013). *Sustainable Bioenergy Production: An Integrated Approach*. Heidelberg: Springer.

World Bioenergy Association. (2014). *Global Bioenergy Statistics*. Available at: www.worldbioenergy.org

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Edrisi and Abhilash. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

