Toward Direct Biosynthesis of Drop-in Ready Biofuels in Plants: Rapid Screening and Functional Genomic Characterization of Plant-derived Advanced Biofuels and Implications for Coproduction in Lignocellulosic Feedstocks

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Blake Lee Joyce

August 2013

Abstract

Advanced biofuels that are “drop-in” ready, completely fungible with petroleum fuels, and require minimal infrastructure to process a finished fuel could provide transportation fuels in rural or developing areas. Five oils extracted from *Pittosporum resiniferum*, *Copaifera reticulata*, and surrogate oils for *Cymbopogon flexuosus*, *C. martinii*, and *Dictamnus albus* in B20 blends were sent for ASTM International biodiesel testing and run in homogenous charge combustion ignition engines to determine combustion properties and emissions. All oils tested lowered cloud point. Oils derived from *Copaifera reticulata* also lowered indicated specific fuel consumption and had emissions similar to the ultra-low sulfur diesel control. Characterization of the biosynthetic pathways responsible for the sesquiterpene-rich *Copaifera*-derived oils could lead to production of these oils in biofuel feedstocks.

The *Copaifera officinalis* transcriptome sequencing, assembly, and annotation identified eight terpene synthase genes in *C. officinalis* and *C. langsdorffii* that produced mono- and sesquiterpene products in functional assays. The terpene synthases characterized produced the major fraction of sesquiterpenes identified in *C. officinalis* leaf, stem, and root tissues as well as the oils tested previously. This initial characterization will support future investigation of sesquiterpene biosynthesis in the *Copaifera* genus to understand how liters of sesquiterpene oils are produced for biotechnology applications and the mechanism responsible for the geographical biochemical variation seen in sesquiterpene-producing New World species compared to diterpene-producing African species.

Lastly, *Cymbopogon flexuosus* and *C. martinii* biomass production in small field trials, as well as oil and ethanol yield from biomass were investigated to determine the feasibility of producing the advanced biofuels in lignocellulosic feedstocks. *C. flexuosus* and *C. martinii* ethanol yields from biomass were lower than *Panicum virgatum*, but had an average oil yield of 85.7 kg ha^{-1} [ha^-1] and 67.0 kg ha^{-1} [ha^-1], respectively. Combined ethanol and oil value for *C. flexuosus* and *C. martinii* were higher than *P. virgatum* ethanol value. This suggests that the oils from *C. flexuosus* and *C. martinii* are more suitable as high-value fermentation coproducts rather than as low-value advanced biofuels. Increasing yield of oil or alternative production schemes could lead to economically feasible advanced biofuel production.

Preface

Our growing world demands more and more from agriculture and natural resources. This reality is unavoidable. Natural population growth coupled with the fundamental search for a better quality of life has led to agricultural demands being levied on the world's natural order. However, through technology and research, yields of food and agricultural products that seemed impossible to imagine 100 years ago are all too common today. In 1907, the average corn yield was 1 706 kg ha^{-1}, but has since increased steadily to a reported 9 300 kg ha^{-1} in 2005 (Egli, 2008). Likewise, soybean yield increased from approximately 1000 kg ha^{-1} to 3 000 kg ha^{-1} in Illinois and Iowa from 1925 to 2006 (Egli, 2008). These documented continuous yield increases result from many interacting factors several of which are well known: fertilizer application, use of herbicides and pesticides, plant breeding and genetic improvement, and enhanced management practices to reference a few (Egli, 2008).

Despite these and other past agricultural successes, a new generation of challenges has risen in the wake of global demands for the comforts of a modern lifestyle traditionally enjoyed primarily by citizens of developed economic powers. To meet these demands future generations will have to provide electricity, transportation, clean water, a stable high-quality food supply which includes protein, and sound housing for billions of people with existing arable land and natural resources. Additionally, modern agriculture will have to meet these demands with the added constraints of sustainable production as climate and land use change are inexorably linked to increasing agriculture production (Godfray et al., 2010; Lambin & Meyfroidt, 2011). Climate and land use change predictions portray reduction of food production

and availability, stability of food supplies, access to food, and utilization of food with these impacts affecting poor developing countries disproportionately (Schmidhuber & Tubiello, 2007). The remaining amount of arable land that exists and the environmental impacts associated with bringing these lands into production are difficult to predict. In one estimate, there are 2.4 billion hectares of land suitable for cereal production of which 1.5 billion are already cultivated (Pingali et al., 2008). Pingali *et al.* note that the majority of the uncultivated land exists in South America and sub-Saharan Africa that require irrigation; however, availability of fresh water will be limited in the future and will likely be a major factor that ultimately determines future agricultural production (Strzepek & Boehlert, 2010). Therefore increasing agricultural production to feed our growing world is not as simple as opening more land for production.

The intersection of agriculture and petroleum

This leads to one seemingly simple conclusion: increasing production capacities on existing agricultural land will have to be the primary method to meet increased demand (Godfray et al., 2010). This has been traditionally accomplished through intensification of agriculture; which in turn leads to greater demands for agricultural inputs, e.g. liquid fuels for modern machinery, fertilizers, and pesticides (McMichael et al., 2007). This remains a complex issue in the face of increased global demand for liquid fuels and the impending predictions of global peak oil. A future peak oil production is now an accepted idea and debate has slowly shifted to timing peak oil production rather than its potential existence (de Almeida & Silva, 2009). A review of peak oil date predictions shows that independent analysts tend to predict peak oil between 2015-

2020 while agencies like the U.S. Energy Information Administration (EIA), International Energy Agency, Shell, and the Cambridge Energy Research Associates (CERA) predict peak oil occurring after 2030 and as late as 2112 (de Almeida & Silva, 2009).

Peak oil predictions, much like in global food production predictions of the past, have been confounded by technological development. Overall, declines in oil production from 2000 to 2008 suggest that about 1.8 million barrels day^{-1} needs to be replaced from countries that are not members of Organization of the Petroleum Exporting Countries (OPEC) (Allsopp & Fattouh, 2011). To meet this production decline, countries have turned to other sources of liquid fuels. Some of these new sources are hydraulic fracturing, or fracking, which breaks source rock to access sequestered gas and crude oil commonly referred to as 'tight oil,' deep-water oil drilling in the Gulf of Mexico, Alaska, and Greenland, functionalization of Canadian oil sands, and biofuel production in the United States and Brazil (Allsopp & Fattouh, 2011; Kerr, 2012). While these new sources of liquid fuels offset declines in supply they are comparatively more expensive to extract than historical sources (Allsopp & Fattouh, 2011; Moerschbaecher & Day Jr., 2011). Additionally, these technologies carry specific environmental risks and will not likely produce at a rate similar to the best oil fields that have already begun to experience production decline (Kerr, 2012).

Fracking is a technique used to access natural gas deposits that are in otherwise impermeable rock. Water and proprietary chemical mixtures are pumped underground to pressurize and fracture bedrock to yield natural gas. Concerns over groundwater contamination from fracking operations have led to public outcry. Fracking fluid surface spills in Dunkard Creek, Hopewell Township, and Dimock, Pennsylvania led to runoff

into nearby water systems resulting in wildlife deaths and, ultimately, fines from state regulators (Rahm, 2011). Increased methane concentrations in New York wells were determined to be from deep thermogenic methane sources based on C-CH_4 isotopes and the presence of ethane and propane which are not usually found in shallow water sources (Osborn et al., 2011). Debate about water quality and public safety versus economic benefits of shale gas production will likely continue into the foreseeable future, but it is unclear how this will affect production rates. The environmental risks involved in deep-water oil drilling have recently been demonstrated in the Gulf of Mexico during the British Petroleum (BP) Deepwater Horizon oil spill (Camilli et al., 2010). Deep-sea oil fields also tend to be depleted faster than other oil fields, and require 10 to 20 years to establish (de Almeida & Silva, 2009). Ultimately, meeting demands for energy, like demands for agricultural products, will not be as simple as increasing production lands.

An unusual time in energy supply and demand

Recently, fracking in the United States has resulted in a boom of natural gas supplies. However, most of the infrastructure in the US uses coal or oil. For instance, in 2008 the majority of electricity produced in the US was derived from coal (48.5%) followed by natural gas (21.3%) and nuclear (19.6%) (Sovacool, 2009). In 2011, the US transportation sector consumed 26.5 x 10^{15} kJ of petroleum fuels while only consuming 775 x10^{12} kJ of natural gas (U.S. Energy Information Administration, 2011). Therefore, utilizing the new supply of natural gas requires building new infrastructure or converting current infrastructure. Additionally, the cost of natural gas source extraction, future

competing technologies, and developing global natural gas markets has led to uncertainties when considering future energy policies but likely results in natural gas reserves competing well in electricity production as the greenhouse gas emissions of natural gas is favorable compared to coal (Paltsev et al., 2011).

Alternatively, petroleum-derived liquid fuels drive transportation infrastructure in the US, which in turn, literally drives goods from production to consumer. In 2011 the US became a net exporter of finished petroleum products from refineries. However, this does not mean that the US has reached energy independence. In fact, the US is still a net importer of *crude* oil in 2011, or a net importer of supply oil to make finished products (Energy Information Administration, 2011). However, due to decreased demand in the US for *finished* petroleum products, e.g. gasoline, highway diesel, jet fuel, etc., use of ethanol in gasoline to offset supply, and increased production of finished product from Gulf Coast refineries resulting from lower cost of local West Texas Intermediate crude compared to Brent crude from the North Sea lead to profitable exportation of finished petroleum products from the US (Bailey & Lee, 2012). In short, net exportation of finished petroleum products from the US resulted from temporary market situations rather than energy supply independence from countries such as Venezuela or the Organization of Petroleum Exporting Countries (OPEC).

The impact of biofuels in the US: an issue of supply

Though biofuels supplied 4.6×10^{15} kJ to the transportation sector in 2011 the total demand was 13.7×10^{15} (U.S. Energy Information Administration, 2011). This estimate includes biodiesel, ethanol, and wood-derived fuels. In short, biofuels were only able to

offset 33.6% of the fuels required to supply the transportation sector of the US. Comparatively, the total energy consumption of the US was 100.3×10^{15} kJ (U.S. Energy Information Administration, 2011). Therefore, biofuels provide only 4.6% of total consumed energy of the U.S. In light of this, two distinct transportation biofuel consumer markets can be broadly considered amongst countries of the world: 1) consumers in developed regions that require large volumes of inexpensive liquid fuels to maintain their daily living standard, and 2) consumers in rural or remote regions that would benefit from any volume of liquid fuel as they are accustomed to living without modern conveniences. To illustrate this, developed countries consume five to ten times more GJ capita^{-1} in residential, industry, and transportation sectors than developing countries (Chow et al., 2003).

While biofuels offset a small portion of energy consumption in the U.S. and other developed countries, rural or consumers in developing countries could benefit from even small inputs in liquid transportation fuels. In developing countries, the highest losses of agricultural production occur during post-harvest from food spoilage as a result of poor transportation infrastructure and storage facilities (Godfray et al., 2010). *Access to liquid fuels for reliable transportation of goods and services, electricity from generators for refrigeration, and non-timber cooking fuels would transform the energy economics and ecological footprint of rural consumers.*

However, biofuels need to have several characteristics to be effective in these rural or developing conditions:

1) The (bio)fuel should yield supplies of liquid fuels frequently, preferably yearly or several times a year to be sustainable;

2) The (bio)fuels should require minimal infrastructure to extract, process, and deliver finished liquid fuel products to consumers;

3) The (bio)fuels should have long-term storage capabilities to meet energy demands throughout boom and bust years;

4) The (bio)fuels should mix with existing supplies of petroleum fuels without restriction (are completely fungible with petroleum fuels) to meet fluctuations in supply, demand, and therefore cost, of petroleum fuels;

Considering these qualities, traditional bioethanol and biodiesel production violate quality 2 and/or quality 4. As such, additional source of ethanol and fatty acid-derived biodiesel need to be considered. However, biofuel production in developing countries is complicated by lack of uniform adoption of modern crop production techniques (Pingali et al., 2008). Alternatively, this could also be an advantage as each country will adopt a production scheme that makes sense locally.

The following chapters detail an investigation into direct use of plant metabolites for “drop-in ready” advanced biofuel production. These biofuels are usable directly extracted from biomass and do not require chemical conversion as bioethanol and biodiesel.

Firstly, I begin this dissertation with a review of relevant topics related to current biofuel production from plant biomass. The topics of how the chemical classes present in petroleum-derived gasoline and diesel affect fuel and engine properties, current methods and emerging technologies for production of biofuel and coproduct chemicals from biomass, and chemical fractions present in lignocellulosic feedstocks such as switchgrass and poplar are discussed. Lastly, current conversion technologies of

lignocellulosic feedstock chemical fractions are discussed, and potential routes for engineering direct production of extractable, ‘drop-in’ ready biofuels from plant biomass are discussed. Chapter I concludes with the following:

> While the relationship between petroleum chemistry and fuel properties are generally understood, further investigation into plant-derived biochemicals will be required as little is known about their fuel properties and biochemicals do not fit neatly into traditional petroleum fuel classes.

In the second chapter, *Copaifera* trees and their oils are discussed specifically. While *Copaifera* production ecology was reported to supply liters of oleoresin tree^{-1} yr^{-1} recent studies of *Copaifera* spp. production suggest each tree produces only a few hundred mL within a lifetime or across many years. Chapter II concludes with the following:

> Study of this unique terpene biosynthesis pathway that produces liters of sesquiterpenes at a time could still yield understanding of terpene production that is useful for other terpenoid-products such as artemisinin or advanced plant-derived hydrocarbon biofuels.

The third chapter investigates the use of plant-derived hydrocarbon biochemicals as advanced biofuels. A total of six hydrocarbon oils were investigated from five different plant species: *Cymbopogon flexuosus* essential oil (comprised primarily of citral; a monoterpene aldehyde), *C. martinii* essential oil (comprised primarily of geraniol; a monoterpene alcohol), *P. resiniferum* oil (comprised primarily of monoterpenes and short-chain alkanes), raw *Copaifera reticulata* oil directly tapped from the tree (cyclic sesquiterpenes and diterpene resin acids), steam distilled *Copaifera reticulata* oil (cyclic

sesquiterpenes), and anethole (an aromatic compound similar to monolignols). These oils were blended in ultra-low sulfur highway diesel #2 (ULSD) to 20% (B20) and the resulting B20 mixtures were subjected to standard ASTM International biodiesel tests. Additionally, the B20 blends were run in homogenous charge compression ignition (HCCI) engines to characterize their engine operability and emissions properties. Chapter III concludes with the following:

> Replacing petroleum-derived liquid fuels with renewable fuels resources has become a major research focus for a number of environmental and political reasons. However, only alkyl esters of fatty acids are being considered for use as biodiesel. Other hydrocarbons from biomass for advanced fuels are of interest but few reports have investigated the fuel properties and combustion properties of these chemicals.

In the penultimate chapter, the sesquiterpene biosynthetic pathway of *Copaifera officinalis* and *C. langsdorffii* is functionally characterized. First, a *de novo* transcriptome from *C. officinalis* leaf, stem, and root tissue was sequenced using the Illumina next-generation sequencing platform. The assembled transcriptome was annotated and analyzed to identify putative genes involved in mono-, sesqui-, or diterpene biosynthesis. These genes were isolated, cloned, and expressed in recombinant *E. coli*. Each terpene synthase was characterized with *in vitro* assays using geranyl pyrophosphate (the monoterpene precursor) and farnesyl pyrophosphate (the sesquiterpene precursor). Chapter IV concludes with the following:

> *Copaifera* species, often referred to as the ‘diesel trees,’ native to the Americas produce a sesquiterpene-rich oleoresin that is collected by tapping the trunks of

mature trees. While the oleoresin has been used as traditional medicine in parts of Central and South America, the oleoresin has also been reportedly used in diesel engines as fuel. While production of biofuels from the *Copaifera* genus might currently not be economically feasible, the characterization of transcriptomes allows for gene discovery related to the unique biosynthetic pathway in *Copaifera* species, which, in turn, should improve our understanding of the copious production of sesquiterpene oleoresins present in these species. Here we describe *de novo* assembled transcriptomics and functional characterization of the *Copaifera officinalis* sesquiterpene biosynthetic pathway. Annotated sesquiterpene synthases of *C. officinalis* and *C. langsdorffii* contained mono-, sesqui-, and diterpene amino acid motifs previously described in gymnosperm and angiosperm terpene synthase classes. Functional characterization of the identified sesquiterpene synthases resulted in production of all major sesquiterpenes found in *C. officinalis* tissues. The *de novo* transcriptome of the northernmost *Copaifera* New World species, *C. officinalis*, was robust enough to isolate sesquiterpene synthases in *C. langsdorffii*: a species native to the southernmost range of the *Copaifera* genus in the New World. Functional characterization of *C. officinalis* suggests a link between ancestral monoterpene and diterpene oleoresin production in gymnosperms and sesquiterpene oleoresins present in angiosperm tree species.

Lastly, in chapter five investigation of the most viable way to produce novel “drop-in ready” advanced biofuels from biomass is needed. To this end we have identified the grass species *Cymbopogon flexuosus*, lemongrass, and *C. martinii*, palmarosa, which

produce terpenoid essential oils in leaf tissues as a model system. Terpenoids have been suggested as potential advanced biofuels since 1980 by the Nobel Prize winning Melvin Calvin (Calvin, 1980). Furthermore *C. flexuosus* and *C. martinii* terpenoid oils were previously characterized in Chapter III in homogenous charge compression ignition (HCCI) engines. Chapter V concludes with the following:

> *Cymbopogon flexuosus* and *C. martinii* are perennial grasses grown to produce essential oils for the fragrance industry. The objectives of this study were (1) to evaluate biomass and oil yields as a function of nitrogen and sulfur fertilization, and (2) to characterize their utility for lignocellulosic ethanol compared to *Panicum virgatum* (switchgrass). Mean biomass yields were 12.83 Mg lemongrass ha^{-1} and 15.11 Mg palmarosa ha^{-1} during the second harvest year resulting in theoretical biofuel yields of 2541 and 2569 L ethanol ha^{-1} respectively compared to reported 1749-3691 L ethanol ha^{-1} for switchgrass. Pretreated lemongrass yielded 198 mL ethanol (g biomass)$^{-1}$ and pretreated palmarosa yielded 170 mL. Additionally, lemongrass yielded 85.7 kg essential oil ha^{-1} and palmarosa yielded 67.0 kg ha^{-1} with an estimated value of (USD) 857 and 1005 ha^{-1}. These data suggest that dual use crops such as lemongrass and palmarosa may increase the economic viability of lignocellulosic biofuels.

TABLE OF CONTENTS

Introduction 1

Abstract 2

1. Introduction 4

2. Replacing petroleum commodities: can we grow barrels of oil? 8

2.1 Plant-derived biofuels: two engines, two fuels, two crops? 10

2.2 Coproduction of biogasoline, biodiesel, green chemicals, and high-value coproducts: towards growing green petroleum 18

3. Engineering plants to make the 'biofuel feedstock buffalo' 20

3.1 Bale to barrel: strategies for engineering the perfect petroleum-replacement feedstock 22

3.2 Improving unused portions of switchgrass biomass 34

4. Conclusions 36

Appendix 38

CHAPTER I 49

Diesel Trees 49

Abstract 50

1 Introduction 51

2 Chemicals Present in *Copaifera* Oleoresins 54

3 Biosynthesis of *Copaifera* Oleoresins: What Conifers Can Teach Us 56

4 Biological Functions of Oleoresin 59

5 Oleoresin Production Ecology 61
6 Comparing Oleoresin to Diesel Fuel 64
7 Future Scope of Research and Development 67
Appendix 70
CHAPTER II 74
Biodiesel fuel properties and combustion characteristics of drop-in ready plant-derived oils that do not contain fatty acids 74
Abstract 77
1. Introduction 80
2. Materials and methods 83
Plant species and their oils 83
Chemical composition determination 85
Fuel property testing 85
HCCI engine specifications and operation 85
3. Theory/calculation 86
Prediction of air intake temperatures at MFB50 top dead center 86
4. Results 87
Plant oil chemical composition 87
Plant oil fuel properties 88
Engine combustion experiments 90
5. Discussion 92
Comparison of the experimental plant oil chemistries and traditional petroleum fuel classes 92

Fuel, combustion, and emission properties of B20 plant oil blends 96
Overall trends in the experimental B20 plant oil blends 100
6. Conclusions 105
Acknowledgements 107
Appendix 108
CHAPTER III 141
De Novo Transcriptome mining of Copaifera officinalis and Functional characterization of sesquiterpene synthases in Copaifera officinalis and Copaifera langsdorffii 141
Abstract 144
Background 146
Results 150
C. officinalis transcriptome sequencing, assembly, and annotation 150
Annotated mevalonic acid, methylerythritol phosphate, and terpene synthase pathway transcripts 151
Isolation of C. officinalis and C. langsdorffii complementary DNA 153
Sequence similarity of isolated cDNAs with known TPS and motifs 153
Functional terpene synthase assays of C. officinalis and C. langsdorffii TPS cDNAs 155
Discussion 156
C. officinalis transcriptome assembly and annotation 156
In silico and functional characterization of C. officinalis and C. langsdorffii terpene synthase activity 158
Implications for future work 163

Conclusions 164
Methods 165
Plants 165
Illumina library preparation and sequencing of C. officinalis 166
De novo assembly of the C. officinalis transcriptome 167
Functional annotation of assembled C. officinalis reads 168
Phylogenetic analysis of C. officinalis and C. langsdorffii terpene synthases 169
Isolation and functional characterization of identified terpene synthase contigs .. 170
Authors' contributions 171
Appendix 174
Figures 175
Tables 188
Chapter IV 199
Ethanol and high-value terpene co-production from lignocellulosic biomass of *Cymbopogon flexuosus* and *Cymbopogon martinii* 199
Abstract 202
1. Introduction 203
2. Materials and Methods 205
2.1. Field experiments 205
2.3. Essential oil extraction 208
2.4. Gas Chromatography–Mass Spectroscopy Quantitative Analysis 208
2.5. Statistical analysis 209

2.6. Pretreatment, ethanol fermentation, and high performance liquid chromatography analysis 211
2.7. Determination of saccharification efficiency and lignin composition 212
3. Results and Discussion 212
3.1. C. flexuosus and C. martinii biomass and essential oil yield in field trials 214
3.2. Lemongrass essential oil responses 217
3.3. Palmarosa essential oil responses 220
3.4. Biomass composition and recalcitrance 221
3.5. Ethanol fermentation and inhibition from biomass metabolites 222
3.6 From field to fermenter: estimated market values of coproduction 224
4. Conclusions 226
Appendix 228
Tables 229
Figures 235
Conclusion 243
List of References 247
VITA 271

LIST OF TABLES

Table 1. Chemical class influence on petroleum distillates' fuel and physical properties and biochemical alternatives. 45

Table 2. Lignocellulosic biofuel feedstock chemical composition in percent dry weight. 46

Table 3. Biofuels and coproducts derived from biomass metabolites and conversion techniques used. 47

Table 4. Calculated seed composition of switchgrass and yield per hectacre of each component. 48

Table 5. Three major sesquiterpenes present in the oleoresins of Copaifera species. . 73

Table 6. Potential sources and chemical classes of plant-derived biofuels. 109

Table 7. Comprehensive list of ASTM International biodiesel tests performed on B20 plant oil blends and ULSD. 110

Table 8. Chemical composition (%) of plant oils divided into petroleum chemical classes based on chemical structure. 111

Table 9. Chemical composition (%) of plant oils divided into biochemical classes based on biosynthesis pathways. 112

Table 10. Standard biodiesel fuel properties of B20 plant-derived oils mixed with ultra-low sulfur diesel. 113

Table 11. Chemicals identified in plant-derived oils used in blending stocks and ULSD. 137

Table 12. *De novo* assembly statistics for *C. officinalis* transcriptome. 188

Table 13. KEGG annotated transcripts in terpenoid biosynthetic pathways for the 1KP and TAMU *C. officinalis* transcriptome. 189

Table 14. Contig length, annotation, and signal peptide prediction of isolated terpene synthases. 190

Table 15. List of volatile terpenes found during *C. officinalis* and *C. langsdorffii* enzyme assays with GPP and FPP. 191

Table 16. Primers used to isolate full length transcripts and cloning terpene synthase into vectors. 192

Table 17. Functionally characterized terpene synthase protein sequences used in the phylogenetic analysis. 194

Table 18. Lemongrass and palmarosa biomass without pretreatment fraction composition and enzymatic saccharification efficiency. 229

Table 19. The repeated measures main and interaction effects of crop, nitrogen (N), sulfur (S) fertilization, and harvest on biomass dry weight yield (DW) and essential oil composition (EO). 230

Table 20. The effect of N, S and harvest on lemongrass EO content, and the composition and yield of β-caryophyllene, (Z)-citral and (E)-citral. 231

Table 21. P-values showing the effect of N, S and harvest on palmarosa EO content, and the composition and yield of geraniol and geranylacetate. 232

Table 22. Biomass fractions remaining in fermentation liquid after SSF for enzyme optimization, dried harvested biomass, and pretreated biomass as determined by HPLC. 233

LIST OF FIGURES

Figure 1. Chemical conversion versus *in-planta* biochemical conversion of biofuel metabolites. 39

Figure 2. Current (solid lines) and future (dotted lines) conversion techniques for production of biofuels and coproducts from fractions of plant biomass. 41

Figure 3. Systematic processing of the ideal biofuel plant feedstock. 43

Figure 4. Identification and tapping of *Copaifera langsdorffii* trees near Nova Odessa, Sao Paulo State, Brazil. 71

Figure 5. Biochemicals present in each plant-derived oil as determined by GC/MS. .. 114

Figure 6. Distillation ranges determined by ASTM International testing for B20 blends of plant-derived oils. 115

Figure 7. B20 blend biodiesel fuel properties as determined by ASTM International cetane number (A), cloud point (B), flash point (C), net heat of combustion (D), viscosity (E), and sulfur content (F) tests. 117

Figure 8. Air intake temperatures used to modify combustion phasing (MFB50) in homogenous charge compression ignition engine modeling to achieve a range of before and after top dead center (represented as 0 °CA) (A), the range of air intake temperatures used for each B20 blend (B), and relation of B20 blend cetane number to combustion phasing. 119

Figure 9. Indicated mean effective pressure (IMEP) controlled for energy content of B20 blends across combustion phasing (A), and indicated specific fuel consumption (ISFC) for B20 oil blends compared to ULSD control (B). 121

Figure 10. Total length of combustion from 10% mass fraction burned to 90% mass fraction burned (MFB10-90) for B20 oils blends in relation to combustion phasing from top dead center (MFB50). 123

Figure 11. Maximum rate of pressure increase in the cylinder (dP/dCA) at different combustion phasing (MFB50). 125

Figure 12. Unburned hydrocarbon emissions from B20 oil blends in relation to combustion phasing. 127

Figure 13. Nitrous oxide (NOx) emissions in relation to MFB50 (A) and air intake temperature (B) of B20 oil blends..... 129

Figure 14. Carbon monoxide emissions in relation to combustion phasing of B20 oil blends and ULSD control. 131

Figure 15. Soot emissions (filter smoke number or mg m^{-3}) in relation to combustion phasing of B20 oil blends. 133

Figure 16. GC/MS chromatograms of plant-derived oils used in B20 blends and the ULSD control..... 135

Figure 17 - Comparison of the 1KP and TAMU assembled transcriptome annotation using the KAAS SBH algorithm. 175

Figure 18 - Unrooted phylogenetic analysis of identified *C. officinalis* and *C. langsdorffii* terpene synthases and previously functionally characterized mono-, sesqui-, and diterpene synthases. 177

Figure 19 - Alignment of isolated *C. officinalis* sesquiterpene synthases with *Medicago truncatula* (+)-δ-cadinene synthase (GenBank XP_003596565.1)..... 179

Figure 20 - Alignment of *C. officinalis* contig 5 with *A. thaliana* (GenBank NP_192187.1) and *Solanum lycopersicum* (GenBank NP_001234008.1) copalyl-diphosphate synthases. 183

Figure 21 - GC-MS chromatograms of *C. officinalis* (Co) and *C. langsdorffii* (Cl) terpene synthase enzyme assay using FPP. 186

Figure 22 - GC-MS chromatograms of *C. officinalis* (Co) and *C. langsdorffii* (Cl) terpene synthase enzyme assays using GPP as a substrate. 187

Figure 23. Main effects and interaction plots for dry weight biomass (kg ha^{-1}) and essential oil (kg ha^{-1}) production across both crops (lemongrass and palmarosa). 235

Figure 24. Means of lemongrass essential oil content, (*E*)-citral content and yield, (*Z*)-citral yield, and â-caryophyllene yield. 238

Figure 25. Main effects and interaction plots palmarosa essential oil content, geranylacetate content and yield, and geraniol content and yield. 239

Figure 26. Bench top simultaneous saccharification and fermentation (SSF) ethanol yield from extracted whole biomass based on filter paper units of enzymes. 240

Figure 27. Final ethanol yield from lemongrass biomass (mg g^{-1}) and palmarosa (mg g^{-1}) that was (EX) or was not (NE) previously extracted for essential oils in comparison to two lots of BioEnergy Science Center (BESC) control switchgrass. 241

Figure 28. Interaction effect of nitrogen×sulfur×harvest on β-caryophyllene and (*Z*)-citral content in lemongrass essential oil. 242

INTRODUCTION

Designing the perfect plant feedstock for biofuel production:

Using the whole buffalo to diversify fuels and products

A version of this chapter was originally published by Blake L. Joyce and C. Neal Stewart, Jr.

B.L. Joyce, and C.N. Stewart, Jr. “Designing the perfect plant feedstock for biofuel production: Using the whole buffalo to diversify fuels and products.” *Biotechnology Advances* (2011): doi:10.1016/j.biotechadv.2011.08.006.

The manuscript was written by Blake L. Joyce, and reviewed by C. Neal Stewart.

Abstract

Petroleum-derived liquid fuels and commodities play a part in nearly every aspect of modern daily life. However, dependence on this one natural resource to maintain modern amenities has caused negative environmental and geopolitical ramifications. In an effort to replace petroleum, technologies to synthesize liquid fuels and other commodities from renewable biomass are being developed. Current technologies, however, only use a portion of plant biomass feedstocks for fuel and useful products. Using the whole “feedstock buffalo” optimally using all portions and biochemicals present in renewable biomass will enhance the economic and environmental feasibility of biofuels and coproducts. To accomplish this optimization, greater understanding of the relationship between liquid fuel and bioproduct properties and plant chemistries is needed. Liquid fuel properties and how they relate to biochemistry and petrochemistry are discussed. Several metabolic engineering strategies for increasing the efficient use of

dedicated feedstock plants such as switchgrass biomass are outlined. Enhanced biofuel yields and high-value commodities from biomass are needed to sustainably replace petroleum-based products.

Keywords

Advanced biofuels, Metabolic engineering, Plant biotechnology, Consolidated bioprocessing, Plant-extractable biofuels, Biobased coproducts

Abbreviations

SSF – simultaneous saccharification and fermentation; FAME – fatty acid methyl ester; FAEE – fatty acid ethyl ester; PHA – polyhydroxyalkanoates; PHB – poly-3-hydroxybutyrate; TAL – tyrosine ammonia-lyase; HCT – hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; DGAT – diacylglycerol acyltransferase; LEC – leafy cotyledon

1. Introduction

Modern, industrialized society relies on a single natural resource to provide a plethora of commodities and conveniences that would be hard to envision living without: petroleum. Petroleum not only provides liquid transportation fuels, but also provides the asphalt which literally paves the way for transportation. Petroleum provides heating fuels, plastics and other materials which have revolutionized everything from how we package and store food to modern medical products (Thompson et al., 2009). Petroleum has in some way contributed to nearly every aspect of modern daily life, but the end of petroleum is in sight (Allsopp & Fattouh, 2011; de Almeida & Silva, 2009). Though predictions for an international peak in oil production vary, there is a general agreement that peak oil will eventually occur (Allsopp & Fattouh, 2011). These predictions have recently been complicated by increased production of oil from nontraditional sources such as deep-water drilling and Canadian oil sands (Allsopp & Fattouh, 2011; Kerr, 2012). While these new sources of liquid fuels offset declines in supply they are comparatively more expensive to extract than historical sources (Allsopp & Fattouh, 2011; Moerschbaecher & Day Jr., 2011). This means that in the future the rate of crude oil production will plateau from relatively cheaper sources which will lead to an increase in price as demand continues to increase.

In addition to eventual supply plateaus, every positive benefit that petroleum has provided there seems to be a negative environmental ramification. The Deepwater Horizon oil spill released 4 million barrels of oil into the Gulf of Mexico which has had a series of ecological and economic impacts on the states and countries lining the Gulf (Camilli et al., 2010). Extensive use of asphalt has created a phenomenon known as ‘urban heat islands’ which increases energy consumption and can increase mortality rates in urban centers (Rizwan et al., 2008). Emissions from combustion engines have led to debate and growing concern over air quality and climate effects such as more intense storms (Knutson et al., 2010). Plastics make up 10% of human wastes, do not readily degrade, and when they do they release toxic chemicals that have started to bioaccumulate across the globe (Thompson et al., 2009). There have also been negative geopolitical ramifications associated with petroleum production and consumption. Included in the list are war, economic and political instability, and increasing disparity between rich and poor countries. Taken together these considerations have led researchers to investigate a number of technologies to replace petroleum-derived commodities with renewable and more environmentally benign substitutes. Replacing petroleum commodities with an inexpensive, renewable resource that can be produced in any country in the world would lead to a second green revolution focused on human needs going beyond food (Mooney, 2009).

Of specific interest, biofuel research has taken aim at replacing petroleum liquid fuels with biochemicals derived from crop and forest residues, algae, and bio-derived waste materials. There have been a number of policies and incentives directed at developing both ethanol (Hoekman, 2009; Martin, 2010) and biodiesel (Hoekman, 2009) into mature cost-effective technologies. However, current biofuels are not ideal liquid fuels when characteristics like fuel properties and compatibility with existing infrastructure are considered. Benefits and drawbacks of biofuel production will be further discussed in Section 2.1. Plant biotechnology and microbial biotechnology have been proven to be useful tools in improving biomass processing and biorefinery product yields (Hermann & Patel, 2007; Octave & Thomas, 2009). Biocatalyst reactions, or reactions driven by enzymes, have advantages over organic chemistry synthesis, e.g. the ability to produce complex molecules efficiently (Wohlgemuth, 2009). Specifically, complex chemicals such as human growth hormone are produced using biocatalysts and fermenters (Di Cesare et al., 2013). Although the use of biocatalysis of chemicals on large scale has been limited, biotechnology and bioprocessing have been applied extensively to biofuel production, which will be discussed in Section 3. The reasons why biofuels have become an attractive solution to replacing petroleum-derived liquid fuels has been addressed in a number of reviews, and as such is beyond the scope of this manuscript e.g. Hoekman, 2009. However, it is important to briefly discuss how first generation

(food crop-derived) and second generation (non-food crop-derived) biofuels were developed so as to understand developing next-generation biofuels.

In 2007, the U.S. Energy Independence and Security Act set incentives and a goal of 144 billion liters of biofuels per year by 2022 (Martin, 2010). Industry and researchers turned to available technologies in an attempt to begin to displace petroleum fuels immediately. In the US, ethanol for use as biofuels was first derived from fermented starch (usually maize grain); biodiesel was derived from alkyl esters of fatty acids derived from animal or plant triglycerides. Starch and plant oils are feedstocks easily accessible with liquid fuel synthesis technologies being well developed (Hoekman, 2009); in the case of ethanol fermentation, humans have been practicing it for millennia. But these technologies led to a now famous public outcry against using food sources to produce fuels. The outcry resulted from food prices that increased 4.0% in 2007 and 5.5% in 2008 compared to a 2.4% increase in 2006 and 2005 (Martin, 2010). In reality, the higher food prices were a result of several factors with corn-based ethanol production accounting for only about a fifth of the total food price increase of 4.0 and 5.5% (Martin, 2010). Despite this, biofuel research shifted more heavily to non-food sources such as corn stover and dedicated biofuel crops such as poplar, switchgrass, and algae.

2. Replacing petroleum commodities: can we grow barrels of oil?

The ultimate goal of biofuels is to completely replace petroleum-derived liquid fuels, especially for the transportation sector. But biomass, like a barrel of oil, contains a diverse array of chemicals that could be used to create many different commodities in addition to liquid fuels. Indeed, fuel could be the essential loss leader in the emerging bioeconomy (Bozell, 2008). Petroleum itself is formed from organic matter such as marine algae and plants heated to specific temperatures in the Earth's crust on geologic time scales. The formation of petroleum occurs throughout the world and the chemicals that are formed differ based on different locations and different source rock (Speight, 1999). Petroleum is so chemically complex and variable between each deposit that it has been traditionally characterized by bulk properties like distillation ranges and total atomic percentage. It was not until recently that individual chemicals present in petroleum could be identified using high resolution mass spectrometry (Marshall & Rodgers, 2008). The chemical complexity of petroleum has led to the petroleum industry adopting a number of technologies, e.g. catalytic reforming, hydrotreating, etc., to separate, refine and alter chemical fractions for specific uses (Matishev, 1994).

By comparison, biomass *is* immature petroleum that needs to be converted and refined into chemicals useful for finished fuel products. In the

current biofuel industry, chemically complex plant biomass is separated, thermally cracked or degraded by enzymes, and then converted into products using chemical synthesis or biological conversion. The key is the efficient conversion of biomass into petroleum-like chemicals on a biological timescale (second to hours) rather than a geologic timescale (millennia). The main factor that distinguishes petroleum from biomass is the use of biotechnology to fundamentally alter enzymes present in biomass; essentially, biotechnology enables researchers to engineer and fine-tune barrels of renewable (biomass-derived) petroleum. To put the concept into petroleum terminology, biotechnology could be considered *in vivo* refining, and can occur in plant biomass, microbes used to ferment the biomass, or a combination of both. A significant amount of work has gone into altering fermentation products in microbes, and several comprehensive reviews are available (Lee et al., 2008; Liu & Khosla, 2010; Peralta-Yahya & Keasling, 2010). Microbial fermentation to produce ethanol or isopropanol has been thoroughly studied (Peralta-Yahya & Keasling, 2010), and has several advantages over *in planta* synthesis of metabolites such as rapid screening on culture chips, short life cycles, ease of engineering resulting from relatively simple metabolic pathways, and more sequence data available (Wohlgemuth, 2009). As dedicated biofuel crops become more commonplace, however, *in planta* synthesis of biochemical fuels offer several advantages such as simple extraction and separation to yield products, and direct (efficient)

synthesis of hydrocarbons and high-value commodities using low-cost solar energy. Cyanobacterial or algal production of biofuels might likely be the best combination of microbial and plant production systems, but there are still significant barriers to these technologies and their use on a sustainable industrial scale remains in the long-term (Wijffels & Barbosa, 2010).

With this in mind, our focus here is on plant feedstock metabolism and biotechnology strategies for producing the 'perfect' dedicated biofuel feedstock. We liken the goal of this bioenergy feedstock design process to the American Indian paradigm of "using the whole buffalo." Prior to European settlers in America, North American plains people hunted buffalo (bison) for food, clothing, fuel, and many other needs in their daily lives. No part of the buffalo went to waste. We envisage, likewise, designer dedicated plant feedstocks that provide a plethora of high value fuels, bioproducts and materials. Biotechnology should be integral in designing this perfect feedstock; an ideal feedstock does not exist in nature (Gressel, 2008).

2.1 Plant-derived biofuels: two engines, two fuels, two crops?

There are many plant-derived biochemicals that could be used to replace petroleum fuels. The complex hydrocarbon fraction of petroleum fuels can be broken down into three general petrochemical fuel classes: paraffins (alkanes), naphthenes (cyclic alkanes), and aromatics. However, several subgroups such

as isoparaffins (branched alkanes) and olefins (unsaturated alkanes) exist (Wallington et al., 2006). Ultimately, fuel and engine operability properties of liquid fuels result from a combination of engine type, chemical composition, environmental conditions, e.g. ambient temperature, and vehicle parameters, e.g. heavy load versus light load, which all vary in real world applications. However, experimentation with simplified surrogate fuel mixtures has led to some understanding of how chemical components affect quality of fuels. Petroleum-derived fuels, physicochemical properties, and the effect of chemical class constituent on those properties are illustrated in Table 1. Each petroleum chemical class present in a fuel yields different physicochemical properties for that commodity; essentially, there is no perfect chemical constituent that translates to a perfect liquid fuel. Market and engineering demands, such as a low cloud point property and a high cetane number in diesel fuel requires mixing chemical classes that counteract each other. For example, aromatics in diesel fuel will provide low cloud point properties but also a low cetane number, whereas paraffins will provide high cetane numbers but will also begin to solidify at higher temperatures. Current liquid fuel demands and environmental regulations require catalytic cracking of heavier petroleum fractions, e.g. aromatics and naphthenes, to form smaller hydrocarbons (Dupain et al., 2003). Blending is a crucial process in petroleum fuel synthesis, because mixing different chemical classes allows for the vast flexibility to meet market and

environmental demands. Biofuels, however, are conspicuously homogeneous in their chemistries, which is in great contrast to plant biochemistry where the typical vascular plant is composed of over 50,000 different chemicals (Hartmann, 2007). Ethanol or butanol for gasoline replacement are, obviously, single chemicals. Biodiesels derived from alkyl esters of either animal or plant oils contain more chemical diversity, but even then there are only 5-15 distinct chemicals based on the source material's composition of fatty acids. As there is a vast range of products derived from petroleum that could be replaced by biochemicals, we will focus on the suitability of different biochemicals as liquid fuels.

Currently in plant-derived biofuels, biomass is either deconstructed and sugars are fermented to produce ethanol/butanol, or oils are collected from oilseed crops to produce biodiesel through alkyl esterification reactions. This separation of biofuel crops seems to stem from availability of first generation biofuels as well as a restriction in technologies to derive both gasoline and diesel replacements in the same crop. The overall suitability of biofuels as a replacement for petroleum-derived fuels will depend on a plethora of factors including fuel properties, combustion and operability properties, emissions, and fungibility or compatibility with existing infrastructure. All of these properties can be linked directly to the liquid fuels' chemical components. Understanding of how fuel chemistry influences fuel properties is still insufficient, and reports of biofuel

effects in petroleum fuel blends are often difficult to compare due to inconsistent experimental designs, vastly different fuel chemistry between studies, and incomplete data sets (Lapuerta et al., 2008). A better understanding of petroleum fuel chemistry and how that relates to fuel properties will lead to a more intelligent design of biofuels (Pitz & Mueller, 2011). Fuel properties of any liquid fuel, whether petroleum or biomass derived, result from a combination of fuel chemistry and combustion engine type (Wallington et al., 2006). As such, we will discuss general gasoline chemistry for spark ignition engines and diesel fuel chemistry for compression ignition engines and how these two chemistries relate to fuel properties in more detail separately.

2.1.1. Current production technologies toward biogasoline

Gasoline is used in spark ignition engines. In these engines, fuel is carburetor-distributed or injected into a combustion chamber and then ignited with a spark at the appropriate time. Gasoline, therefore, needs to have a high volatility to combust instantly in presence of a spark but not as volatile as to prematurely detonate or to be explosive in storage. This range makes predicting optimal chemical composition for biogasoline difficult as each chemical class can have chemicals inside or outside the volatility range depending on carbon number, chemical structure, or side groups (Table 1). Petroleum gasoline distills at temperatures between 30 °C and 200 °C which contains the lower molecular

weight paraffins, naphthenes, isoparaffins, olefins, and aromatics in crude oil (Speight, 1999; Speight, 2002). Gasoline distilled directly from petroleum has a low octane rating, and as such, requires upgrading and blending with other refinery hydrocarbon streams (Pitz et al., 2007). Olefins, or unsaturated alkanes, are not present in significant amounts in crude oil, but are refined and blended with gasoline fractions to meet market requirements for fuel and emission properties (Speight, 1999).

As there are a few bio-based chemicals being investigated to be replace petroleum-derived gasoline, including ethanol, butanol, and hydrocarbons from thermochemical conversion, they will be referred to collectively as biogasoline. Thermochemical conversion of biomass, such as Fisher-Tropsch synthesis, will be discussed further in Section 2.1.3. Ethanol and butanol have different advantages and disadvantages as biofuels, and there is a debate centered on which is more suitable. Ethanol also has positive fuel characteristics when blended with gasoline such as reduced emissions of CO, CO_2, and hydrocarbons (Demirbas, 2009a). However, ethanol has a lower energy density than gasoline or butanol which means that ethanol will carry a car a shorter distance liter for liter; reduced kilometers per liter have led a lack of economic incentive for consumers to switch to using E85 fuel blends and flex-fuel cars (Martin, 2010). When compared to ethanol production, butanol has lower final concentrations (2% versus 15% for ethanol) and longer fermentation times which reduce its

usefulness in meeting widespread demand for liquid fuels (Pfromm et al., 2010). Despite these restrictions, butanol has greater energy density and is more hydrophobic which means that it is more suitable as a drop-in replacement for gasoline and more compatible with existing infrastructure. The benefits and drawbacks of ethanol and butanol as fuels directly result from their oxygen content. Currently, both fuels are fermented from biomass whether it is starch or sugars derived from lignocellulosic feedstocks. Though lignocellulosic biofuels are not widespread currently, biotechnology improvements have led to better lignocellulosic feedstocks for ethanol production (Fu et al., 2011). These improvements will likely lead to lignocellulosic biofuels being industrially viable in the near future, and apply to all potential products derived from lignocellulosic sources.

2.1.2. Current production of biodiesel

Diesel fuel is used in compression ignition engines. In these engines, fuel is injected into a combustion chamber where it is compressed until it reaches a specific pressure which causes the fuel to heat and ignite producing mechanical work. High cetane diesel fuels will ignite quickly to produce the maximum amount of work or transferable power to the engine. Jet fuel distills from nearly the same petroleum fraction as diesel fuel with a few more restrictions such as a limit on the percentage of aromatics and the need for low temperature operability to -40

°C (Carlsson, 2009). Currently, diesel fuel demand is growing at 3.5% which is greater than gasoline, kerosene, or jet fuel (Agency, 2011).

There are two major diesel replacement technologies that use plant biomass: the production of fatty acid alkyl esters (usually methyl esters or FAMEs) and ‘green diesel.’ Biodiesel has been primarily derived from four oilseed plants: soybean, oil palm, canola, and sunflower, although there are other crops being used in smaller amounts or being considered, e.g. *Camelina sativa*, cotton, and *Crambe abyssinica* (Carlsson, 2009). Plant-derived oils, which are primarily composed of acylglycerides, are too viscous to be used as fuel directly in engines without chemical structure modification or without heating to reduce viscosity. Synthesis of alkyl fatty acid esters requires an esterification reaction involving an alkyl alcohol, usually methanol, and a catalyst such as sodium hydroxide or a lipase biocatalyst (Demirbas, 2009b). The addition of a methyl ester group to a fatty acid does not radically alter the original fatty acid chemical structure, however separation from the glycerol backbone reduces viscosity and is the primary goal of this reaction (Figure 1A). Ethanol for production of fatty acid ethyl esters (FAEEs) has gained some interest because it can be produced from biomass; although methanol is far more common for economic reasons (Demirbas, 2009b). This reaction produces glycerol as a byproduct in a 1:9 ratio with biodiesel. Glycerol must be separated from the biodiesel product requiring processing and then disposal of alkaline or acid catalyst wastes (Du et al., 2008).

Biotechnology research to improve biodiesels has largely focused on developing products from byproduct glycerol and altering the fatty acid profile of oilseed crops to modify biodiesel properties. Glycerol has been used in chemical conversion (Thompson et al., 2009) and biological conversion (García et al., 2008; Rahmat et al., 2010; Zhang & Memelink, 2009) to make new products which will be discussed in more detail in Section 3.1.2. However, direct biosynthesis of biofuels *in planta* would allow for byproducts to reenter the metabolic pathways and reduce waste catalyst and water processing (Figure 1B). Altering the fatty acid composition of oilseeds has been suggested as a way to optimize biodiesel fuel properties (Agarwal, 2007; Knothe, 2009), but this approach will always limit biodiesel producers to the inherent properties of long-chain oxygenated alkanes and alkenes. In addition, annual food crops such as soybean, sunflower, and canola have unfavorable net energy output (Yuan et al., 2008c). Increasing unsaturated fatty acids in biodiesel improves cold operability characteristics but increased hydrocarbon and NOx emissions, and lowered cetane rating (Benjumea et al., 2010). Butanol and ethanol have also been blended with diesel, biodiesel, and even raw canola oil to enhance fuel properties such as lowering viscosity and increasing cold temperature operability characteristics (Demirbas, 2009a; Laza & Bereczky, 2011). However, there is still uncertainty among reports regarding biodiesel fuel properties and emissions (Xue et al., 2011).

2.2 Coproduction of biogasoline, biodiesel, green chemicals, and high-value coproducts: towards growing green petroleum

To date thermochemical conversion of biomass is the only way to produce biogasoline, biodiesel, and chemical commodities from the same feedstock. Thermochemical conversion is more commonly used to process woody biomass that has higher lignin content than herbaceous crops because lignin increases biomass recalcitrance to degradation into fermentable sugar monomers for bioconversion. Each thermochemical conversion process breaks down and reforms biomass into small molecular building blocks to yield biosyngas or biocrude (Demirbas, 2009c). Biosyngas and biocrude can then be reformed or upgraded to produce drop-in fuels with fuel properties essentially identical to existing liquid fuels, e.g. high energy content through removal of oxygen. However, there are disadvantages to thermochemical conversion that reduce their economic and product efficiency which include the need for high temperature reactions, loss of energy from biomass to entropy, and catalyst fouling (Carroll & Somerville, 2009). Thermochemical conversion reaction conditions range from 450-950 °C depending on which thermochemical conversion process is being used (Demirbas, 2009c; Ong & Bhatia, 2010). Lower temperature conversion processes usually require catalysts which are eventually fouled by coke formation and require replacement (Kleinert & Barth, 2008b). Additionally, thermochemical conversion favors construction of large reactors to

take advantage of economy of scale and make thermochemical processing economically viable. However, biomass typically is produced in relatively small quantities across large areas leading to a low concentration of feedstocks which favor construction of many smaller biorefineries to minimize transportation costs (Carroll & Somerville, 2009). Combining the quality of liquid fuels from thermochemical conversion with the product specificity, low energy inputs, and scalability of bioconversion will result in higher quality and economically viable renewable liquid fuels.

Lignocellulosic feedstock biomass has traditionally relied on pretreatment, which facilitates hydrolysis of cellulose, hemicellulose and lignin polymers into monomers with supplemented enzymes that cleave the polymers into sugar monomers for fermentation (Figure 2). The concept of consolidated bioprocessing, or simultaneous saccharification and fermentation (SSF), incorporates fermentation and pretreatment of biomass into one process which usually includes production of necessary conversion enzymes by fermentation microbes (Fu et al., 2011). This opens a sizable fraction of plant biomass for conversion to biofuels while simplifying the overall fermentation process and making bioconversion more economic. However, production of liquid fuels ranging from gasoline to diesel and jet fuel in a single fermentation vessel from a biomass source will take the next step in making consolidated bioprocessing truly consolidated. Coproduction of several biofuels has been discussed in other

reports, and in itself is not a novel concept. Hydrogen, methane, and ethanol coproduction from biomass in a biorefinery has been considered and even patented under the name 'Maxifuel Concept' (Ahring & Westermann, 2007). However, coproduction of biogasoline and biodiesel using biological conversion has rarely been considered. The first significant step in biogasoline and biodiesel coproduction from biomass was realized when a process to convert sugars into fatty acid esters using engineered *Escherichia coli* fermentation was coupled with hemicellulases (Steen et al., 2010). Engineering lignocellulosic feedstocks for maximum conversion of biochemicals to suitable biofuels will lead to a new generation of biogasoline and biodiesel production.

3. Engineering plants to make the 'biofuel feedstock buffalo'

The first step in engineering the most suitable biofuel feedstock is the choice of the optimal crop for mass production of biofuels and bioproducts (Yuan et al., 2008b). This can be a difficult choice to make because biomass will be produced around the world and each environment and climate will have varying requirements and adaptation for production. Several species have been considered as dedicated lignocellulosic crops. The major biomass feedstock candidates in the United States are poplar (*Populus spp.*), switchgrass (*Panicum virgatum*), miscanthus (*Miscanthus x giganteus*), sweet sorghum (*Sorghum bicolor*), and microalgae. Oilseed crops are not considered here because it is not clear whether oilseeds will ever be dedicated biofuel crops, though there has

been interest in *Camelina sativa* in recent years (Moser, 2010). Corn stover, the remnants of the corn plant after harvest, has been investigated as a major source of lignocellulosic biomass and represents agricultural wastes. Forest and agricultural residues will likely constitute a large portion of biomass supply for biofuel production (Perlack et al., 2005), but by their nature, will most likely not be genetically engineered and as such will not be discussed here. Perennial dedicated lignocelluosic feedstocks, e.g., switchgrass and miscanthus, have benefits compared to other potential crops that include requirement of fewer energy inputs for stand establishment, good nutrient- and water- use efficiency, and environmental benefits that include soil carbon deposition and ecosystem services (Carroll and Somerville, 2009). Certain tree species such as poplar and willow have been considered for perennial dedicated feedstocks, but they require large amounts of water which will ultimately limit their use (Allison et al., 2010). Algae might ultimately be the best feedstock for biofuel production as it will not compete for arable land and has a large lipid fraction, but production engineering considerations and large capital outlay for production facilities will most likely put algae for biofuel production in the long term (Carlsson, 2009). In the southeastern United States, the dedicated lignocellulosic feedstock of choice will most likely be a perennial grass species such as switchgrass, miscanthus, or energy cane. Of these, switchgrass has received a lot of research attention as a dedicated biofuel crop. The BioEnergy Science Center (BESC) selected

switchgrass and poplar as primary research species, and companies such as Ceres and Metabolix have ongoing research projects that feature switchgrass. Recently, genetic modification of the lignin biosynthetic pathway in switchgrass has successfully produced plants which yielded 38% more ethanol than unmodified switchgrass (Fu et al., 2011). The first public field trials of transgenic switchgrass were started at the University of Tennessee in 2009 which will bring switchgrass a step closer to being a viable dedicated biofuel feedstock by determining regulation needed for transgenic switchgrass in the future. With this in mind, we will focus on engineering approaches to engineering switchgrass and perennial grass feedstocks.

3.1 Bale to barrel: strategies for engineering the perfect petroleum-replacement feedstock

The first way to improve biofuel production yields and economic viability is to convert all the chemicals present in biomass into useful liquid fuels or high-value commodities, and secondly to use all the biomass generated by the dedicated feedstock. Strategies for using and improving all portions of switchgrass biomass will be discussed further in Section 3.2. Efficient engineering of feedstocks and conversion of all biochemicals present in biomass requires in-depth knowledge of metabolites natively present in feedstocks. However, identifying all chemical constituents in biomass is difficult, and moreover, highly variable depending on season, biomass fraction, and extraction techniques employed (Yan et al., 2010;

Yang & Ohlrogge, 2009). Biomass composition and metabolites present in major feedstocks are compiled in Table 2. Data were selected based on a single late season harvest for crops that have otherwise been considered for multiple harvests in a year, e.g. during senescence of switchgrass. As described above in Section 2.2, technical advances in lignocellulosic ethanol production and consolidated bioprocessing have opened up the possibility of using entire aboveground biomass for production of biofuels. While much research has focused on the composition of cellulose and lignin present in switchgrass feedstocks, there has been little compositional analysis of the other portions of switchgrass biomass, namely the 'extractives' fraction.

Switchgrass has an extractives fraction that ranges from 11% -17% of the biomass depending on cultivar, and 13.3 % - 21.0% in different portions of the plant itself (Carroll & Somerville, 2009; Mann et al., 2009). However, the term 'extractives fraction' simply equates to a miscellaneous grouping used to describe the portion of biomass metabolites that is not lignocellulosic biomass and not inorganic components, i.e., ash. Remarkably, few studies on the chemical composition of the extractives fraction have been carried out in switchgrass. This could result from the highly variable nature of the extractive fraction. Switchgrass extractives percentages of dry biomass changes during storage, whether sheltered or outside (Wiselogel et al., 1996). The percentage of total dry biomass the extractives fraction of switchgrass also changes depending

on the extraction procedure itself. One study has shown that 95% ethanol extractives fractions include fatty acids, sterols, triglycerides, sugars, and other metabolites (Yan et al., 2010). However, this report only examined the composition of metabolites from one extraction method. These fractions likely include other secondary metabolites such as isoprenoids and phenylpropanoids, but no studies have been published on secondary metabolites present in switchgrass (typically considered extractives). Further investigation into existing metabolites and metabolic pathways in switchgrass will aid biofuel crop engineering efforts.

Ideally, dedicated plant biomass feedstocks would be processed for biofuel and coproduct production in three steps: 1) simple extraction or distillation to recover a liquid portion of biomass that would be drop-in ready biofuels, 2) the resulting lignocellulosic fraction would be deconstructed and fermented to produce liquid fuels and chemicals for chemical synthesis precursors, 3) residual biomass would then be thermochemically converted to produce hydrocarbons for liquid fuels, coproducts, or heat for generation of electricity (Figure 3). Drop-in ready biofuels in plant biomass that can be extracted or collected through simple distillation will allow for biofuel production in rural and non-industrialized areas. Additionally, converting more of the feedstock biomass to usable products, e.g., combustible metabolites for liquid fuels, that can be simply extracted will increase overall biofuel yields from biomass and increase biorefinery production efficiency

while not requiring more infrastructure or investment. Bioconversion microbes are also subject to plant metabolite toxicity, and so extraction of biomass before fermentation would also remove potentially toxic metabolites from feedstocks. Engineering plant metabolism for the production of simple extraction drop-in ready biofuels will discussed further in
Section 3.1.1.

Modifying lignin content and structure in biomass feedstocks to reduce bioconversion recalcitrance has been one of the primary focuses of biofuel feedstock engineering. While lignin reduces the efficiency of biomass hydrolysis into sugar monomers and subsequently fermentation, significant reduction of lignin content could also lead to lodging, increased susceptibility to pathogens, and decreased drought tolerance of feedstocks. These considerations have led researchers to investigate ways to alter lignin monolignol composition rather than drastically decrease total lignin. Biotechnology approaches to increasing product yields from biomass with specific focus on biofuels will be discussed in Section 3.1.2.

Thermochemical conversion has been usually been considered a competing technology to bioconversion, but in most biorefinery designs both technologies are included (Cherubini & Jungmeier, 2010; Lyko et al., 2009). Bioconversion and separation of products from feedstocks before thermochemical conversion allows for the production of high-value native

coproducts and complex biochemical metabolites that are not feasible for chemical synthesis. Biocrude oil produced from fast pyrolysis is chemically diverse and needs to be catalytically upgraded (Yaman, 2004). Selectively removing large portions of biomass as extractable or fermentable biofuels and coproducts before thermochemical conversion could lead to biocrude with simpler chemistry and higher product specificity.

3.1.1. Plant metabolites for extractable biofuels

Production of switchgrass feedstocks with extractable portions of drop-in biofuels requires two key traits: 1) production of metabolites with suitable fuel properties for combustion in modern gasoline or diesel engines, and 2) storage of metabolites in high concentrations that will not be toxic to plant tissues. Plants produce a range of hydrocarbons that could be used as drop-in ready biofuels and coproducts (Table 3). Plants produce an incredible diversity of C_{10}, C_{15}, and C_{20} isoprenoids, also called terpenoids, which are derived from precursors comprised of isoprene units. The chemical structures of monoterpenes (C_{10}), sesquiterpenes (C_{15}), and diterpenes (C_{20}) are highly diverse and are primarily isoalkanes/enes, and cyclic alkanes/enes. Sesquiterpene synthesis has been shown to occur primarily in cytoplasm, whereas mono- and diterpene synthesis occurs primarily in plastids (Chen et al., 2011). The diverse array of terpenoid isoparaffins and naphthenes produced in plants is reminiscent of the gasoline

and diesel fractions of petroleum (Table 1). Catalytic conversion of the monoterpene pinene yielded a biofuel that had similar net heat of combustion and density as jet fuel, but a higher freezing point (Harvey et al., 2009). A number of terpenoid or terpenoid derived metabolites may have potential as high-value extractible coproducts. Taxol and artemisinin are expensive drugs used in the treatment of cancer and malaria, respectively, which have moved to production through tissue culture or heterologous expression of plant genes in microbes (Kirby & Keasling, 2009). Many mono- and sesquiterpenes are volatile organoleptic compounds responsible for the taste and smell of fruits and flowers, and as such are commodities in the food and cosmetic industries. However, concentrations of terpenoids in most plants range from 1%-2% (Singh, 1999; Singh, 2001; Tholl, 2006; Zheljazkov et al., 2011). As such, engineering larger fractions of biomass fractions should be considered for land plants.

Phenylpropanoids makeup one of the largest pools of plant metabolites and are involved in pathogen defense, ultraviolet light protection, and biosynthesis of lignin (Besseau et al., 2007). Lignin monomers are aromatic rings with oxygenated side group when compared to aromatic compounds found in petroleum fuels (Speight, 1999). Lignin has been widely considered at best, a byproduct of biofuel production that should be burned or converted to liquid fuels by thermochemical conversion (Kleinert & Barth, 2008a), and, at worst, a large fraction of plant biomass that interferes with biofuel product, and, as such, is a

candidate for decreased biosynthesis (Chen & Dixon, 2007; Fu et al., 2011). However, phenolic compounds show promise as precursors for bioplastics (Kleinert & Barth, 2008a), carbon fibers (Baker et al., 2009), and even antioxidants in diesel fuel (Kleinert & Barth, 2008a). Directing phenylpropanoid metabolites for storage in cellular compartments would create an aromatic biofuel fraction that would enhance properties in biogasoline such as lower (net) energy per volume, and cold flow properties in biodiesel (Table 1).

Identification and characterization of novel enzymes involved in unique reactions has been identified as an important line of research that will lead to the development of future biorefinery processes and industrial chemical synthesis (Wohlgemuth, 2009). For biofuels, production of short-chain alkanes from biomass could be the most important as they make up the largest chemical fraction of gasoline and diesel (Table 1). There are two known plants that produce short-chain alkanes: *Pinus jeffreyi* and *Pittosporum resiniferum*. *Pittosporum spp*. produce a range of n-alkanes including heptane, nonane, dodecane, and undecane (John et al., 2008). *P. jeffreyi* only synthesizes n-heptane in tissues and oleoresins; preliminary radiolabeled substrate feeding experiments suggested that n-heptane is formed from octanal precursors coming from fatty acid biosynthesis (Savage et al., 1996). However, no genes involved in either *Pittosporum* or *P. jeffreyi* alkane biosynthesis are known. Recently, identification and recombinant expression of cyanobacterial genes identified as

an acyl-ACP reductase and an aldehyde decarbonylase led to tridecane, pentadecane, and heptadene biosynthesis in *E. coli* (Schirmer et al., 2010). Further investigation into these unique biosynthetic pathways will lead to applications in biofuel property and combustion characteristic enhancement, and extractable drop-in fuels.

Alkane and isoprenoid biofuels could also be enhanced through modifications such as additions of methyl groups to create isoparaffin-like biofuels and which have higher octane values for biogasoline, and better cetane number and cloud points for biodiesel. Methyltransferases have been identified that add methyl groups to a wide range of metabolites including sterols derived from terpene metabolism (Zhou et al., 2008), fatty acids to make FAMEs (Yang et al., 2006), and tocopherol (Bergmüller et al., 2003). Screening of methyltransferases with n-alkanes will be required to determine if any known enzymes will catalyze the formation of isoalkanes. Terpenoid substrates have also been modified in bacterial using both native and plant-derived cytochrome P450 genes (Misawa, 2011). Terpenoid engineering work has focused primarily on mono- and sesquiterpenoid biosynthesis and modification, and as such should provide a fundamental basis for engineering terpenoids in biofuel feedstocks.

Ultimately, the effectiveness of extractable biofuels will depend on the extent to which metabolites can be synthesized and stored in large quantities in

feedstock biomass. Investigation into increasing secondary metabolite concentrations in plants have yielded mixed results. Overexpression of substrate synthesis genes and localization of terpene synthases in non-native cell organelles have showed remarkable increases in specific terpenoid products (Kirby & Keasling, 2009). Investigation into unique species could also shed insight onto mechanisms for increasing production of terpenoids in plants. Trees in the genus *Copaifera* produce a sesquiterpene-rich oleoresin when their trunks are tapped, and can produce anywhere from 0.46 to 1.8 L at a time (Medeiros & Vieira, 2008; Plowden, 2003). However, these researchers noted that production of oleoresin from these trees is unstable and varies with age of tree and environment. Investigation of *Copaifera* saplings grown in greenhouse conditions showed *in planta* sesquiterpene production varied with age and in tissues (Chen et al., 2009). The primary sesquiterpene detected in tissues and oleoresins was β-caryophyllene, a compound that is directly comparable to a bicyclic naphthene. *Copaifera* oleoresins have been reportedly used directly in diesel engines for transportation and production of electricity in remote areas of the Amazon (Calvin, 1983; da Costa et al., 2007). There are several challenges to terpenoid metabolic engineering, namely: cross-talk between terpene synthases and other metabolic pathways that can lead to uncertain product synthesis, and a large diversity but low overall concentration of individual products. The terpenoid biosynthesis pathway is highly complex, and a single terpene synthase can have

multiple products. Cellular localization of terpene synthases can also lead to a change in their products, and has been suggested as a way that terpene biosynthesis has evolved from lower plants to flowering plants (Chen et al., 2011)**.** However, terpene synthases can be engineered by altering amino acids present in the reaction pocket to influence product specificity (Köllner et al., 2006). Using a maize sesquiterpene synthase that natively had two major products, Köllner et al. were successful in creating amino acid mutations that could alter the enzyme activity to one specific major product or the other. Recombinant expression of terpene synthases that have been engineered for product of a single or a few select major products would enable biofuel production from this portion of plant metabolism. Expression of a *tyrosine ammonia-lyase* (*TAL*) from *Rhodobacter sphaeriodes* in *A. thaliana* shunted more carbon into the phenylpropanoid by synthesizing p-coumaric acid from tyrosine (Nishiyama et al., 2010). Down regulation of the *hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase* (*HCT*) gene in *A. thaliana* resulted in increased accumulation of flavonoids and altered lignin profiles (Besseau et al., 2007). However, the accumulated flavonoids interfered with normal auxin transport in transgenic plants resulting in a dwarf phenotype.

Storage of biofuel metabolites *in planta* without toxicity to the cell is the second key step in engineering plant extractable biofuels. Modification of the phenylpropanoid biosynthetic pathway to reduce lignin resulted in dwarf *A.*

thaliana growth from perturbation of auxin transportation through cells (Besseau et al., 2007). *In planta* synthesis of polyhydroxyalkanoates (PHAs) polymers causes a measurable reduction in seed set and growth (Suriyamongkol et al., 2007). Storage of metabolites must be considered when engineering plants for specific applications. The vacuole in plant cells is usually the largest organelle, and stores a host of secondary metabolites generated by the cell during its life cycle. Therefore it is a perfect target for sequestration of novel biofuel metabolites. Vacuolar H^+-ATPase and vacuolar pyrophosphatase transporters are responsible for transport of a large fraction of metabolites into the vacuole (Roytrakul & Verpoorte, 2007). These vacuolar transporters and others are targets for engineering extractable biofuel metabolites accumulation, and will require further investigation in switchgrass and other feedstocks.

3.1.2. Enhancing production of biofuels from lignocellulosic biomass

Lignocellulosic fractions of switchgrass are currently being studied so that they can be modified to reduce recalcitrance to degradation into simple sugars. Several review papers addressing lignin biosynthesis and engineering strategies to modify lignin for enhanced biofuel and coproduct production have been written (Pauly & Keegstra, 2010; Simmons et al., 2010). Both the down-regulation of genes in the lignin biosynthesis pathway, and addition of novel monolignols, such as ferulic acid and coniferyl ferulate, to remodel lignin structure have been

considered to enhance biofuel production from lignocellulosic feedstocks. Successful reports achieving reduced recalcitrance in switchgrass are just beginning to be published (Fu et al., 2011). Investigation of native switchgrass lignin biosynthesis genes have shed light on useful targets for down regulation, and perhaps more importantly, genes that are important in plant defense that should not be knocked down (Escamilla-Treviño et al., 2010). The down-regulation of lignin biosynthesis in *Medicago sativa* led to reduced growth and overexpression of drought tolerance genes and those encoding pathogen defense proteins (Gallego-Giraldo et al., 2011). Free monolignols and other phenylpropanoids may be present in higher concentrations in switchgrass biomass engineered for reduced lignin, and as such technologies to convert the aromatic chemicals or store them for extractable biofuels will need to be developed more fully.

Technology to synthesize useful biofuels and coproducts from sugars and metabolites present in switchgrass biomass is still rather new. Interest in producing better biofuels from biomass has led researchers to develop a myriad of microbial, chemical, and thermochemical conversion techniques (Table 3). Conversion of levulinic acid, derived from acid treatment of hexose sugars, to alkenes using catalysts for use as biogasoline and biodiesel has been demonstrated (Bond et al., 2010; Lange et al., 2010). Therefore, conversion of

sugars into biodiesel could yield production of gasoline and diesel fungible fuels in a single biomass feedstock.

3.2 Improving unused portions of switchgrass biomass

Efficient utilization of all biomass of the dedicated feedstock will enhance biofuel yields and economic viability. However, data on the composition of other portions of switchgrass biomass, specifically seed composition, is scarce. One study found switchgrass seeds contained 62.9% dry weight carbohydrates, 7.4% fiber, 8.6% ash, 8.2% lipid, 12.9% protein (Christian & Lederle, 1984). Switchgrass seed yield has been calculated in South Dakota at 338 and 283 kg ha^{-1} for the cultivars Summer and Sunburst, respectively (Boe, 2007). Harvest of seed biomass in combination with leaf and stem biomass would add an additional source of high quality feedstocks such as starch (carbohydrates), protein for animal feed, and press extractable lipids for production of biofuels and/or coproducts (Table 4). Further investigation is needed to determine whether the production of biofuels from switchgrass seed biomass would outweigh the cost of harvest and processing. However knowledge from transgenic improved oilseed crop seed composition could be applied to switchgrass to increase the breadth of it utility. Overexpression of diacylglycerol acyltransferase (DGAT) in *Brassica napus* changed metabolic flux in the fatty acid biosynthetic pathway and increased overall seed oil accumulation (Weselake et al., 2008). Additionally, overexpression of a maize transcription factor involved in triglyceride

biosynthesis increased seed oil content 46% but reduced seed starch content by 60% (Shen et al., 2010). Shen *et al*. (2010) also reported that expression of *ZmLEC1* increased oil concentrations in seeds, but delayed and decreased seed germination. A similar phenotype in switchgrass could be used as an interesting transgene containment phenotype. Expression of a fungal *DGAT2* gene increases oil in maize seed (Oakes et al., 2011), and coexpression of these genes or orthologs in switchgrass may be a viable strategy for enhanced seed quality for biofuel production. This strategy to enhance biofuel characteristics is not limited to seed biomass. Expression of *DGAT* and *LEC2* from *Arabidopsis thaliana* showed a two-fold increase of triglyceride content in *Nicotiana tabacum* leaf tissues (Andrianov et al., 2010).

After senescence, switchgrass leaves still have 10.6 μg mg^{-1} fatty acids in extractable fractions (Yang & Ohlrogge, 2009). Increased biosynthesis and storage of fatty acids in leaf tissues could be achieved as discussed above for seed tissues. Additionally, direct synthesis of fatty acid ethyl esters (FAEEs) from glucose has been achieved in *E. coli* (Steen et al., 2010). Expression of a recombinant thioesterase for production of free fatty acids was coupled with expression of a recombinant pyruvate decarboxylase and alcohol dehydrogenase to produce FAEEs. Furthermore, hemicellulose excretion was engineered into the FAEE producing strains to liberate xylose from biomass which further enhanced FAEE production. The FAEE composition could be controlled by

expressing thioesterases with different substrate specificity. While this strategy may or may not be feasible to use directly in switchgrass to produce FAEEs, increasing fatty acid content in biomass would most likely increase the efficiency of FAEE production during fermentation using these engineered strains of *E. coli*.

4. Conclusions

Current visions of biofuel production that would convert only a portion of biomass to liquid fuels may be considered unsustainable. Converting latent metabolites into valuable coproducts and biofuels will lead to not only a more robust biobased products industry, but reduced reliance on petroleum feedstocks for chemical synthesis and liquid fuels. Additionally, engineering production and storage of biofuel metabolites that are extractable from biomass using simple techniques such as distillation or cold pressing will enable liquid fuel production, and perhaps even isolation of coproducts in rural or undeveloped areas. Production of biofuels in rural areas and farmlands of the United States will help to reduce costs associated with transportation of biomass to biorefineries, and lend more incentives to farmers to grow dedicated feedstock biomass. Any sustainable biorefinery concept will reach far beyond simple liquid fuels such as ethanol.

To create plant-extrable biofuels, we need a greater understanding of how biochemical structures combust in engines and we must be able to manipulate unique metabolite biosynthetic pathways, such that for short-chain alkane

biosynthesis. Additionally, genes and engineering strategies useful in transporting and storing large amounts of metabolites in plant organelles will be needed to avoid toxicity issues. Once these technologies are developed, they will then be applicable to any biomass feedstock for biofuel production being considered across the world. Use of biotechnology for optimization of biofuel feedstocks is critical in replacing petroleum as a natural resource. As such, strategies for transgene biocontainment and mitigation of gene flow and research to help inform and guide proper regulation of transgenic feedstocks are crucial in developing the biofuel industries' infrastructure (Kausch et al., 2010).

Acknowledgements.

We wish to acknowledge funding from UT-Batelle and the UT Science Alliance as well as funding from the BioEnergy Science Center (BESC). BESC is a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the US DOE Office of Science.

APPENDIX

Figure 1. Chemical conversion versus *in-planta* biochemical conversion of biofuel metabolites.

A) Conventional chemical methyl esterification conversion of fatty acids for biodiesel production whether free or bound to glycerol. Methyl esterification reduces biodiesel viscosity while creating glycerol, alkaline catalyst waste, and waste water that needs to be processed. B) Conversion of metabolites with biocatalysts, bioconversion, in planta would produce biofuels that are extractable from plant biomass directly. The resulting metabolic byproducts such as glycerol would reenter plant metabolism for recycling and reduce processing waste.

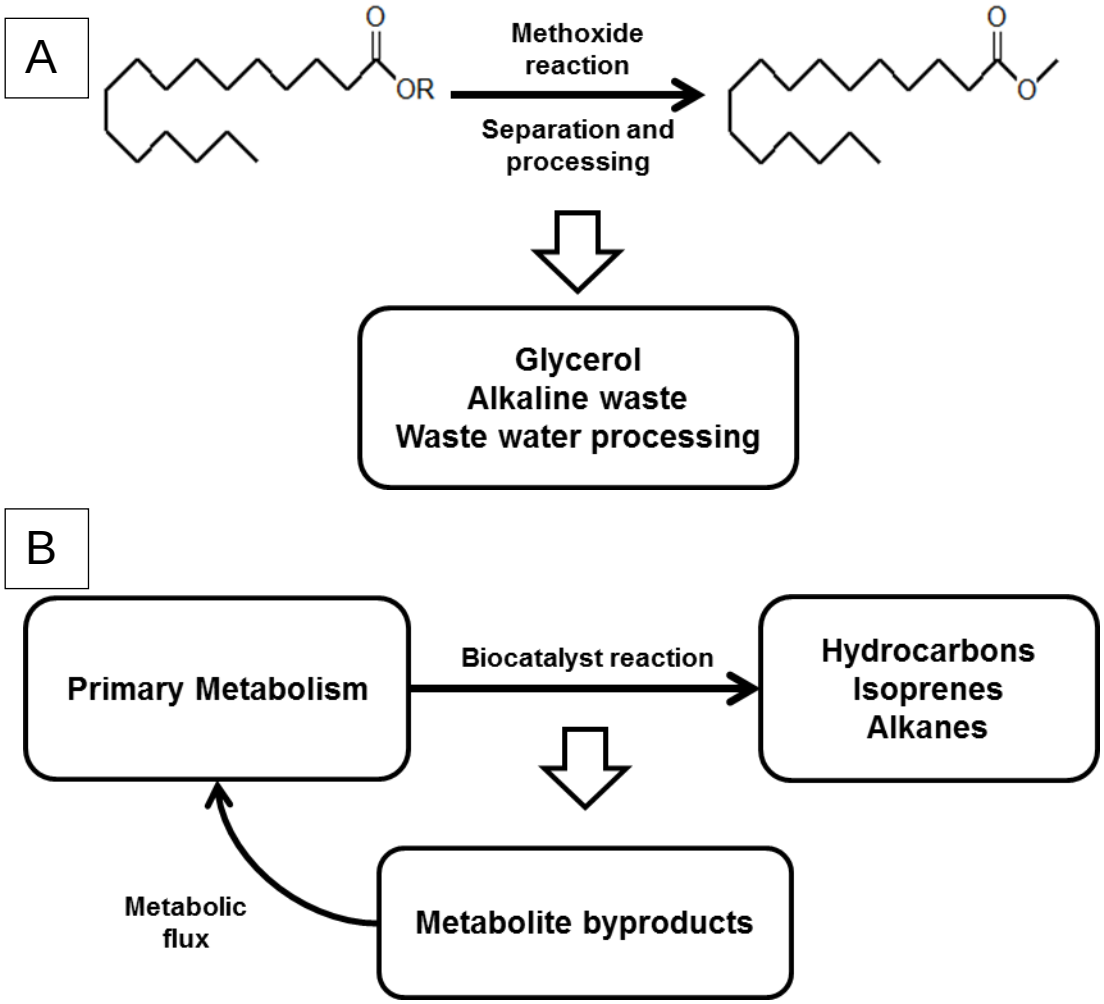
A
O
OR
Methoxide
reaction
Separation and
processing
O
O
Glycerol
Alkaline waste
Waste water processing
B
Primary Metabolism
Biocatalyst reaction
Hydrocarbons
Isoprenes
Alkanes
Metabolic
flux
Metabolite byproducts

Figure 1.

Figure 2. Current (solid lines) and future (dotted lines) conversion techniques for production of biofuels and coproducts from fractions of plant biomass.

Cellulose and hemicellulose are currently converted to sugars through pretreatment and enzyme degradation which are then fermented to produce alcohol biofuels. Simultaneous saccharafication and fermentation (SSF) techniques will allow for consolidated bioprocessing to reduce inefficiencies resulting from multistep processing. Biodiesel is generated from fatty acid chemical conversion. Lipases and other biocatalysts are being developed to enhance the esterification reaction, and conversion of glycerol into useful biofuel and coproduct chemicals. Novel biocatalysts have also been developed to produce biodiesel fatty acid ethyl esters (FAEEs) from sugars. Lignin is currently hermochemically converted into biocrude, syngas, or electricity/heat. Production of commodities such as carbon fibers and bioplastics are being developed from lignin fractions. Extractable hydrocarbon metabolites such as isoprenoids and alkanes are being considered for biofuels and coproducts. Monolignols would also be suitable as an aromatic biofuel fraction.

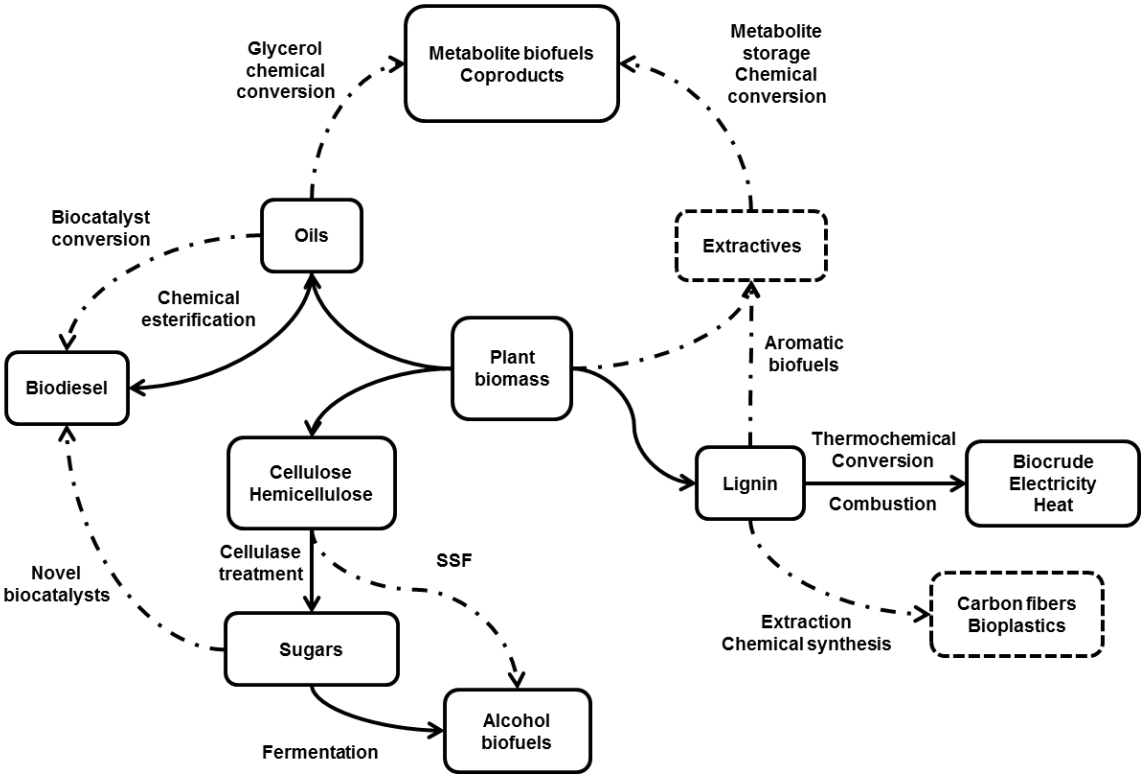
Metabolite biofuels
Coproducts
Glycerol
chemical
conversion
Metabolite
storage
Chemical
conversion
Biocatalyst
conversion
Oils
Extractives
Chemical
esterification
Plant
biomass
Aromatic
biofuels
Biodiesel
Thermochemical
Conversion
Cellulose
Hemicellulose
Lignin
Biocrude
Electricity
Heat
Combustion
Cellulase
treatment
SSF
Novel
biocatalysts
Carbon fibers
Bioplastics
Extraction
Chemical synthesis
Sugars
Alcohol
biofuels
Fermentation

Figure 2.

Figure 3. Systematic processing of the ideal biofuel plant feedstock.

Simple distillation or extraction of biomass would yield in planta biofuels and coproducts such as bioplastics (PHAs, PHBs), pharmaceuticals (artemisinin, taxol), and food and cosmetic additives (limonene, geraniol, citral). Biofuel feedstocks with this characteristic would allow for production of biofuels and biobased products in rural and areas without biorefinery capabilities, and help to offset costs associated with transportation of biomass to biorefineries. Lignocellulosic conversion would occur in areas with a biorefinery infrastructure. This segment of biofuel processing would allow for more complete conversion of biomass and produce a host of coproducts such as green chemical precursors that require either microbial fermentation or further processing to develop valuable coproducts. Examples include ethanol, butanol, carbon fibers, succinic acid, lactic acid, and valeric acid for biodiesel. The residue remainder of the biomass that cannot be bioconverted will be processed using thermochemical conversion to generate syngas, biocrude, and/or combusted to produce heat or electricity. This will reduce the volume of biomass that has to be converted at high temperatures, and reduce the chemical complexity of biocrude generated from fast pyrolysis.

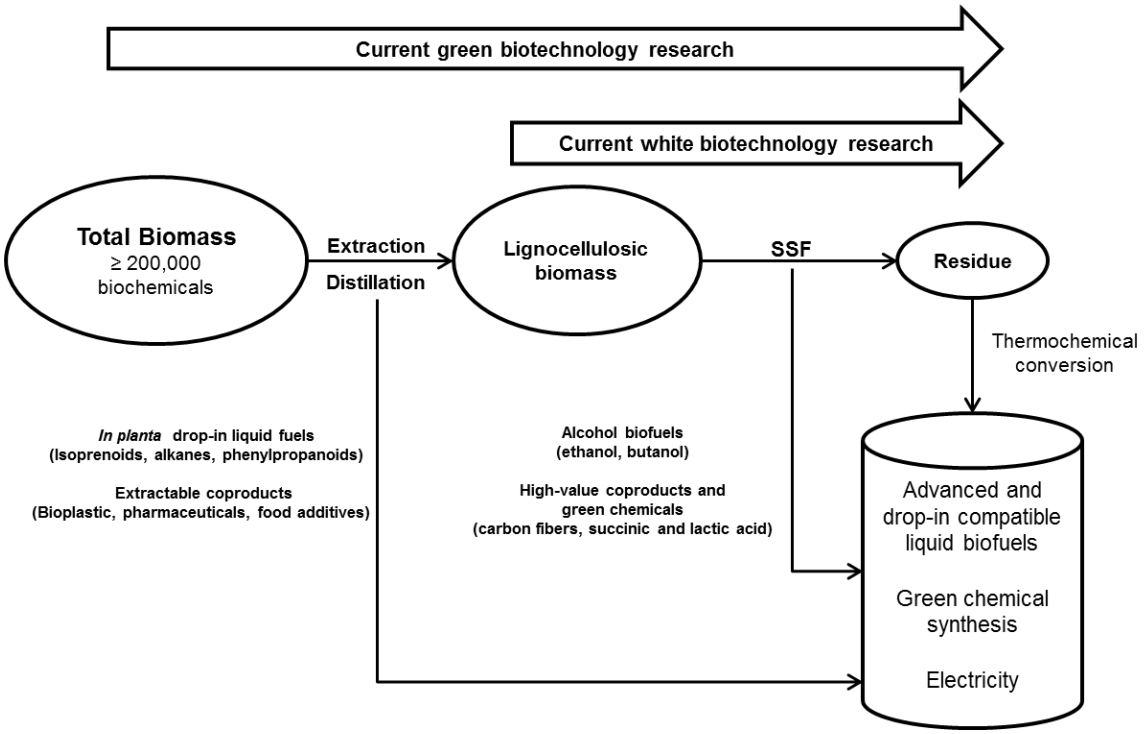
Current green biotechnology research
Current white biotechnology research
Total Biomass
≥ 200,000
biochemicals
Extraction
Distillation
Lignocellulosic
biomass
SSF
Residue
Thermochemical
conversion
In planta drop-in liquid fuels
(Isoprenoids, alkanes, phenylpropanoids)
Extractable coproducts
(Bioplastic, pharmaceuticals, food additives)
Alcohol biofuels
(ethanol, butanol)
High-value coproducts and
green chemicals
(carbon fibers, succinic and lactic acid)
Advanced and
drop-in compatible
liquid biofuels
Green chemical
synthesis
Electricity

Figure 3.

Table 1. Chemical class influence on petroleum distillates' fuel and physical properties and biochemical alternatives.

Petroleum Distillate (range °C)	Fuel Property	Paraffin	Iso-paraffin	Olefin	Naphthene	Aromatic	Oxygenates[a]	Saturated Fatty Acid Esters	References
Gasoline (30-200) C_4-C_{12} 50-60% paraffin	Octane number	-	+	±	- , ±	NR	+		(Pitz, Cernansky, 2007, Speight, 1999)
	Lower heating value (kJ L^{-1})	-	±	±	+	+	-		(Pitz, Cernansky, 2007, Speight, 1999)
	Volatility	+	+	±	±	±	- , ±		(Pitz, Cernansky, 2007, Speight, 1999)
Kerosene (140-320) C_{10}-C_{16}	Smoke emission	+	±	-	±	-	NR	NR	(Dagaut and Cathonnet, 2006, Speight, 1999)
Diesel (126-258) C_8-C_{18}	Cetane number	+	- , +	-	±	-	-	+	(Bacha et al. , 2007, Benjumea, Agudelo, 2010, Li et al. , 2005)
	Cloud point	-	+	±	+	+	+	-	(Bacha, Freel, 2007, Benjumea, Agudelo, 2010, Li, Zhen, 2005)
	Lower heating value (kJ L^{-1})	-	NR	±	+	+	-	NR	(Bacha, Freel, 2007, Li, Zhen, 2005)
	Freezing point[b]	±	+	± , +	+	+	+	- , ±	(Bacha, Freel, 2007, Dagaut and Cathonnet, 2006)

[a] oxygenates denote ethanol, butanol
+ denotes a positive affect
± denotes a neglible or mixed affect
- denotes a negative affect
NR – not reported

Table 2. Lignocellulosic biofuel feedstock chemical composition in percent dry weight.

Biomass Fraction	Fraction Metabolite(s)	Biomass Composition of Lignocellulosic Feedstocks			
		***Panicum virgatum* Cv Alamo (% dry weight)**	***Miscanthus x giganteus* (% dry weight)**	***Zea mays* Stover (% dry weight)**	***Populus* (% dry weight)**
Cellulose		33.48 - 33.75	46.93 - 49.41	37.12 - 39.4	42.2 - 48.95
Hemicellulose		26.1 - 27.04	29.68 - 32.26	24.18	16.6 - 23.24
	Glucose	37.0	50.47	36.8	39.23
	Xylose	20.42, 28.8	21.68	22.2	13.07
	Arabinose	2.75, 3.7	2.78	5.5	0.89
	Galactose	0.92, 1.3	0.35	2.9	0.88
	Mannose	0.29	NR	NR	1.81
	Uronic acid	NR	NR	NR	4.31
Lignin		16.8 - 17.35, 22.7	11.97-13.24	23.1	21.4 - 29.1
Extractives		11.0[a], 15.50[a], 18.4[b], 10.2[c]	1.13[d], 14.03	3.9[a], 5.61	2.4[c], 6.89
	Fatty acids	1.54[a], 5.5[d]	3.93 – 4.53	NR	NR
	Sterols	1.0[a]	2.75 – 9.49	NR	NR
	Trehalose	2.2[a]	NR	NR	NR
Ash (inorganic elements)		5.76	3.2	10.06	2.03
References		(Carroll and Somerville, 2009, Sannigrahi et al. , 2010, Yan, Hu, 2010)	(Allison, Robbins, 2010, Le Ngoc Huyen et al. , 2010, Villaverde et al. , 2009)	(Carroll and Somerville, 2009, Hu et al. , 2010, Sannigrahi, Ragauskas, 2010)	(Carroll and Somerville, 2009, Sannigrahi, Ragauskas, 2010)

[a] ethanol extraction
[b] hot water
[c] alcohol-benzene extraction
[d] dichloromethane extraction
[e] toluene extraction
NR – not reported

Table 3. Biofuels and coproducts derived from biomass metabolites and conversion techniques used.

Metabolite Precursor	Chemicals Generated	Fuel Chemical Class	Products	Method	Refs
Glucose	Ethanol	Oxygenated alkane	Biogasoline Biodiesel	Microbial fermentation	(Agarwal, 2007)
	Valeric esters C_8-C_{16} alkenes	Olefins	Biogasoline Biodiesel	Chemical conversion	(Bond, Alonso, 2010, Lange, Price, 2010)
	Fatty acid esters	Oxygenated paraffin/olefin	Biodiesel	Microbial fermentation	(Steen, Kang, 2010)
	Lactic acid	For green chemical synthesis	Solvents, resins, antifreeze	Microbial fermentation	(Octave and Thomas, 2009)
	Succinic acid	For green chemical synthesis	Bioplastics, paints, food additive	Microbial fermentation	(Lyko, Deerberg, 2009)
Xylose	Ethanol	Oxygenated alkane	Biogasoline Biodiesel	Microbial fermentation	
Monolignols	C8-C12 alkanes	paraffins and aromatics	Biogasoline Biodiesel	Thermochemical conversion	(Kleinert and Barth, 2008)
	Direct coproduct and biofuel	Aromatics	Biogasoline Biodiesel	Plant biomass	Proposed
Fatty acids	Fatty acid esters	Paraffin or olefin	biodiesel, lubricants, surfactants, food additives	Plant biomass Microbial fermentation	(Agarwal, 2007, Demirbas, 2009b)
	Short-chain alkanes	Paraffins	Biogasoline Biodiesel	Plant biomass Microbial fermentation	Proposed
	Glycerol	Oxygenated paraffin	Adhesives, polymers, plasticizers, ethanol, succinate, hydrogen, butanol, bioplastics	Microbial fermentation Chemical synthesis Catalytic cracking	(García, Laca, 2008, Zhang, Shanmugam, 2010)
Trehalose	Direct coproduct	Direct coproduct	Food and pharmaceutical coproduct	Plant biomass	(Börnke and Broer, 2010)
Terpenoids (isoprenoids)	Direct coproduct and biofuel	paraffin, olefin, isoparaffin	Biogasoline, biodiesel, jet biofuel, pharmaceuticals, food additives	Plant biomass	Proposed, (Harvey, Wright, 2009, Lee, Chou, 2008)

Table 4. Calculated seed composition of switchgrass and yield per hectacre of each component.

Fraction	Calculated content of *Panicum virgatum* seed (kg ha^{-1})	
	Cv Summer	**Cv Sunburst**
Carbohydrates	212.60	178.01
Fiber	25.01	20.94
Ash	29.07	24.34
Lipid	27.72	23.21
Protein	43.60	36.51

CHAPTER I

DIESEL TREES

A version of this chapter was originally published by Blake L. Joyce and C. Neal Stewart:

B.L. Joyce, H. Al-Ahmad, F. Chen, and C.N. Stewart. 2011.Chapter 24. Diesel trees. Pp. 615-625 in C. Kole, C.P. Joshi, D.R. Shonnard, R. Tressider (eds); Handbook of Bioenergy Crop Plants, Taylor and Francis.

B.L. Joyce wrote the chapter. H. Al-Ahmad, F. Chen, and C.N. Stewart provided paragraphs and reviewed the chapter.

Abstract

When trees in the genus *Copaifera* are tapped they produce an oleoresin, which is rich in sesquiterpenes. These oleoresins are used in cosmetics, the food industry, as herbal medicine in South America, and as fuel to power diesel engines. Melvin Calvin originally highlighted these trees in the 1980s as a potential source of plant-derived hydrocarbons, but since this time they have not received formal investigation into these properties. Collection of *Copaifera* oleoresins has been suggested as a way to supplement native people's incomes in the Amazon River Basin without the need to clear cut the forest. This practice and traditional forestry production of the oleoresin for the biodiesel market seems unfeasible due the long generation time, low yields, and the tropical nature of the trees that limits their range. Despite this fact, these oleoresins represent an interesting chemistry new to biodiesels as they are comprised mostly of cyclic

hydrocarbons from the isoprenoid pathway. While the chemical constituents of the oleoresins are well documented, the biochemical pathway and molecular biology of these plants has not yet been studied. More understanding of the *Copaifera* isoprenoid synthesis pathway could lead to use of these genes in temperate oilseed crops to improve their suitability for use as biodiesel. Additionally, scientific investigation into the oleoresin fuel properties is needed to determine whether the oleoresins will function as biodiesel in modern engines and their potential use in blending with other biodiesels.

1 Introduction

The natural history of diesel trees has a long interaction with humans in the realm of economic botany. Trees in the genus *Copaifera* belong to the subfamily Caesalpinioideae in the family Fabaceae. In total, there are more than 70 species of *Copaifera* distributed throughout the world with at least 30 species found in South and Central America, primarily in Brazil, four species in Africa, and one in Malaysia and the Pacific Islands (Dwyer, 1951; Dwyer, 1954; Hou, 1994). The first species in the genus *Copaifera* was described by George Marcgraf and Willem Pies in 1628, but no formal species name was ascribed to the plant, though later it was deemed *Copaifera martii* based on the description by Veiga Junior and Pinto (2002). Oleoresin from a *Copaifera* tree was listed as a drug in the London Pharmacopoeia in 1677 and to the United States Pharmacopoeia in 1820, and Linnaeus first described the genus *Copaifera* in 1762 (Plowden, 2004).

Later, more descriptions of *Copaifera* species were completed by Hayne in 1825 and Bentham in 1876 (Dwyer, 1951). The current taxonomy of the genus has been largely defined by Dwyer and Léonard who resolved the differences between the genera *Copaifera* and *Guibourtia* and further developed both the New World and African species descriptions in the early 1950s (Dwyer, 1951; Dwyer, 1954; Léonard, 1949; Léonard, 1950). Some species are still difficult to identify in the field, even to specialists, because of an incomplete taxonomy and esoteric species differences that rely on intricate flower morphology and other transient characteristics than can be difficult to ascertain or collect, compared to leaf morphology. To complicate this situation further, *Copaifera* trees have been known to only flower once every two or three years in Amazônia (Alencar, 1982; Pedroni et al., 2002). Furthermore, most up-to-date references on *Copaifera* taxonomy are in Portuguese which hampers the interchange of information amongst the mainstream of scientists.

Copaifera species found in Africa are biochemically distinct from those discussed above because they produce resins that harden into a solid copal, which fossilizes into amber, whereas New World species produce a liquid oleoresin owing to the higher concentrations of sesquiterpenes (Langenheim, 1973). Oleoresin which results from tapping *Copaifera* trees was listed as a drug in the London Pharmacopoeia in 1677 and to the United States Pharmacopoeia in 1820. In Brazil, the oleoresin produced by *Copaifera* trees has been used by

native people as a local medicine for healing wounds, an antiseptic, to relieve pain, and a host of skin, respiratory, and urinary ailments (Plowden, 2004). They have also been used for more esoteric purposes such as a snake bite remedy, aphrodisiac, removal of intestinal parasites, and as a contraceptive.

More recently, several scientific studies have verified the medicinal properties of various *Copaifera* oleoresin fractions for anti-inflammatory activity (Veiga et al., 2006), stomach ulcers and intestinal damage mitigation (Paiva et al., 2004; Paiva et al., 1998), anticancer activity (Gomes et al., 2008; Lima et al., 2003; Ohsaki et al., 1994), reduced pain sensitivity (Gomes et al., 2007), and increased rate of wound healing (Paiva et al., 2002). The oleoresin and oils of *Copaifera* species have also been used in varnishes and lacquers, as lumber, cosmetic products, and tracing paper (Lima & Pio, 2007; Plowden, 2004).

Additionally, in 1980 the Nobel Prize winning chemist Melvin Calvin noted the oleoresin from *Copaifera* trees was being used as diesel fuel directly from the tree with minimal processing (Calvin, 1980). Calvin began his search for plants that could produce liquid fuels to be used directly in engines after the 1973 oil embargo. He later wrote two more papers in 1983 and 1986 on the potential for production of hydrocarbon fuels from living plants, the issue of global warming, and the pressing need to address United States' foreign oil dependency which now, some 20 years later, seems almost prophetic. Plantations of *Copaifera* trees were established in Manaus, Brazil to test the viability of biofuel production

in the 1980s, but were later shifted to focus on production of timber and the oleoresin for pharmaceutical and industrial purposes (Plowden, 2004). The direct reasons for this shift were undoubtedly economic when diesel fuel returned to being relatively cheap.

2 Chemicals Present in *Copaifera* Oleoresins

Copaifera oleoresins, in general, are unique because they contain a greater fraction of sesquiterpenes compared with mono- and diterpenes. In *Copaifera multijuga*, roughly 80% of the oleoresin is comprised of sesquiterpenes, whereas in *Copaifera guianensis* only about 44% of the oleoresin was comprised of sesquiterpenes (Cascon & Gilbert, 2000). These authors also noted that the majority ratio of diterpene acids and sesquiterpenes oscillated back and forth throughout the growing season in *Copaifera duckei*.

A wealth of original articles and review papers has focused on describing terpene biosynthesis. As such, only a brief description of the major terpene constituent characteristics and their biosynthesis in relation to conifer and *Copaifera* structures will be attempted here.

In short, isoprene units, the building blocks of terpenoids, are derived from either the mevalonic acid (MVA) pathway present in the cytosol of cells, or the 2-C-methylerythritol-4-phosphate (MEP) pathway, also known as the non-mevalonate pathway, which occurs in plastids (Lichtenthaler, 1999). Condensation of isopentenyl diphosphate and its isomer dimethylallyl

diphosphate, the products of the MVA and MEP pathways, leads to the formation of geranyl pyrophosphate (GPP), farensyl pyrophosphate (FPP), or geranyl geranyl pyrophosphate (GGPP), which are the common precursors for mono-, sesqui-, and diterpenes, respectively. These three intermediates are catalyzed to form mono-, di-, and sesquiterpenes by the action of terpene synthases (TPSs). Individual TPSs can generate either one product or multiple of products which, in turn, can be linear or cyclic. Mono- and diterpenes are thought to be derived primarily from isoprenes made in the plastid through the MEP pathway, while sesquiterpenes are derived from isoprenes made in the cytosol where the MVA pathway occurs. Movement of intermediates between these two pathways has been demonstrated in plants (Cheng et al., 2007).

The chemical compounds present in *Copaifera* oleoresin varies not only with tissue type (Chen et al., 2009; Gramosa & Silveira, 2005), but also seasonally (Cascon & Gilbert, 2000; Zoghbi et al., 2007), and amongst species (Veiga Junior et al., 2007). Therefore, any future genomics-based characterization of *Copaifera* trees must be coupled with close biochemical analysis to correctly match major compounds present in each tissue at the time of sampling. Identification of particular chemicals responsible for the pharmaceutical effects of *Copaifera* oleoresins will be necessary in the future because high chemical variability within samples, seasons, and species will inherently affect the effectiveness, dosage, and safety for patients.

3 Biosynthesis of *Copaifera* Oleoresins: What Conifers Can Teach Us

Not much is known about the biosynthesis of *Copaifera* oleoresins as the majority of studies have been focused on traditional ecology and forestry of the genus. Conifer resins, however, have been thoroughly studied for over 40 years. These oleoresins are essentially made of the same basic constituents as *Copaifera* oleoresin: mono-, di-, and sesquiterpenes. Conifer oleoresins usually have equal part of mono- and diterpene compounds, and with lower concentrations of sesquiterpenes (Martin et al., 2002).

Monoterpenes are volatile components found in oleoresins. Monoterpene synthases have been extracted from woody stems of 10 conifer species, and their activities measured (Lewinsohn et al., 1991). Species with resin ducts showed the highest levels of monoterpene cyclase activity from wood extracts, suggesting that monoterpene synthesis for oleoresins occurs in epithethial cells surrounding the resin ducts. Diterpenoids themselves are not typically found in conifer oleoresins in large quantities. Instead, modifications such as hydroxylation and oxidation occur, so the alcohol, aldehyde, and predominantly, acid products, are present (Keeling & Bohlmann, 2006b). These modified diterpene products harden the resin and form rosin after the volatile constituents evaporate. Sesquiterpenes, like monoterpenes, are volatile and are major constituents of *Copaifera* oleoresin. The three major chemical constituents,

based on percentage, of oleoresin from different species are presented in Table 5. Although the percentages vary, β-caryophyllene is the major sesquiterpene product of oleoresins throughout *Copaifera* species that have been studied to date. Other than these major three sesquiterpenes in each species, there is a great diversity of terpenoids produced in the oleoresin. Nuclear magnetic resonance studies have found previously undescribed diterpenes (Monti et al., 1999; Monti et al., 1996) that seem to be unique in biology.

Conifers produce a myriad of specialized tissues to store and secrete oleoresins that range from simple resin blisters to intricate networks of resin ducts (Martin et al., 2002). *Copaifera* trees form resin ducts throughout their xylem tissue that can easily be seen in cross-sections (Calvin, 1980). *Copaifera*, *Hymenaea*, and *Daniella* resin ducts display many structural similarities (Langenheim, 2003).

Conifer oleoresins accumulate in resin ducts throughout their lifetimes, but a local response can also be induced during mechanical damage, herbivory, or even fungal inoculation. This response activates epithelial cells in resin ducts, signals for formation of special traumatic resin ducts in stem xylem tissue, and induces diterpene biosynthesis gene transcripts (Keeling & Bohlmann, 2006b). Methyl jasmonate can also induce this response (Zhou et al., 2008). Oleoresin production can also be induced in *Copaifera* species. Younger trees that do not produce oleoresin on the first attempt have been known to produce a small

amount on a second tapping, putatively through induction by mechanical damage (Medeiros & Vieira, 2008; Plowden, 2003). Medeiros and Vieira (2008) were also able to draw a weak correlation between trees with termite infestations and production of oleoresin suggesting that insect damage can induce production of oleoresins.

The cellular mechanisms involved in transport, storage, and secretion of oleoresin constituents against the concentration gradient present in resin ducts are not well understood (Keeling & Bohlmann, 2006b; Langenheim, 2003). Synthesis of terpenoids present in conifer oleoresins typically involves terpene synthases (TPS) and cytochrome P450 oxygenases (P450). A conifer diterpene synthase (PtTPS-LAS) and the first diterpene P450 (PtAO) have been localized to plastids and the endoplasmic reticulum (ER) using a green fluorescent protein in tobacco leaf cells (Ro & Bohlmann, 2006). Based on the lack of accumulation of diterpenes in cells, these authors suggest that a transport mechanism must be in place to move the diterpenes into the ER or cytosol of cells.

Although *Copaifera* oleoresin exudes from resin ducts during tapping, no experiments have confirmed which tissues are responsible for production of chemical constituents in the oleoresin. Calvin (1980) hypothesized that the constituents in *Copaifera* oleoresin must be synthesized in the canopy of the tree and seep down through the resin ducts. In Norway spruce, diterpene synthases have been localized to epithelial cells surrounding resin ducts using protein-

specific antibodies (Keeling & Bohlmann, 2006b). Recently, we have found that the sesquiterpenes present in *C. officinalis* oleoresin in leaves and stem tissue of seedlings as well as leaves, stems, and roots of two year old saplings (Chen et al., 2009). The presence of sesquiterpenes in different tissues at different ages could indicate transport or changes in regulation of TPS gene scripts signaled by development. In addition, the terpenes detected in oleoresins also appear in other tissues such as seeds (Gramosa & Silveira, 2005). The seeds also have different sesquiterpenes that are not seen in oleoresins such as γ-muurolene perhaps suggesting that different terpene synthases function in different tissues.

4 Biological Functions of Oleoresin

The principal chemical constituents of *Copaifera* oleoresin are terpenoids. Therefore, understanding the biological/ecological roles of terpenoids will allow us to understand the roles of *Copaifera* oleoresin. Terpenoids are the largest class of secondary metabolites produced in the plant kingdom. Approximately 50,000 of these have been structurally identified (McCaskill & Croteau, 1997). This diverse group of plant metabolites is important for many aspects of plant biology and ecology (Tholl, 2006; Yuan et al., 2009). For instance, some terpenoids function in plant defenses against herbivores and microbial pathogens (Gershenzon & Croteau, 1991). Other terpenoids produced by flowers as volatiles are involved in attracting insect pollinators for plant cross-pollination (Odell et al., 1999). Some volatile terpenoids are emitted from herbivore-

damaged plants and function as cues to attract natural enemies of the feeding herbivores (Yuan et al., 2008a). *Copaifera* oleoresin is generally believed to be involved in plant defenses that can be mainly attributed to terpenoids. Depending on the mechanisms of production, oleoresins may act in either constitutive defense or induced defense, or both. *Copaifera* oleoresin could be toxic to herbivorous insects, bacteria, or fungi. Because of high volatility, the terpenoids in *Copaifera* oleoresin may be released from the tree as infochemicals, which can deter potential insect pests. Oleoresin may also flow out of the wound to physically push the invading insects out of the entry wound or entomb them so the insects cannot cause further damage. The wound caused by insect herbivory can be a natural site for invasion of microbial pathogens which would need to be defended against. *Copaifera* oleoresin and its constituents have been documented to have antimicrobial and antifungal activity (Braga et al., 1998; Howard et al., 1988). *Copaifera* oleoresin produced upon insects feeding may therefore prevent further damage caused by pathogens.

Studies on *C. langsdorffii* populations have showed that seedlings have a higher sesquiterpene concentration than their parent trees (Macedo & Langenheim, 1989c). Additionally, there was a 48% mortality rate of first generation oecophorid larvae and pupae when they were reared on seedling leaves, but no mortality seen on oecophorids reared on parent leaves. The oecophorids that survived feeding on seedling leaves also exhibited a

significantly lower weight gain than those feeding on parent leaves. Seedlings had twice as much caryophyllene, the major sesquiterpene present in most species' oleoresin, in leaves when compared to their parents. It is still unknown how tapping *Copaifera* trees for oleoresin affects tree health in the long term. Initial tapping, or even multiple tapings, could harm the tree by removing a source of chemical defense against pathogens and insects and must be considered in future studies.

5 Oleoresin Production Ecology

Extractive collection of the oleoresin from wild populations of *Copaifera* trees has long been touted as a means to supplement income for native people in rural and forest areas instead of participating in the destructive practices such as slash-and-burn agriculture and timbering. The viability of this practice, however, has been called into question because of intermittent presence of oleoresin amongst individual trees, low yields of oleoresin per tree, along with reduced and questionable secondary harvests of trees that produce oleoresins on the first tapping (Medeiros & Vieira, 2008; Plowden, 2003). Sustainable production of quality oleoresin for medicine and other uses has many problems that must be considered. First, a management system must be found that will maximize production and minimize impact on the forest where harvest is occurring must be described (Rigamonte-Azevedo et al., 2004). This matter is complicated by the fact that the genus *Copaifera* is made up of many species that can produce

useful oleoresin, and each of these species will naturally respond differently to each possible management strategy. In addition, anecdotal evidence suggested that each tree could produce between 20-30 liters of oleoresin from one drill hole every 6 months (Calvin, 1980); however, these stories seem to be more myth than fact.

In a study of 43 *C. multijuga* individuals in the Adolpho Ducke Forest Reserve in Manaus, Brazil about half produced some volume of oleoresin during three tapings (Medeiros & Vieira, 2008). Six of these individuals never produced oleoresin at all. On average, trees with a diameter at breast height (dbh) $\geq$ 41 cm produced 1.8 L of oleoresin per tree on the first tapping and 0.5 L during the second tapping one year later. Trees of 30 and 41 cm dbh produced an average of 0.13 L during the first tapping and 0.16 L during the second tapping. Plowden (2003) studied *Copaifera* oleoresin production from three different species in Pará, Brazil on the Alto Rio Guamá Indigenous Reserve. Trees 55 to 65 cm dbh yielded the most oleoresin averaging 459 ml after two holes were drilled.

Some of the highest recorded average yields per tree were seen in the southwestern Brazilian Amazon in *C. reticulata* and *C. paupera* trees with 2.92 L and 1.33 L, respectively (Rigamonte-Azevedo et al., 2006). These numbers, however, were averages amongst oleoresin producing individuals only. Only 27% of *C. reticulata* trees and 80% of *C. paupera* trees produced oleoresin. It is not clear whether the lack of uniformity in oleoresin production stems from tapping

methodology or whether the oleoresin itself is just not produced constitutively in all trees. Significant variation, both natural and in response to herbivory, in chemical composition of *Copaifera langsdorffii* leaves has been noted (Macedo & Langenheim, 1989a; Macedo & Langenheim, 1989b; Macedo & Langenheim, 1989c). This variation, compounded by variation in climate, nutrient availability, and other factors, could also cause sporadic oleoresin production and therefore explain the variation seen in oleoresin collection.

Multiple harvests have also been considered to increase oleoresin yields. Cascon and Gilbert (2000) tapped 300 to 550 ml of oleoresin from a single *C. duckei* tree ten consecutive times at four month intervals, but never depleted the tree of oleoresin at any point. However, it is impossible to determine how much oleoresin collected at each interval was residual material that had been stored in the tree and how much had been synthesized and replaced between tappings. Most studies suggest that primary tapping accesses oleoresin from accumulations in heartwood that have built up over long periods of time (Plowden 2004), and, therefore, would not quickly regenerate for a secondary major harvest as Calvin originally hoped. The density of trees also ranges from 0.1 to 2.0 ha^{-1} depending on location and forest type (Rigamonte-Azevedo et al., 2004).

It is unknown how phenology plays a role in oleoresin production. As mentioned before, the chemical composition of the *Copaifera* oleoresins change throughout the year, but no specific cause has been identified as the factor

driving this change. Phenology studies of *Copaifera* species are rare and focus more on the flowering, seed set, and leafing patterns (Pedroni et al., 2002). The majority of these types of studies have been in *Copaifera langsdorffii*, a species native to the southern parts of Brazil. Oleoresin collection for commercial products, however, occurs more commonly in the northern half of Brazil and South America. From our experience, the species *Copaifera multijuga* and *C. reticulata* are most commonly available for purchase outside of Brazil, though they are often mislabeled as *C. officinalis*.

In a recent visit to Brazil during July, we were able to observe the oleoresin collection process (Fig. 4). The trees had to be drilled by hand, and reaching the core of the tree to access the heartwood where the oleoresins are stored was not easy. We observed the tapping of 12 *Copaifera langsdorffii* trees, none of which produced oleoresin. It was suggested that these trees may not produce oleoresin at all, or that they may not be in season as July is during the winter or dry season. This again reinforces the notion that tree species native to the northern parts of Brazil are more suitable for production of oleoresin, or at least traditionally there is a more widespread culture of oleoresin collection in the north.

6 Comparing Oleoresin to Diesel Fuel

Diesel fuel, like gasoline, consists of many different compounds isolated from only one fraction of the greater mixture known as crude oil. Diesel fuel distills

from crude oil between the temperatures of 200 and 350°C. Not all diesel fuel comes directly from primary distillation; processes like catalytic cracking, which breaks larger denser molecules into smaller ones, have been developed to generate more liquid fuels from crude oil barrels (Bacha et al., 2007). In general, diesel fuel is made up of paraffins (alkanes), naphthenes (cycloalkanes), olefins (alkenes), and aromatics. As mentioned before, *Copaifera* oleoresins consist primarily of sesquiterpenes hydrocarbons.

The important properties of diesel fuel are the cetane rating, low temperature operability, and volumetric heating value. Diesel engines produce combustion by compressing air, which, in turn, heats the air; at a designated moment of compression, fuel is injected into the chamber as tiny droplets, which vaporize and ignite. The cetane rating measures the ignition quality of fuels, or how readily the fuel burns. A fuel's quality of ignition can have implications in starting engines in cold conditions, as well as emissions, smoothness of operation, noise, and misfires (Bacha et al., 2007).

Low temperatures can cause some constituents in diesel fuels to solidify (such as the paraffins). This, in turn, can clog the fuel filter and stop the flow of fuel to the engine. This effect is measured with 'cloud points', the temperature when the waxes in the mixture begin to solidify, or the 'pour point', the temperature when the fuel becomes so thick it will no longer pour. The volumetric heating value measures how much energy the fuel has per volume.

Volumetric heating values influence torque, horsepower, and to some degree, fuel economy.

Both monoterpenes and sesquiterpenes are volatile cyclic hydrocarbons. The major sesquiterpene present in most *Copaifera* species, β-caryophyllene, has a chemical structure most similar to a cyclic olefin, or a naphthene, which contains two double bonds. In general, naphthenes have a midrange cetane rating, good low temperature properties, and an acceptable volumetric heating value. Biofuels from oilseed sources like soybean and canola have a pour and cloud point around 0°C making them impractical in areas with cold climates. In addition, fuel additives to improve low temperature properties are not very effective because of the high level of saturated compounds present in the oils (Tyson, 2004). Addition of terpenoid components, such as sesquiterpenes, to these types of biofuels could increase their low temperature properties and complement their high cetane ratings.

Not much is known about the chemical and physical properties of *Copaifera* oleoresin as a diesel fuel. Calvin (1980) submitted a sample of *Copaifera* oleoresin to the Mobil Corporation and obtained a cracking pattern: 50% aromatics, 25% liquid petroleum gas (LPG), 3-4% low-molecular-weight fuel gas, and coke. Later, cracking of *Copaifera officinalis* oleoresin with a zeolite catalyst, ZSM-5, led to production of over 200 compounds from 34 sesquiterpenes present in the original oleoresin (Stashenko et al., 1995). The

great variety of resulting products could indicate the utility of these oleoresins in not only fuels but also additional value-added products from a renewable resource. As mentioned before, the seeds of *Copaifera* species produce sesquiterpene hydrocarbons, but also produce various fatty acids when pressed and extracted (Neto et al., 2008; Stupp et al., 2008). In *Copaifera langsdorffii*, oleic acid (C18:1) made up 33.1% of the fatty acid profile while palmitic acid (C16:0) made up 20.2% of the fatty acid profile. According to Stupp et al. (2008), the major fatty acid that was extracted was linoleic acid (C18:2) which made up 45.3% of the fatty acids and oleic acid making up 30.9%. It would be interesting to test oil pressed from these seeds against other biodiesels, and to compare their overall chemical structure to see how the percentages of sesquiterpenes versus fatty acids are present in the seed oil.

7 Future Scope of Research and Development

For reasons described earlier, it does not seem economically feasible to create plantations of *Copaifera* trees to produce oleoresin for biodiesel markets. In brief, long generation times, low and sporadic yields per tree, and their tropical nature limit production of oleoresin. Instead, characterization of the unique terpenoid biosynthesis pathway and expressing it in other species already suited for production of biodiesel offers a more reasonable avenue.

Why these oleoresins produce higher amounts of certain terpenoids, sesquiterpenes mostly, is not well understood. Possible mechanisms include

differential regulation of sesquiterpene synthesis, or even higher TPS efficiency. While there has been a lot of work accomplished to characterize the chemical nature of oleoresins, there has been a surprising lack of molecular biology and biochemistry research as to how these oleoresins are created, stored, and transported. Identification, isolation, and characterization of the TPS responsible for the production of the oleoresin constituents will be crucial in first determining how these proteins function, but also localizing them within tissue types to understand production of oleoresin constituents.

Detailed studies on the emission and performance standards of oleoresins as a diesel fuel will also be necessary. The host of traditional diesel classification tests for physical and chemical properties including density, cloud and pour points, viscosity, heat of combustion, cetane number, etc., should be performed to gauge the usefulness of these compounds in today's markets with current engine technology. This work will also need to include several different possible species, because each one has a varying chemical makeup and properties associated with that makeup. These studies will be instrumental in determining whether oleoresin constituents are better suited as a stand-alone biodiesel fuel, or as an additive for other petroleum or biofuels.

There are, however, barriers to further research on *Copaifera* species. Many publications and historical records about *Copaifera* trees are in Portuguese; this presents a barrier to the larger scientific community and

hampers research efforts to assess available information. In addition, obtaining germplasm outside of the native range of the *Copaifera* genus has been exceptionally difficult. Collaborations on an international level will be crucial to establishing successful research initiatives.

We are performing genomics and biochemistry research to characterize and exploit the *Copaifera* terpene biosynthetic pathways. There are at least two end-goals of research. First, we need a better understanding of the basic biochemistry of this interesting genus and its oleoresins. Second, genes and gene regulation responsible for hydrocarbon production could be valuable with regards to their use in production of bio-products and fuels. For example, key genes might be transferred to temperate oilseed species to complement and increase their biofuel production. Taken together, we expect the diesel trees to contribute to new fuels and products beyond diesel.

Appendix

Figure 4. Identification and tapping of *Copaifera langsdorffii* trees near Nova Odessa, Sao Paulo State, Brazil.

A) A Copaifera langsdorffii tree growing near a farm in Nova Odessa, Brazil. The trees grow as single individuals rather than in stands making it difficult to locate and tap multiple trees. B) Tapping a Copaifera langsdorffii tree with a manual drill. The oils collect in the heartwood and so the hole must be drilled to the very center of the tree making collection difficult. C) Botanical characterization of Copaifera langsdorffii. Pictures of leaves, seeds with fleshy aril, and seed pods were taken to correctly identify the genus and species of the trees.

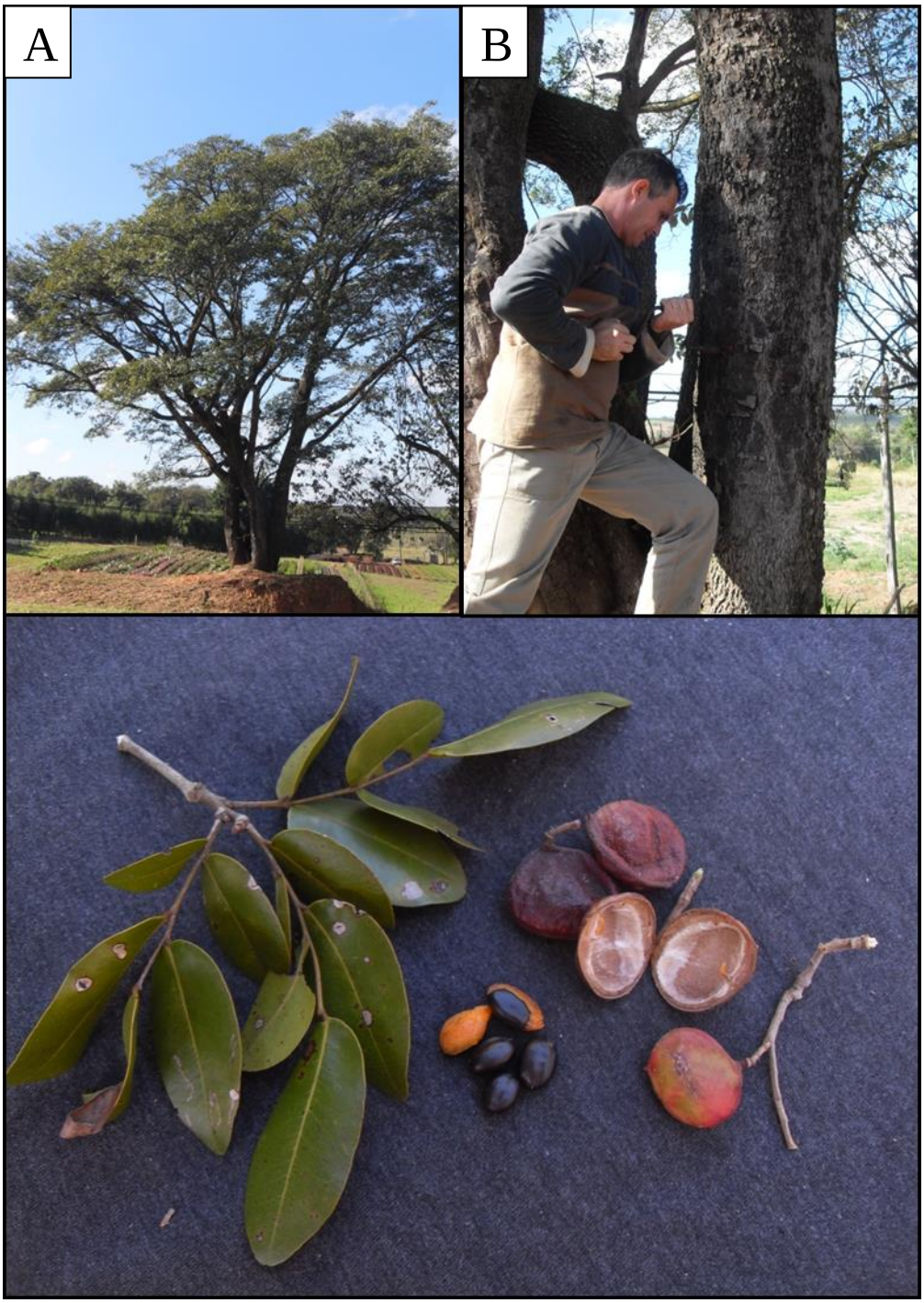

Figure 4.

Table 5. Three major sesquiterpenes present in the oleoresins of Copaifera species.

Species	Reference	Compound	Oleoresin Concentration (%)
Copaifera langsdorffii	Gramosa and Silveira 2005	β-caryophyllene	53.3
		germacrene B	8.7
		β-selinene	6.5
Copaifera martii	Zoghbi et al. 2007	β-caryophyllene	42.6 †
		δ-cadinene	15.7 †
		β-elemene	5.0 †
Copaifera multijuga	Veiga Junior et al. 2007	β-caryophyllene	57.5
		α-humulene	8.3
		α-bergamotene	2.6
Copaifera cearensis	Veiga Junior et al. 2007	β-caryophyllene	19.7
		α-copaene	8.2
		β-bisabolol	8.2
Copaifera reticulate	Veiga Junior et al. 2007	β-caryophyllene	40.9
		α-humulene	6.0
		α-bergamotene	4.1
Copaifera trapezifolia	Veiga Junior et al. 2006	β-caryophyllene	33.5
		germacrene D	11.0
		spathulenol	7.6

† Number represents an average of 11 sampling dates

CHAPTER II

BIODIESEL FUEL PROPERTIES AND COMBUSTION CHARACTERISTICS OF DROP-IN READY PLANT-DERIVED OILS THAT DO NOT CONTAIN FATTY ACIDS

This manuscript is formatted for submission to Fuel. B.L. Joyce helped to design the project, identified and collected the novel plant oils, analyzed the engine data, and wrote the manuscript. M. Bunce conducted HCCI engine trials, processed the collected data, and analyzed the data. S. Lewis ran GC/MS analysis of the novel plant oils and processed data. B. Bunting helped design the project, provided equipment, funding, mechanical engineering expertise, conducted HCCI engine, and edited the manuscript. C.N. Stewart, Jr. helped design the project, provided funding, and edited the manuscript.

B.L. Joyce[1], M. Bunce[2], S. Lewis[2], B. Bunting[2], C.N. Stewart, Jr[1]

Affiliations

1. Department of Plant Sciences, University of Tennessee, Knoxville, TN

2. Fuel, Engines, and Emissions Research Center, Oak Ridge National Laboratory

Abstract

Alkyl esters of fatty acids are the only compounds that are currently used in biodiesels. Other hydrocarbons from biomass for advanced fuels are of interest but few reports have investigated the fuel properties and combustion properties of these chemicals. Five oil-producing plant species were selected to analyze their compounds as potential diesel components: *Copaifera reticulata*, *Cymbopogon flexuosus*, *Cymbopogon martinii*, *Dictamnus albus*, and *Pittosporum resiniferum*. The compounds produced by these plants represent a range of traditional petroleum classes and are produced by biochemical pathways that could yield fractions of advanced biofuels from biomass that have not been converted into biofuels. Here we report ASTM International biodiesel fuel properties and HCCI engine combustion modeling of plant-derived oils in B20 blends with ultralow-sulfur-diesel fuel #2 (ULSD) in order to rapidly screen for suitability for use as biodiesel compounds. Four of six B20 blends failed ASTM International oxidation stability testing, and three of six failed cetane number testing. *C. flexuosus* oil extracts improved both oxidation stability and cetane number compared to ULSD control. All B20 blends had the same or lower cloud point than ULSD which is an unusual trait for biodiesel blends. The plant oils we studied did not require methyl esterification for use in blends with ULSD which separates them from traditional fatty acid methyl esters. These results

suggest that many biochemicals are suitable for use as biodiesel and greater chemical diversity might be useful to increase chemical flexibility in biodiesel fuel.

Keywords

Plant-derived fuel, Homogenous charge compression ignition engine, Advanced biofuel, Rapid screening, ASTM International biodiesel testing

Abbreviations

CO	Carbon monoxide
FAME	Fatty acid methyl ester
HC	Hydrocarbon
ISFC	Indicated specific fuel consumption
IMEP	Indicated mean effective pressure
NO_x	Nitrogen oxides
TDC	Top dead center
ULSD	Ultra-low-sulfur diesel

1. Introduction

Replacing petroleum-derived liquid fuels with renewable sources has become a major research focus for a number of reasons. Biofuels are expected to play an important role in replacing liquid fuels primarily used for transportation (). However, there are barriers to creating economically viable biofuels in abundant quantities that meet market demands which include biomass conversion efficiency, biomass supply, and product suitability for replacing existing fuels.

Ethanol, butanol, and fatty acid methyl esters (FAMEs), the major biofuels currently being produced, do not have ideal fuel properties. Butanol and other higher alcohols have been proposed as the next generation of biofuels for gasoline replacement and in some cases diesel replacement. However ethanol and longer carbon chain alcohols such as butanol still have technical limitations as both gasoline and diesel-replacement fuels (Kohse-Höinghaus et al., 2010; Laza and Bereczky, 2011; Pfromm et al., 2010). FAMEs also are not ideal fuels, but they are, by definition, the only biological chemicals currently allowed in biodiesel. Specifically, ASTM International has set the definition of biodiesel as “mono-alkyl esters of long chain fatty acids derived from vegetable oils or animals fats” (ASTM Standard D6751, 2012). This limits the chemical class diversity which also limits their flexibility and fungibility for petroleum fuel replacement. In contrast with the emphasis to produce drop-in gasoline replacements there has been much less effort to produce ‘advanced’ biodiesel.

Nonetheless, there are reports of biomass-derived chemicals other than fatty acid methyl esters that could be useful as a diesel replacement (Bond et al., 2010; Lange et al., 2010).

Current petroleum fuel chemistries include several classes: paraffins, olefins, naphthenes, and aromatics. The effect of petroleum chemistry classes on fuel properties, combustion characteristics, and emissions are generally understood (Table 6). However, biochemicals do not fit neatly into these traditional petroleum chemical classes. Biochemists classify them based on their biosynthetic pathways more often than on strict chemical structures. Identification of novel plant-derived hydrocarbons suitable as fuels and their corresponding biocatalysts might yield chemically diverse biofuels that could be capable of directly replacing a portion of petroleum-derived liquid fuels.

Plants synthesize myriad chemicals which include hydrocarbons such as isoprenes and even short-chain alkanes. Isoprenoids, also called terpenoids, were suggested in the early 1980s as fuel candidates that could replace petroleum fractions (Calvin, 1981). To date, a few plant-derived terpenoid oils have been investigated as fuel supplements, such as eucalyptus, orange oil, and turpentine (Karthikeyan et al., 2010; Poola et al., 1994; Tamilvendhan and Ilangovan, 2011; Yumrutas et al., 2008). As such, these studies are similar in nature to surrogate fuel experiments that have been conducted in an effort to understand how petroleum fuel chemistry affects fuel properties and engine

operability (Dagaut and Cathonnet, 2006; Pitz et al., 2007). Even though terpenoid oils are often a complex mixture of 10-30 biochemicals, typically only a few terpenes predominate the composition of a single plant's oil. This means that even complex terpenoid oils are less chemically diverse than petroleum fuels. In the case of terpenoid oils, the majority of the oil can have the same chemical formula, e.g. $C_{10}H_{16}$ for monoterpenes. However monoterpene isomers can be branched or cyclic, have single or double bonds, and can be modified after they are synthesized in plants to contain functional groups, e.g. alcohols and aldehydes, all of which will have an impact on the fuel properties of the oils. To illustrate this point, citral is a monoterpene aldehyde that contains double bonds, branched side chains, and an oxygen atom, which technically makes it an olefinic isoparaffin oxygenate in petroleum chemical nomenclature.

The next-generation of biomass feedstocks and their products are under development. There is a wide range of plant species and fuel conversion platforms under consideration and advanced biofuel chemicals are still not clearly defined. On the other hand, petroleum chemistry and chemical engineering is a mature field. Engine modeling and fuel chemistry research has been performed for decades to investigate petroleum-derived fuel chemistry, properties, combustion characteristics, and emissions (Pitz and Mueller, 2011). Comparatively, investigation of biofuel properties, combustion characteristics, and emissions is a recent endeavor and databases to investigate biological

chemicals as surrogates for fuels do not currently exist (Kohse-Höinghaus et al., 2010). Therefore, rapid screening of putative biofuels could identify the novel metabolites from plants suitable for advanced biofuels. Once suitable fuel biochemicals are identified biotechnological engineering can focus on modifying feedstocks and biofuel conversion microbes to produce better biofuels.

In this work we have identified five plant species that produce extractable oils that contain a variety of traditional petroleum chemical classes. We tested oils from *Copaifera reticulata*, *Cymbopogon flexuosus*, *Cymbopogon martinii*, *Dictamnus albus*, and *Pittosporum resiniferum* against standard ASTM International biodiesel fuel properties in 20% v/v blends (B20) with standard ultra-low sulfur highway diesel #2 (ULSD). Each B20 mixture was then compared to ULSD. Combustion characteristics and emissions of the B20 oil blends were determined using a homogenous charge compression ignition (HCCI) engine to determine their suitability as biodiesel.

2. Materials and methods

Plant species and their oils

Five plant species were selected based on biochemical composition of the plants, or previously described flammability of their oils. *Dictamnus albus*, commonly called the gas plant, has flammable flower essential oil composed of

methyl chavicol, anethole, and dictagymnin (Fleisher and Fleisher, 2004). Anethole was purchased as a surrogate because large amounts of *D. albus* essential oils could not be purchased or extracted directly from the plant for logistical reasons. *Pittosporum resiniferum,* 'the petroleum nut,' sequesters flammable oil in fruit tissues. The oil consists primarily of pinene and a small percentage of short-chain alkanes (Nemethy and Calvin, 1982). Petroleum nut oil was collected by hand-squeezing fruits into plastic containers in La Trinidad, Philippines. The oil formed a white waxy precipitate after storage at room temperature for two weeks. The liquid oil was decanted for mixing with ULSD to create the B20 *P. resiniferum* blend used in this work. *Copaifera reticulata,* a species in the genus that contains 'diesel trees,' produces an oleoresin rich in sesquiterpenes when the trunk of the tree is drilled and tapped (Veiga Junior and Pinto, 2002). *C. reticulata* oleoresin is composed of over thirty sesquiterpenes (Veiga Junior and Pinto, 2002). *C. reticulata* oleoresin was purchased both in the raw form direct from tree tapping and after processing through steam distillation. *Cymbopogon flexuosus*, lemongrass, is a grass species that produces essential oils that are primarily composed of citral, and *Cymbopogon martinii*, palmarosa, produces an essential oil that primarily contains geraniol (Chowdhury et al., 1998). Therefore, citral and geraniol were purchased as surrogates for the complex *C. flexuosus* and *C. martinii* essential oils, respectively, because liters of the complex oil were not available for purchase. Plant-derived oils were blended

with ULSD in a 20:80 ratio based on volume to prepare B20 blends. B20 blends were stored at 4°C for fuel property testing and HCCI engine testing.

Chemical composition determination

Plant oils were analyzed using GC-MS as previously described (Prikhodko et al., 2012). Individual chemicals were quantified based on area under each peak as a percent of the total area. Any chemical with a peak area percent less than 0.5% or a fragmentation pattern with a quality of match score to the National Institute of Standards and Technology (NIST) mass spectral library of less than 80 was considered unknown. Chemicals were arranged into tradition petroleum classes based on chemical structure, and into biochemical classes based on the pathways responsible for biosynthesis of the identified compounds in plant species. These classifications are intended to provide a summary of each plant species' oil chemistry for comparison to fuel properties.

Fuel property testing

B20 oils and ULSD #2 were submitted for standard ASTM International biodiesel testing (Herguth Laboratories, Inc, Vallejo, CA). A complete list of ASTM International tests performed is presented in Table 7.

HCCI engine specifications and operation

The HCCI engine used was as previously described to investigate emissions and combustion characteristics of the B20 oil blends in comparison to ULSD (Bunting

et al., 2011). In our experiments, fuel flow was held constant during experiments and air intake temperature was varied to control combustion phasing as indicated by the crank angle of the piston where 50% of the fuel has undergone combustion which is referred to as the 50% mass fraction burned point (MFB50). Crank angle is reported as 0° being top dead center (TDC). Air flow and fuel to air ratio (lambda) changed in relation to air intake temperatures.

3. Theory/calculation

Prediction of air intake temperatures at MFB50 top dead center

Lines were fitted to each B20 blend and ULSD data set from Figure 8A to predict the air intake temperature that resulted in an MFB50 at top dead center for Figure 8B and 7C. Any fitted line with an R value lower than 0.98 was not used. The function of the fitted line was then used to calculate the air intake temperature at MFB50 at TDC (x-intercept) of each B20 blend data set presented in Figure 9A.

4. Results

Plant oil chemical composition

Chemical analysis was needed to verify that the identified plant-derived oils were composed primarily of hydrocarbons rather than fatty acids as in standard biodiesel. Initially kukui nut oil, isolated from the plant species *Aleurites moluccana*, was also investigated but it was removed because it was composed primarily of triglycerides and free fatty acids (results not shown). All other oils investigated comprised hydrocarbons which were then classified into traditional petroleum fuel classes based on chemical structure (Table 8), and biochemical classes responsible for the biosynthesis of identified compounds in plant species (Table 9).

The compounds in the plant oils (C_{10}-C_{20}) we tested were smaller than mono-, di-, and triglycerides that are used to make standard biodiesel. As such, these oils may be considered “drop-in” ready advanced biofuels. However *C. reticulata* oils were investigated both in the “raw” oil direct from the tree and after steam distillation which removes heavy compounds and contaminants from the extraction process. Likewise, *P. resiniferum* oils were allowed to settle before testing. A wax layer precipitated at the bottom of *P. resiniferum* oil containers and the liquid was decanted for use in all fuel and engine testing. The chemical composition of the wax precipitate could not be identified in this work. The decanted oil still had a large fraction (20.84%) of GC/MS library peaks with

identification quality scores <80 and were therefore classified as ‘unknown’ (Table 8). This may be as a result of the waxy substrate still suspended in solution.

Plant oil fuel properties

Distillation ranges are used in petroleum fuels to determine fuel volatility. Distillation profiles for the B20 blends and ULSD control were similar for all B20 blends except for those originating from *P. resiniferum* and *Cymbopogon martinii* (Figure 6). *C. flexuosus* and *P. resiniferum* B20 blends had lower initial distillation points than ULSD. The 10% distillation point of steam distilled and raw *Copaifera reticulata* B20 blends was 8 °C higher than the ULSD control.

The plant oils were first screened in B20 blends by standard ASTM International biodiesel testing to determine their suitability as novel advanced biofuels. All B20 blends of plant oils failed at least one ASTM International fuel property test (Table 10). The test failed most commonly was oxidation stability. The B20 blends of *C. reticulata* steam distilled, *D. albus*, *C. martinii*, and *P. resiniferum* oils failed this test. Oxidation stability (EN 14112) is tested to estimate storage length or shelf life of biodiesel. In short, air is bubbled through the B20 sample until the sample begins to produce volatile peroxides and carboxylic acids which then passed into a conductivity measuring vessel filled with distilled water. As the peroxides and carboxylic acids dissolve in the distilled water conductivity is increased. The time this process takes is measured and

reported in hours with a minimum of 12 hours before oxidation required to past the test. *C. flexuosus* B20 passed this test with over 24 hours while *C. martinii* B20 failed it with only 1.3 hours.

Some fuel properties for the B20 blends and ULSD are summarized in Figure 7. *C. reticulata* raw, *D. albus* and *C. martinii* B20 blends had lower cetane numbers than the ULSD control, but all other oils had similar or higher cetane values than ULSD (Figure 7A). Cloud point measures the temperature that biodiesels begin to solidify and lower numbers mean that fuels will remain liquid in colder climates. Cloud point remained the same as ULSD controls in *Cymbopogon flexuosus* and *C. martinii* B20 blends and was lower in *Copaifera reticulata* steam distilled, *C. reticulata* raw, *D. albus*, and *P. resiniferum* B20 blends (Figure 7B). The B20 blend of *P. resiniferum*-derived oil was the only one to fail the flash point test (Figure 7C). Net heat of combustion was lowered in all B20 blends except for *Copaifera reticulata* steam distilled which increased net heat of combustion by 3% (Figure 7D). The *D. albus* B20 blend had the lowest net heat of combustion of the B20 blends studied. A B20 blend of *D. albus* reduced net heat of combustion by approximately 2000 kJ kg^{-1} which is a 4% reduction in energy content. *Cymbopogon flexuosus* and *P. resiniferum* B20 blends reduced net heat of combustion by 3% and 2%, respectively. All B20 blends had similar viscosity to ULSD except for B20 *Copaifera reticulata* raw which failed the viscosity test (Figure 7E). Sulfur content of all the B20 blends

was similar within measurement uncertainty and all met the 15 µg g^{-1} requirement.

Engine combustion experiments

B20 plant oil blends were run in an experimental HCCI engine to determine their suitability as fuels. *C. martinii* and *C. flexuosus* oils had the shortest combustion phasing range (MFB50) which occurred across the lower range (185-260 °C) of intake temperatures investigated (Figure 7B, 7C, and 8A). *C. flexuosus* and *C. martinii* B20 blends also had the lowest IMEP controlled for net heat of combustion differences which were 0.639 bar kg J^{-1} to 0.702 bar kg J^{-1} and 0.668 bar kg J^{-1} to 0.766 kg J^{-1}, respectively (Figure 9A).

Indicated specific fuel consumption (ISFC) is a measure of fuel efficiency. The B20 oil blends had different ISFC (g fuel $(kW\ hr)^{-1}$) with both raw and distilled *C. reticulata* B20 blends having the lowest ISFC followed by ULSD, *C. martinii*, *P. resiniferum*, *D. albus*, and *C. flexuosus* with the greatest ISFC (Figure 9B). ISFC was also investigated on a per volume basis (mL $(kW\ hr)^{-1}$) rather than a per weight basis as in Figure 9, but no difference in the trends was found (data not shown). In all cases, the local minimum of fuel consumption was found when combustion phasing (MFB50) was near 5° after TDC. The length of the combustion from 10% mass burned to 90% burned was also measured in the total degrees of rotation the piston moved across the combustion event (MFB10-MFB90) measured in °CA at different combustion phasing (MFB50) (Figure 10).

At all MFB50, *C. flexuosus* and *C. martinii* B20 blends had longer combustion times than the other B20 blends (Figure 10). The *C. flexuosus* B20 blend also had the lowest maximum increase in cylinder pressure (dP/dCA) for all MFB50 (Figure 11). *C. martinii* B20 blends dP/dCA measurements were only slightly lower than ULSD and other B20 plant oil blends.

Emissions were collected during the engine tests including unburned hydrocarbons (HC), nitrous oxides (NO_x), carbon monoxide (CO), and soot. *C. reticulata* B20 blends had comparable HC emissions to ULSD, but all the other B20 blends resulted in higher HC emissions than ULSD (Figure 12). For all fuels, NO_x emissions increased as combustion phasing advanced before TDC (Figure 13A). NO_x emissions were also graphically represented based on air intake temperature as these emissions increase as the combustion temperature increases (Bunting et al., 2009). The B20 blends of *C. flexuosus*, *C. martinii*, and *P. resiniferum* did show an increase in NO_x emissions at higher combustion temperatures like ULSD, *C. reticulata* oils, and *D. albus* (Figure 13B). *C. flexuosus* and *C. martinii* B20 blends had the highest CO emissions, whereas the two B20 blends of *C. reticulata* had lower CO emissions than ULSD (Figure 14). Soot is characterized as a particulate emission that is formed by incomplete combustion at low temperature and oxygen levels. While soot emissions were low for all B20 blends, a group of higher and lower soot emission oils can be observed (Figure 15). In general, *C. martinii*, *D. albus*, and *P. resiniferum* B20

blends had higher soot emissions across MFB50 whereas ULSD, *C. flexuosus*, and both *C. reticulata* oils had lower soot emissions (Figure 15).

5. Discussion

Comparison of the experimental plant oil chemistries and traditional petroleum fuel classes

The surrogate compound for *C. martinii* oil was composed primarily of the C_{10} monoterpene alcohol geraniol while the surrogate for *C. flexuosus* oil was composed primarily of the monoterpene aldehyde citral. Therefore, the primary difference between these two oils was the functional group found on the monoterpene backbone. This matches previous reports of *C. flexuosus* and *C. martinii* essential oils (Chowdhury, 1998; Kulkarni et al., 1992; Singh, 2001). *C. martinii* contained a small fraction of other monoterpene compounds but ultimately these compounds would have made up approximately 1% of the final B20 blends and therefore do not likely contribute to overall fuel properties.

P. resiniferum fruit oil was also composed primarily of monoterpenes, however the monoterpenes were cyclic rather than linear. Pinene was the major compound found in *P. resiniferum* oil and is also the major chemical constituent of pine tree turpentine (Lewinsohn et al., 1993; Mirov, 1952). *P. resiniferum* oil also contained the C_9 straight-chain alkane nonane. Nonane is is a paraffin and

one of the only biochemicals that was easily classified into petroleum fuel classes. This was the only investigated plant-derived oil that contained alkanes. In general, most land plants make long-chain alkanes in the form of leaf epicuticular waxes, however short-chain alkanes have only been reported in *Pinus* (Anderson et al., 1969), *Pittosporum* (John et al., 2008; Nemethy and Calvin, 1982), and *Bursera* species (Evans and Becerra, 2006). Alkanes also made up the largest fraction of chemical compounds in ULSD and nonane is a compound found in petroleum fuels. This suggests that *P. resiniferum* is capable of producing and storing petroleum fuels in small fractions. Regrettably, nearly twenty percent of the total chromatogram peak area was not identified. This could have resulted from the wax precipitate and other long-nonpolar compounds present in the oil. The chemistry of *Pittosporum* species has been previously studied using either chemical properties (Nemethy and Calvin, 1982) or volatile fractions collected from hydrodistillation (John et al., 2008; Medeiros et al., 2003; Weston, 2004). These reports identified the same monoterpenes and short-chain alkanes as described in this work. Therefore further in-depth chemical analysis of total fruit oils derived from simple compression of fruit tissue will be needed in future work to determine the wax precipitates and other compounds *in P. resiniferum*-derived oil.

Raw *Copaifera* oleoresin failed acid number and cold soak filtration and had a lower cetane number. The raw oils also contained a higher fraction of

monocyclic naphthenes and unknown compounds. Additionally, the raw oleoresin had a higher carbon residue than steam distilled oleoresin. In comparison, the steam distilled *Copaifera* oleoresins passed all of the tests that were failed by the raw oleoresins. However, steam distilled B20 blend failed oxidation stability while the raw oleoresin B20 blend passed.

Both raw and steam-distilled *C. reticulata* oils were investigated. The key difference between raw and steam distilled *C. reticulata* oil was a shift in the percentage of monocyclic sesquiterpenoids to bicyclic sesquiterpenoids after steam distillation because monocyclic sesquiterpenes such as bisabolene have a lower boiling point (90 °C) compared to bicyclic sesquiterpenes such as β-caryophylene (130 °C). However, this does not account for the lower viscosity in steam distilled *C. reticulata* B20 blends as monocyclic sesquiterpenes generally have a similar or lower density than bicyclic sesquiterpenes. Steam-distillation also lowered acid number of *C. reticulata* oil which suggests the presence of an organic acid. *Copaifera* oleoresins are known to have small concentrations of diterpene resin acids which contain carboxylic acid functional groups (Veiga Junior and Pinto, 2002). Diterpene resin acids are also are more dense than either monocyclic or bicyclic sesquiterpenes. Therefore the lower acid number and viscosity in *C. reticulata* steam distillation B20 blends suggests that part of the unknown fraction of *C. reticulata* oil were diterpene resin acids which were removed by steam distillation.

The final oil selected represented the aromatic petroleum chemical class. The *D. albus* surrogate anethole has a chemical structure similar to monolignols present in lignin. This compound was selected for two reasons. Firstly, it has a similar chemical structure to methyl chavicol (estragole) found in high concentrations in the combustible flowers of *D. albus*. Secondly, lignin makes up nearly a third of dry plant biomass in most lignocellulosic feedstocks such as switchgrass and poplar (Mann et al., 2009; Sannigrahi et al., 2010). Effective techniques to separate lignin from bulk biomass have been sought to reduce inhibitory effects of lignin on biofuel production methods and reduce heterogeneity of substrate for conversion of biomass to products and biofuels (Bozell et al., 2011). As such, monolignols separated from biomass may become widely available in the future as either a waste stream or sold as a bulk commodity for chemical conversion into products. Aromatic biofuels derived from monolignols could present an opportunity to utilize compounds that previously inhibited biorefinery processes as biofuels.

Although the plant oils were composed of fewer chemicals than ULSD, the biochemicals proved to be difficult to sort into traditional petroleum fuel classes, i.e. paraffins, olefins, naphthenes, and aromatics (Table 8). For instance, citral from *C. flexuosus* oil is a monoterpene aldehyde that has unsaturated bonds, is branched, and contains an oxygen atom. This makes the one compound an isoparaffin, an olefin, and an oxygenate. Likewise geraniol from *C. martinii* has a

similar chemical structure to citral, but has an alcohol group instead of a ketone. *P. resiniferum* fruit oil was composed of true paraffins, heptane and nonane, as well as cyclic naphthenes and a large fraction of compounds that could not be identified using the GC/MS. *Copaifera reticulata* oleoresin was composed primarily of cyclic naphthenes. Currently, the definition of what can be sold as biodiesel is derived from the official ASTM International definition. However, fatty acids represent only a narrow window of renewable chemicals that can be derived from plant biomass (Table 6). The five plant oils described in this work do not contain FAMEs. As such, they do not fit the standard definition of biodiesel.

Fuel, combustion, and emission properties of B20 plant oil blends

Fuel chemical structures affect fuel properties and combustion characteristics in an engine leading to different emissions, fuel consumption, and work output (Speight, 1999). The chemical profiles of the investigated oils match with general understanding of how petroleum-derived diesel fuel chemistry affects fuel characteristics (Speight, 1999). *C. martinii* and *C. flexuosus* were C_{10} olefinic isoparaffins with carbon numbers between the usual range of gasoline (C_4-C_{12}) and diesel fuels (C_8-C_{18}) (Joyce and Stewart Jr, 2012). Typically, higher cetane number leads to a longer combustion time in HCCI engines which is measured in crank angle from mass fraction burned 10% to 90% (MFB10-90) (Starck et al., 2010). *C. martinii* had the lowest cetane number of the oils tested, but had the second longest MFB10-90. *C. flexuosus* B20 oils did follow the

previously reported trend having the second highest cetane number and the longest MFB10-90. *C. flexuosus* also had the lowest dP/dCA (Figure 11) and the highest ISFC (Figure 9) of all the oil blends investigated. The high fuel consumption in both B20 blends likely results from the long combustion time that produces a slow increase in cylinder pressure and/or from fuel energy lost through CO emissions that were higher than ULSD in both B20 blends. The primary difference between the two oils is that citral in *C. flexuosus* has an aldehyde functional group whereas geraniol in *C. martinii* has an alcohol group. *C. flexuosus* produced more CO emissions and had lower ISFC than *C. martinii* which could result from the double bonded oxygen atom present in the aldehyde functional group present in citral. Combustion could free the aldehyde to form a triple bond resulting in CO.

Soot forms when combustion occurs at low temperatures and oxygen levels. *C. martinii* and *C. flexuosus* B20 blends had the highest soot emissions of all the B20 blends despite being the only blends with oxygen atoms. Both of these B20 blends had the lowest NOx emissions of the oils investigated (Figure 13). NOx emissions increase in high temperature combustion or advanced combustion timing (Bunting et al., 2009). Therefore the soot and NOx emission from *C. flexuosus* and *C. martinii* B20 blends likely result from the low temperature combustion. This suggests the combustion conditions have a

stronger influence on emissions rather than the presence of oxygen atoms in the fuel.

These two B20 blends also had the lowest range of air intake temperatures studied which means that these B20 blends were more ignitable than other B20 blends studied (Figure 8C). Eucalyptus oil is comprised primarily of pinene, phellandrene, 1,8-cineole, globulol, and terpen-4-ol (Tamilvendhan and Ilangovan, 2011) which are structural isomers of major compounds present in *C. flexuosus* and *C. martinii*. However, combustion of eucalyptus oil blends in compression ignition engines resulted in higher NOx emissions and lower hydrocarbon than reference diesel fuel which was the opposite trend observed in this work (Tamilvendhan and Ilangovan, 2011). This serves as a reminder that while plant oils are not as chemically complex as petroleum fuels the overall fuel composition and engine settings will influence combustion characteristics.

P. resiniferum oil had a large fraction of α- and β-pinene which was classified as a monocyclic naphthene (Table 8). Turpentine is also comprised primarily of α- and β-pinene (Mirov, 1952) and *P. resiniferum* B20 blends exhibited low flash points similar to turpentine (Karthikeyan et al., 2010). The calorific value (kJ kg^{-1}) of B100 orange oil was found to be lower than diesel (Purushothaman and Nagarajan, 2009) which was also seen in most B20 blends studied (Table 10). Hydrogenated myrcene and limonene have also been investigated previously (Tracy et al., 2009). In B10 blends with diesel fuel,

myrcene reduced the cetane number by 0.9 while limonene reduced it by 2.8. Myrcene increased the viscosity of the fuel while limonene reduced viscosity (Tracy et al., 2009). However, blending *P. resiniferum*-derived oil into ULSD increased cetane number 0.8 and did not affect viscosity. *P. resiniferum* B20 blends had an ISFC higher than but similar to ULSD. *P. resiniferum* B20 blends had a delayed increase of the rate of NOx formation at all air intake temperatures studied and the rate of NOx emissions increased at higher air intake temperatures than any other B20 blend or ULSD control (Figure 13B). *P. resiniferum* had one of the shortest combustion lengths (MFB10-90) which could lead to less time to form NOx intermediates. The results reported here are contradictory to previous reports of emissions from eucalyptus oils which have a fraction of pinene and result in lower hydrocarbon emissions and higher NOx (Starck et al., 2010). Pinene was previously investigated for high-altitude fuel blends and had characteristics comparable to JP-10 (Harvey et al., 2009).

Both raw and steam-distilled *C. reticulata* B20 blends exhibited the lowest ISFC of all fuel blends investigated and ULSD (Figure 9). Steam distilled *C. reticulata* had the highest cetane number of the B20 oils tested (Table 8) and the shortest MFB10-90 (Figure 10). Combustion of *Copaifera* oils therefore is similar in nature to ULSD under these engine conditions and may explain the lower ISFC observed. Hydrocarbon, NOx, CO, and soot emissions of both steam distilled and raw B20 *C. reticulata* blends were similar to ULSD. Low emissions of

hydrocarbons and CO likely led to the lower ISFC of B20 *C. reticulata*. The observed fuel property differences between raw and distilled *C. reticulata* oils likely result from removal of diterpenoid resin acids.

Lastly, the surrogate for *D. albus* oil anethole was classified as an aromatic compound (Table 8). The *D. albus* B20 blend failed cetane number and oxidation stability ASTM International tests. However, cloud point temperature and sulfur content were lowered. The change in sulfur content most likely results from a dilution effect as anethole does not contain sulfur. During combustion *D. albus* B20 blends exhibited the third highest ISFC, a MFB10-90 statistically similar to ULSD, and a lower dP/dCA than ULSD. The *D. albus* B20 blend had the lowest net heat of combustion of all the B20 oil blends investigated. The low energy content of the fuel and low maximum pressure generated (dP/dCA) would likely result in the high specific fuel consumption observed. *D. albus* B20 blends had higher hydrocarbon emissions than ULSD, but all other emissions were similar to ULSD.

Overall trends in the experimental B20 plant oil blends

All of the B20 plant oils investigated in this work had similar or lower cloud points than ULSD with a range of -16 to -21 °C (Table 8). In comparison, B100 FAMEs have a range of cloud points from -9 to 5 °C (Knothe, 2009; Pinzi et al., 2011). This is expected as only long-chain paraffins and fatty acid methyl esters increase the temperature that fuels begin to solidify (increase cloud point

temperature) in traditional petroleum fuel classes (Speight, 1999). Generally, cloud point temperature increases with increasing length of the paraffin which explains the underlying chemical predisposition long-chain saturated FAMEs have to high temperature cloud points and resulting poor cold flow properties (Knothe, 2009). Currently, cold flow properties of biodiesel are primarily changed by using different mixtures of fatty acids from plant species or by using additives. Recently, nine combinations of fatty acid methyl esters from *Glycine max* (soybean), *Gossypium* spp. (cotton), *Jatropha curcas*, and *Attalea martiana* (babassu) were mixed to determine the temperature which the blends would no longer pour (pour point) which ranged from -5 °C to 18 °C (Freire et al., 2012). Pour points are reached when biodiesel blends are nearly solidified and are therefore lower than cloud points. Cold flow improvers are chemicals added to biodiesel which lowers the temperature at which biodiesel begins to freeze. Many cold flow improvers exist but vary in their effectiveness to lower cloud point. Polyglycerol ester and commercial DEP were added to palm oil biodiesel but only lowered the temperature at which the biodiesel passed through a filter, or cold filter plug point, from 16 °C to 9 °C (Lv et al., 2013). Both strategies lower cloud point through a similar mechanism of reducing FAME crystal formation at low temperatures (Ng et al., 2010). For example, unsaturated FAMEs are bent by double bonds that saturated FAMEs do not have and therefore do not pack as tightly as the relatively straight-chain saturated FAMEs. This reduces crowding of

the FAMEs molecules when temperatures are lowered effectively reducing crystal formation and therefore lowering the cloud point. The majority of plant-derived biochemicals used in this work are cyclic, branched, or contain numerous double bonds which make their three dimensional structures nonlinear when compared to FAMEs. These types of structures will inherently have lower cloud points comparative to straighter and longer chain FAMEs. To our knowledge, the experimental fuel blends in our study had the lowest cloud point temperatures reported for biodiesel blend stocks. The B20 blends studied here were blended with ULSD and as such were expected to have a lower cloud point than B100 FAME biodiesels. Future investigation of these plant oils blended with FAME biodiesels would yield additional useful data for further comparison of these novel biofuels to traditional FAME biodiesels.

Several of the B20 blends failed the oxidation stability test. However, this test assumes the presence of FAMEs. The methyl esters in biodiesel will degrade into peroxides and then into volatile carboxylic acids under test conditions. However, there are no FAMEs present in the B20 blends studied and so it is not clear whether the test results have meaning. Typically, alcohols will oxidize to aldehydes and then aldehydes will oxidize to carboxylic acids in the presence of water. However, geraniol used as a surrogate for the *Cymbopogon martinii* B20 blend failed the oxidation stability test while the citral used as a surrogate for the *C. flexuosus* B20 blend passed. The aldehyde group in citral

likely does not convert into a volatile carboxylic acid because there is no water present in the B20 blend. Aldehydes can undergo auto-oxidation to acids in the presence of air alone, but this was not detected in this test. Anethole and geraniol have been shown to oxidize in the presence of molecular oxygen and heat to form a peroxide which explains why the *D. albus* and *C. martinii* B20 blends failed (Elgendy and Khayyat, 2008; Hagvall et al., 2007). Although these B20 blends failed the oxidation stability testing it does not necessarily mean that they would become rancid like biodiesel. Further investigation of how oxidation affects the plant-derived oils will be needed to determine whether they have short storage times.

Further investigation of non-fatty acid based plant oils will be required to fully understand whether these novel plant-derived biochemicals are suitable for use as biofuels. Specifically, investigation of individual biochemicals will help to illuminate how biochemistry relates to fuel and combustion properties. This same work has been done with surrogate petroleum fuels to investigate how fuel chemistry and fuel properties relate. As a result, the relation between petroleum chemical classes and fuel chemistry is generally understood (Speight, 1999). In this work, the small biochemical difference between an aldehyde functional group (citral in *C. flexuosus*) and alcohol group (geraniol in *C. martinii*) resulted in a significant difference in fuel properties such as cetane number and oxidation stability as well as combustion and emission properties such as ISFC, dP/dCA,

and CO emission. Both of these compounds would be deemed ‘oxygenates’ in traditional fuel classes, but have significantly different fuel properties. This highlights the relative unknown connection between biochemicals and fuel properties.

Identifying novel biofuels could increase the biofuel supply through on site use of plant oils produced as secondary products, e.g. eucalyptus and cedar oils from timber and paper production. Currently, a major barrier to lignocellulosic biorefineries is recalcitrance which results from resistance of plant biomass to digestion (Octave and Thomas, 2009). Lignin makes up 20% to 30% of lignocellulosic biomass and must be removed or digested to access sugars for biofuel production (Carroll and Somerville, 2009; Sannigrahi et al., 2010). In this work, anethole from *D. albus* was investigated as a surrogate for monolignols as it has a chemical structure similar to p-coumaryl alcohol.

The next generation of biofuels revolves around the source material of the biofuel because current production of biofuels is too limited to meet demand for liquid transportation fuels. As such, biodiesel has two proposed generations: 1) fatty acid methyl esters derived from animal or oilseed plants oils and 2) fatty acid methyl esters from algae-derived triglycerides (Mata et al., 2010; Melero et al., 2010). Therefore the major focus in bioenergy research seems to involve increasing or changing the source of biodiesel, but consideration of biodiesel fuel properties is necessary to produce advanced biodiesel. Alkyl esters of long chain

fatty acids represent a single petroleum fuel class that will inherently have a narrow range of possible fuel and combustion properties. FAME fuel properties include both positive properties such as cetane numbers above ULSD, and negative properties such as cloud points near 0 °C (Demirbas, 2009b). Expanding the diversity of chemicals suitable for use as biodiesel would allow optimization of biodiesel emissions, fuel, and combustion properties. The biochemicals in these five oils will also not require methyl esterification. This has the additional advantage in reducing or eliminating the need to chemically convert biodiesel which would result in reducing waste water and processing equipment required for methyl esterification of current generation biodiesel fuels. The plant species that produce the studied oils are not productive enough to meet demand for liquid fuels in countries such as the United States. However, metabolic engineering could be employed to add these chemicals to biodiesel feedstocks to create an advanced biodiesel that can be optimized to meet market demands.

6. Conclusions

This work represents an initial effort to investigate novel “drop-in ready” plant-derived advanced biodiesels. The oils in this study do not require methyl esterification as is needed in most fatty acid-derived biodiesels and as such are considered drop-in ready. The biochemicals in the oils have fundamentally

different chemical structures when compared to FAMEs ranging from paraffins, olefins, and aromatics. As a result, they have unique biofuel properties such as lower cloud point than FAME biodiesels. These unique biofuel properties could make biodiesel blends more flexible to meet market demands and bring biodiesel closer to being a true petroleum diesel replacement. However, fuel and combustion properties are complex and involve many variables that interact with each other outside of fuel chemistry. As such, this work is intended to be an initial screening of the different fuel chemistries available in plant species for advanced biofuels.

All B20 plant oil blends had the same or lower cloud point as ULSD (*P. resiniferum*, *C. reticulata*, and *D. albus*) ranging from -16 °C to -21 °C. Additionally, distilled *C. reticulata* B20 blends had lower ISFC than ULSD suggesting better fuel economy under certain engine conditions. All B20 blends studied failed either one or two ASTM International tests for biodiesel. The majority of B20 blends studied failed cetane number (*C. martinii*, raw *C. reticulata*, and *D. albus*) or oxidation stability (*C. martinii*, *P. resiniferum*, *C. reticulata* distilled, and *D. albus*) tests. *C. flexuosus* improved these fuel properties and may prove useful as a cetane improver and antioxidant in biodiesel blends. However, it is important to note that EN 14112 measures oxidation stability based on the assumption that the biofuel is composed of fatty acid methyl esters that will break down to peroxides and carboxylic acids. It is

unclear how these novel biofuel chemistries will oxidize under test conditions and what products are being volatilized from oxidation. Therefore results from this test most likely do not correspond directly to oxidation stability.

The CO and NOx emissions of the studied B20 blends were similar to ULSD, but unburned hydrocarbon emissions were higher for most B20 blends studied (*C. martinii*, *C. flexuosus*, *P. resiniferum*, and *D. albus*). Of the studied B20 blends, *C. reticulata* oil was most suitable for use as biodiesel. *C. martinii* and *D. albus* were suitable for use as biodiesel but may be more suitable in lower blending ratios. This work suggests that there are many biochemicals that are suitable for use as biodiesel in addition to other chemicals synthesized from biomass such as valeric esters (Bond et al., 2010; Lange et al., 2010). Therefore the definition of biodiesel that designates only fatty acid alkyl esters as biodiesel may need to be reconsidered in the future.

Acknowledgements

We would like to thank the University of Tennessee Science Alliance and Oak Ridge National Laboratory for funding the project. We would also like to thank Dr. Michael Bengwayan, Clyde Awal, Manuel Aranas, and Jing Godio for hosting B. Joyce in La Trinidad, Philippines and supplying *P. resiniferum* oil. We would also like to thank Dr. Valtcho Jeliazkov for supplying *Cymbopogon flexuosus* and *C. martinii* oil.

Appendix

Table 6. Potential sources and chemical classes of plant-derived biofuels.

Biochemical Class	Plant-derived Chemical	Petroleum class	Plant Source	Conversion Required for Fuels	Fuel Testing
Monoterpenes (C_{10})	citral	Paraffin aldehyde	*Cymbopogon flexuosus* *Litsea cubeba*		
	geraniol	Paraffin alcohol	*Cymbopogon martinii*		
	Pinene	Bicyclic naphthene	*Pinus* spp. (turpentine)		Karthikeyan et al., 2010 Yumrutas et al., 2008
	Limonene	Naphthene	Citrus processing by-product		Poola et al., 1994
	1, 8-cineole (eucalyptol; monoterpene ether)	Bicyclic oxygenated naphthene	*Eucalyptus* trees	Used directly in mixtures	Tamilvendhan et al., 2011
Sesquiterpenes (C_{15})	Farnesol	Branched olefin		Methylation	
	Beta-caryophyllene	Bicyclic naphthene	*Copaifera* spp.		
Diterpenes (C_{20})	most likely not useful for fuels	Asphaltenes			
Complex terpene oils	Multiple	Complex	*Chrysopogon zizanoides*		
	cedrol; thujopsene	Tricylic naphthene alcohol	*Juniperus virginiana* *Thuja plicata*		
	linalool methyl chavicol	Paraffin alcohol	Basil oil		
Short-chain alkanes	nonane, undecane	Paraffins	*Pittosporum resiniferum*		
Long-chain alkanes	Epicuticular wax	Paraffins			
Fatty acids	Acetal compounds		Oilseeds - glycerol	Chemical conversion	Yes
Monolignols	p-coumaryl alcohol	Aromatics	Lignin	Monomerization (pyrolysis; enzyme degradation)	
	coniferyl alcohol	Aromatics	Lignin	Monomerization	
	sinapyl alcohol	Aromatics	Lignin	Monomerization	
	Vanilin	Aromatics	Lignin	Monomerization	
Valeric esters	ethyl valerate	Paraffin acid	Glucose	Chemical/catalytic conversion	Lange et al., 2010
	γ-valerolactone alkenes	Olefins	Levulinic acid	Chemical conversion from cellulosic biomass	Bond et al., 2010
Dimethyloctane	Functional Genomix patent		Geraniol	Microbial	

Table 7. Comprehensive list of ASTM International biodiesel tests performed on B20 plant oil blends and ULSD.

ASTM Test Number	Fuel Property	Explanation
ASTM D93	Flash point	Minimum temperature fuel/air mixture will ignite
ASTM D2709	Sediment and water	Fuel cleanliness
ASTM D445	Kinematic viscosity	Viscosity of fuel that determines flow properties and affects droplet size
ASTM D613	Cetane number	Diesel fuel ignition delay that estimates ignition properties
ASTM D874	Sulfated ash	Metal content in fuel
ASTM D130	Copper corrosion	Compatibility test for copper alloy systems
ASTM D2500	Cloud point	Temperature which fuel begins to solidify and form a cloudy precipitate
ASTM D664	Acid number	Concentration of acids present in fuel
D6751	Cold soak filtration	Weight of fuel particulates from cold filtered fuel that estimates fuel filter clogging potential
EN14112	Oxidation stability	Estimates shelf life by determining acid production in fuel
ASTM D86	Distillation of petroleum	Broad volatility profile
ASTM D240	Calorific value	Amount of energy present in a volume of the fuel
ASTM D4052	Specific gravity/API gravity	Fuel density
ASTM D524	Ramsbottom carbon residue	Carbon residue left after pyrolysis of fuel
	Total glycerin	Concentration of glycerin in biodiesel fuel blend
ASTM D5291	Elemental composition	Carbon, hydrogen, nitrogen, oxygen mass percent

Table 8. Chemical composition (%) of plant oils divided into petroleum chemical classes based on chemical structure.

	Linear alkanes			Cycloalkanes			
Plant Source of Oil	**Paraffin**	**Olefin**	**Iso-paraffin**	**Monocyclic Naphthene**	**Bicyclic Naphthene**	**Aromatic**	**Unknown**
C. flexuosus (surrogate)		85.98				4.35	
C. martinii (surrogate)		76.96	6.16				
P. resiniferum	2.12			30.72	44.81	1.51	20.84
C. reticulata raw			0.78	24.76	64.15	0.58	9.735
C. reticulata distilled				9.47	87.19		3.34
D. albus (surrogate)						96.51	3.94
ULSD	58.74		13.86	3.29	3.33	13.64	

Table 9. Chemical composition (%) of plant oils divided into biochemical classes based on biosynthesis pathways.

	Monoterpene (C_{10})	Sesquiterpene (C_{15})	Short-chain Alkane	Phenylpropanoids
C. flexuosus	90.33			
C. martinii	94.63	2.35		
P. resiniferum	44.81	30.72	2.12	1.51
C. reticulata raw		90.67		
C. reticulata distilled		96.66		
D. albus				96.51
ULSD			75.89	13.63

Table 10. Standard biodiesel fuel properties of B20 plant-derived oils mixed with ultra-low sulfur diesel.

Plant species of B20 Blend	Specific gravity ($g\ mL^{-1}$)	Sulfated Ash (% mass)	Copper Corrosion (hour)	Ramsbottom Carbon Residue (% mass)	Acid Number (mg KOH g^{-1})	Oxidation Stability by Rancimat (hour)	Cold Soak Filtration (s)
ULSD (control)	0.85	<.005	1A 3 hours	0.09	0.01	>12	74
***C. martinii* (surrogate)**	0.85	<.005	1A 3 hours	0.1	0.26	1.3*	61
***C. flexuosus* (surrogate)**	0.86	<.005	1A 3 hours	0.38*	0.2	>24	38
P. resiniferum	0.85	<.005	1A 3 hours	0.08	0.04	0.8*	70
***Copaifera* raw**	0.87	0.014	1A 3 hours	0.24	8.62*	>12	720*
***Copaifera* steam distilled**	0.86	<.005	1A 3 hours	0.14	0.01	0.2*	88
***D. albus* (surrogate)**	0.88	<.005	1A 3 hours	0.1	0.01	1.1*	43
B20 Standard			≤3 hr	≤0.35%	≤0.30 mg KOH	≥6 hr	≤360 s

***** denotes failed ASTM International test

Figures

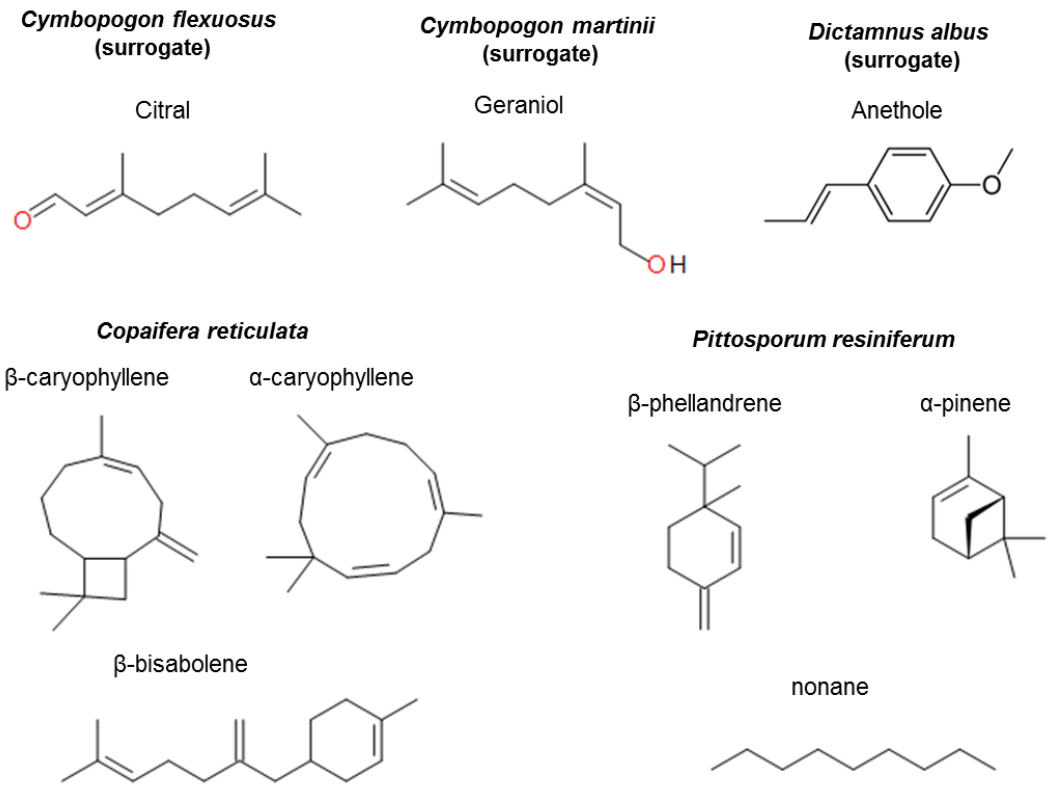

Figure 5. Biochemicals present in each plant-derived oil as determined by GC/MS.

Figure 6. Distillation ranges determined by ASTM International testing for B20 blends of plant-derived oils.

Most B20 blends had a distillation profile similar to ULSD except for *Pittosporum resiniferum* and *Cymbopogon martinii*. These two B20 blends consisted of biochemicals that distilled at lower temperature than ULSD or the rest of the plant-derived oils.

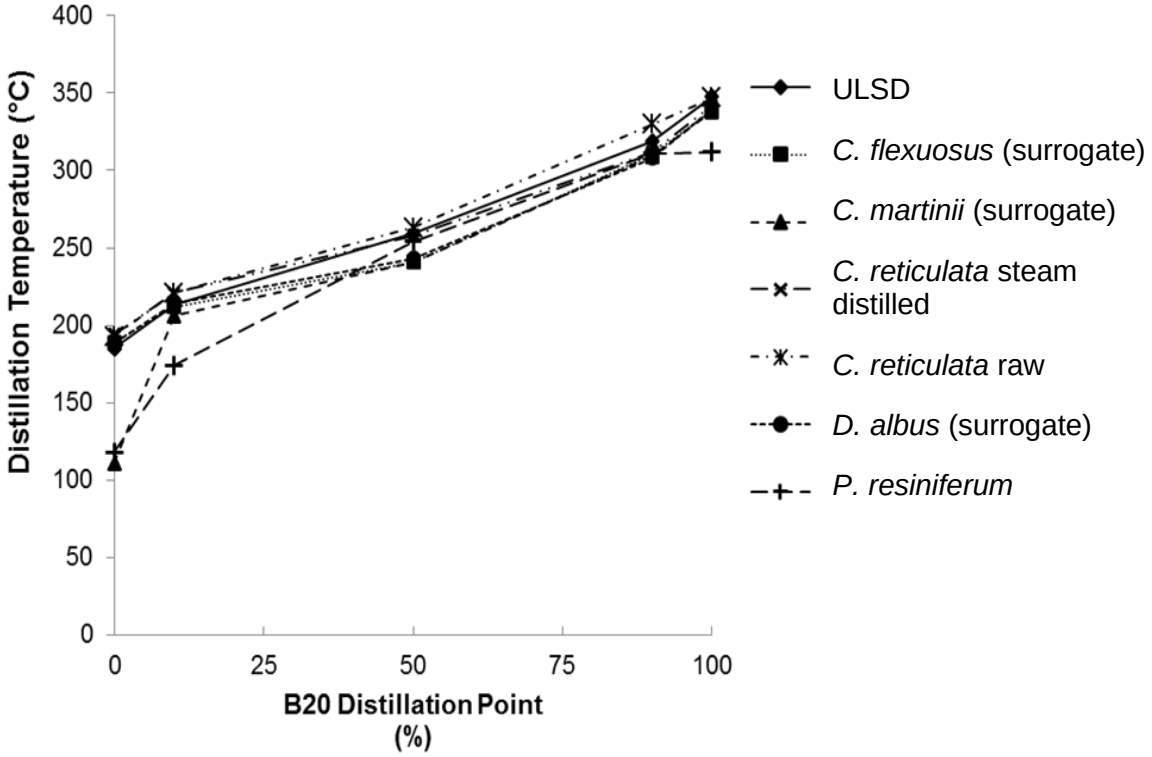
400
350
300
250
200
150
100
50
0
Distillation Temperature (°C)
0
25
50
75
100
B20 Distillation Point (%)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 6.

Figure 7. B20 blend biodiesel fuel properties as determined by ASTM International cetane number (A), cloud point (B), flash point (C), net heat of combustion (D), viscosity (E), and sulfur content (F) tests.

A) *Cymbopogon martinii*, *Copaifera reticulata* raw, and *D. albus* B20 blends failed the cetane number test (cetane number lower than 40 represented by the solid line). B20 blends of *Cymbopogon flexuosus*, *Copaifera reticulata* steam distilled, and *P. resiniferum* had a higher cetane number than the ULSD control. B) All B20 blends had a similar or lower cloud point than ULSD. This suggests the B20 blends would have similar or better cold temperature operability. C) All B20 blends had a flash point high enough to pass ASTM International testing except for *P. resiniferum*. *P. resiniferum* flash point would be considered too low for safe storage with other biodiesel fuels. D) All B20 blends had a lower net heat of combustion than ULSD except for *C. reticulata* raw. The B20 blends lowered net heat of combustion 1000 kJ kg^{-1} to 2000 kJ kg^{-1} which is roughly 50 kJ kg-1 to 100 kJ kg^{-1} for each biofuel blend percent. E) All B20 blends passed kinematic viscosity testing except for *C. reticulata* raw. F) All B20 blends passed sulfur content testing. *C. flexuosus*, *D. albus*, and *P. resiniferum* reduced sulfur content lower than predicted as these plant-derived oils have a lower viscosity (density) than ULSD control which diluted the sulfur content of the ULSD blending stocks.

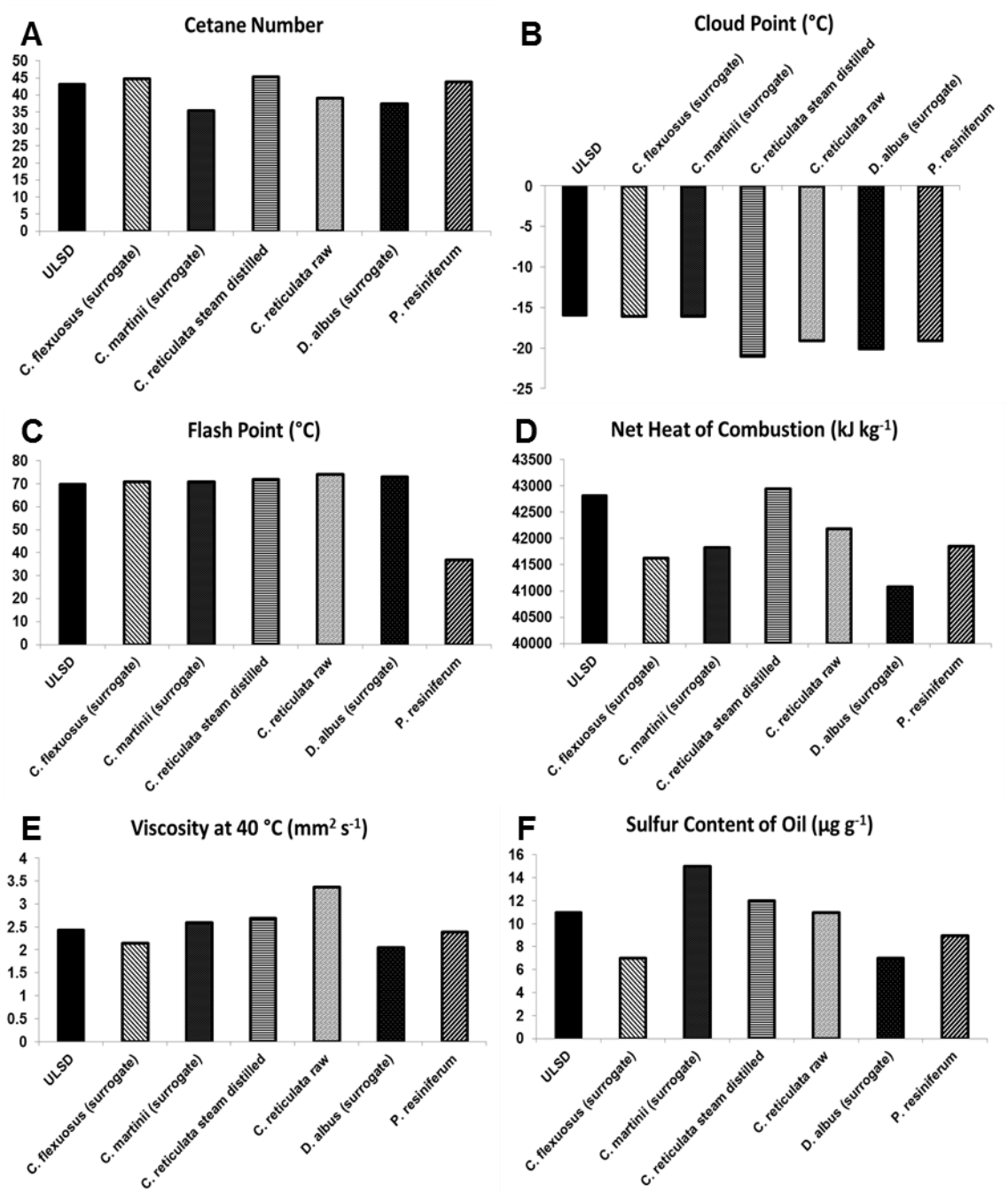
A
Cetane Number
50
45
40
35
30
25
20
15
10
5
0
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum
B
Cloud Point (°C)
0
-5
-10
-15
-20
-25
C
Flash Point (°C)
80
70
60
50
40
30
20
10
0
D
Net Heat of Combustion (kJ kg-1)
43500
43000
42500
42000
41500
41000
40500
40000
E
Viscosity at 40 °C (mm2 s-1)
4
3.5
3
2.5
2
1.5
1
0.5
0
F
Sulfur Content of Oil (µg g-1)
16
14
12
10
8
6
4
2
0

Figure 7.

Figure 8. Air intake temperatures used to modify combustion phasing (MFB50) in homogenous charge compression ignition engine modeling to achieve a range of before and after top dead center (represented as 0 °CA) (A), the range of air intake temperatures used for each B20 blend (B), and relation of B20 blend cetane number to combustion phasing.

(A) B20 blends had MFB50 at air intake temperatures similar to the ULSD control except for *Cymbopogon flexuosus* and *C. martinii* which required lower air intake temperatures to modify combustion phasing. (B) *C. flexuosus* and *C. martinii* B20 blends had lower temperature combustion ranges than all other B20 blends and ULSD control. Diamonds in the range represent MFB50 at top dead center. C) No relationship between cetane number and combustion phasing (represented as air intake temperature to achieve MFB50 at top dead center) was found. *Cymbopogon flexuosus* and *C. martinii* B20 blends improved ignition and had a similar combustion phasing despite having different cetane numbers. This suggests that chemical variation in the B20 blends changes combustion phasing.

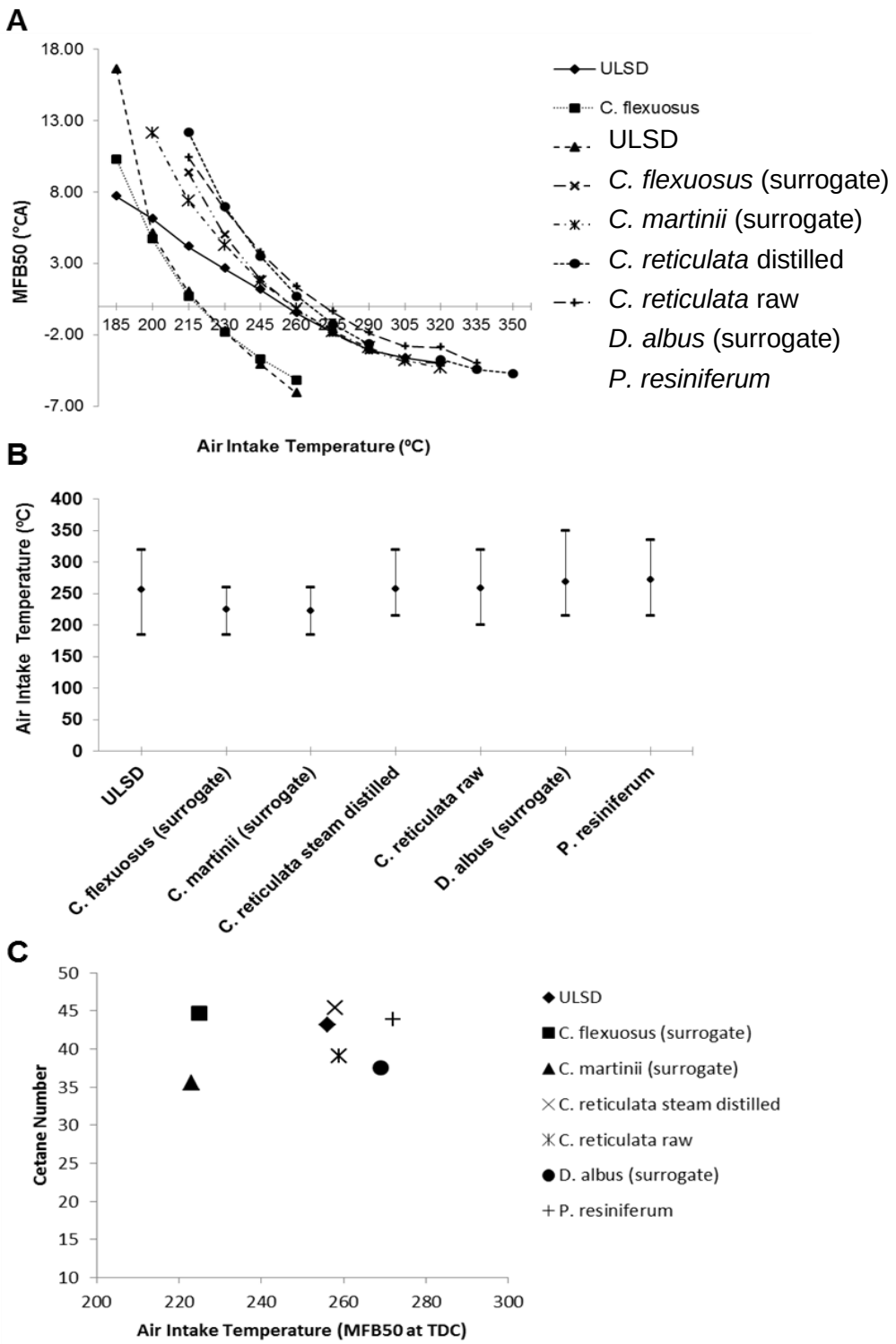
A
18.00
13.00
8.00
3.00
-2.00
-7.00
MFB50 (°CA)
185 200 215 230 245 260 275 290 305 320 335 350
Air Intake Temperature (ºC)
ULSD
C. flexuosus
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum
B
400
350
300
250
200
150
100
50
0
Air Intake Temperature (ºC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum
C
50
45
40
35
30
25
20
15
10
Cetane Number
200 220 240 260 280 300
Air Intake Temperature (MFB50 at TDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 8.

Figure 9. Indicated mean effective pressure (IMEP) controlled for energy content of B20 blends across combustion phasing (A), and indicated specific fuel consumption (ISFC) for B20 oil blends compared to ULSD control (B).

(A) Most B20 blends produced similar average in-cylinder pressure (work) as ULSD after controlling for the different energy content of each B20 blend. *Copaifera reticulata* steam distilled and *C. reticulata* raw produced more pressure (more work for each cycle) than ULSD. (B) *C. reticulata* steam distilled and *C. reticulata* raw B20 blends had a lower ISFC than ULSD that corresponds to the higher average pressure produced in-cylinder. The maximum in-cylinder pressure for all B20 blends was produced near 5 ° after top dead center and so the minimum fuel consumption was also near 5 °ATDC.

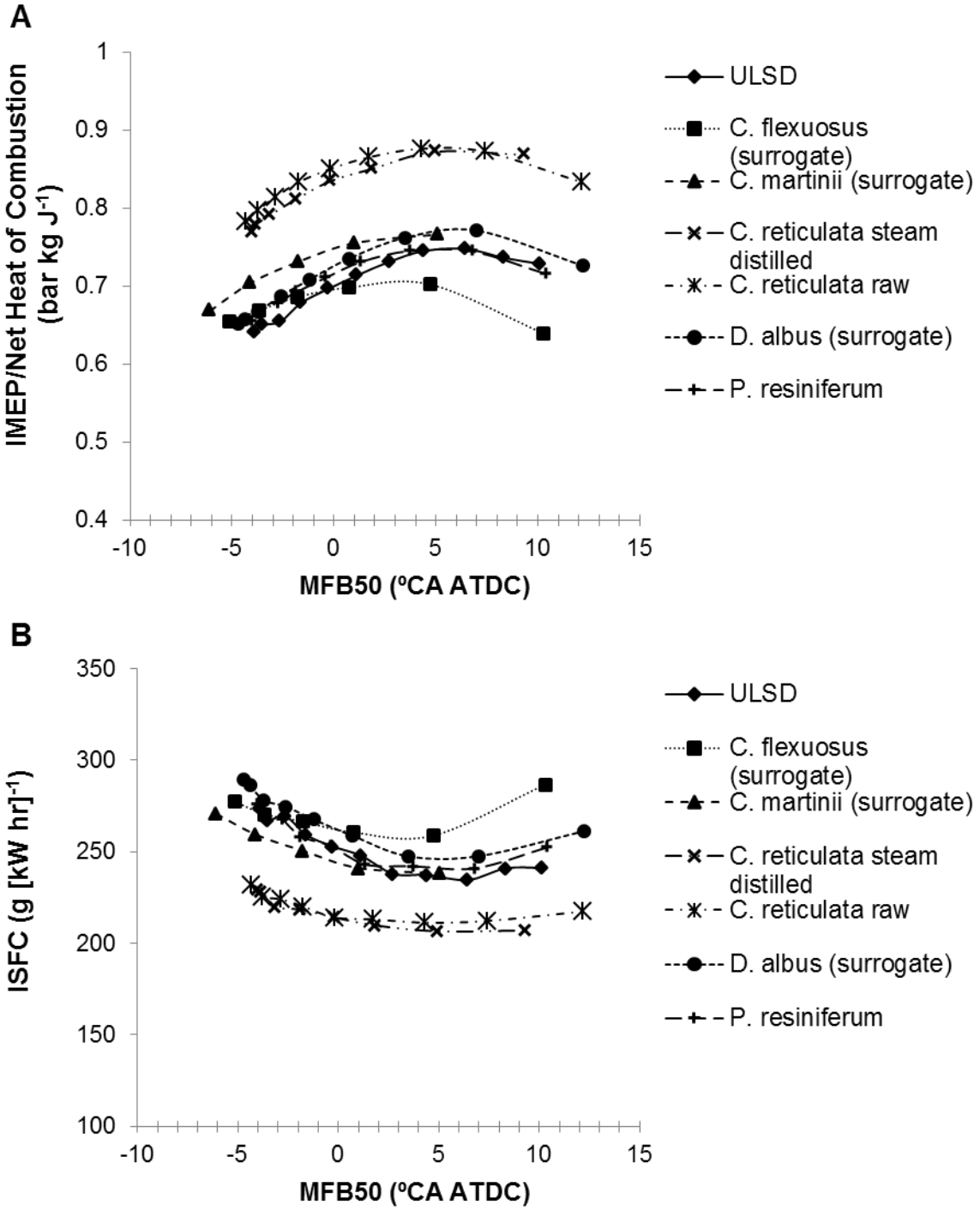
A
IMEP/Net Heat of Combustion (bar kg J-1)
1
0.9
0.8
0.7
0.6
0.5
0.4
-10
-5
0
5
10
15
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum
B
ISFC (g [kW hr]-1)
350
300
250
200
150
100
-10
-5
0
5
10
15
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 9.

Figure 10. Total length of combustion from 10% mass fraction burned to 90% mass fraction burned (MFB10-90) for B20 oils blends in relation to combustion phasing from top dead center (MFB50).

Cymbopogon flexuosus and *C. martinii* had the longest combustion events across all MFB50. All other B20 blends had similar combustion lengths similar to ULSD.

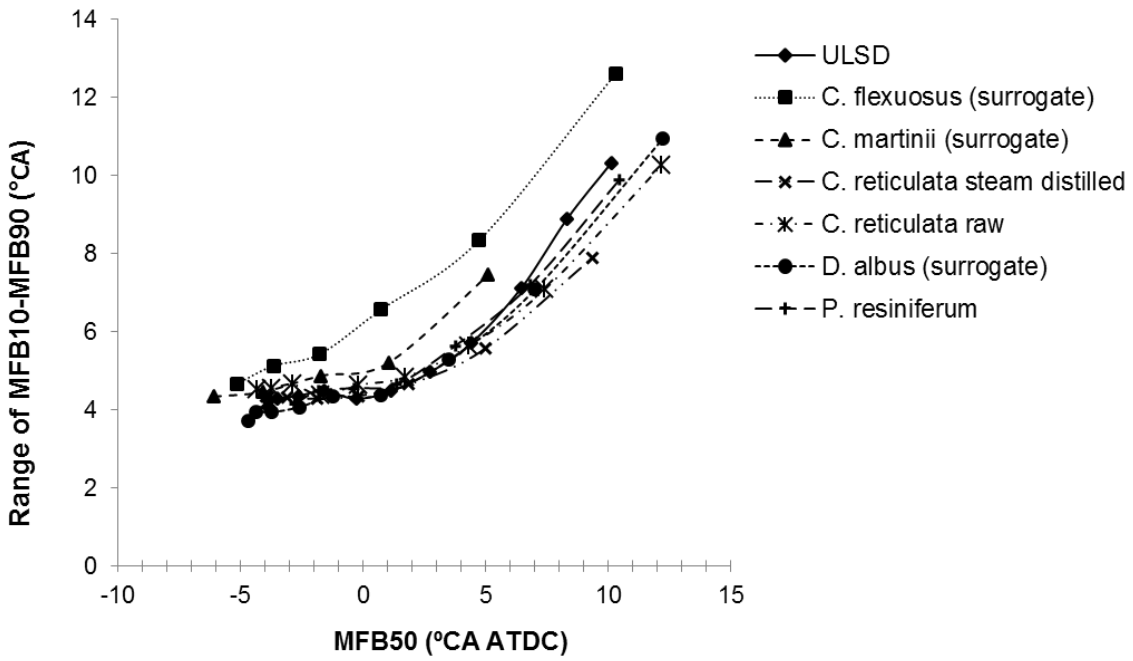
14
12
10
8
6
4
2
0
-10
-5
0
5
10
15
Range of MFB10-MFB90 (°CA)
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 10.

Figure 11. Maximum rate of pressure increase in the cylinder (dP/dCA) at different combustion phasing (MFB50).

Cymbopogon flexuosus and *C. martinii* had the lowest rate of in-cylinder pressure increase while *Copaifera reticulata* steam distilled and *C. reticulata* raw had the highest rate which was similar to ULSD control.

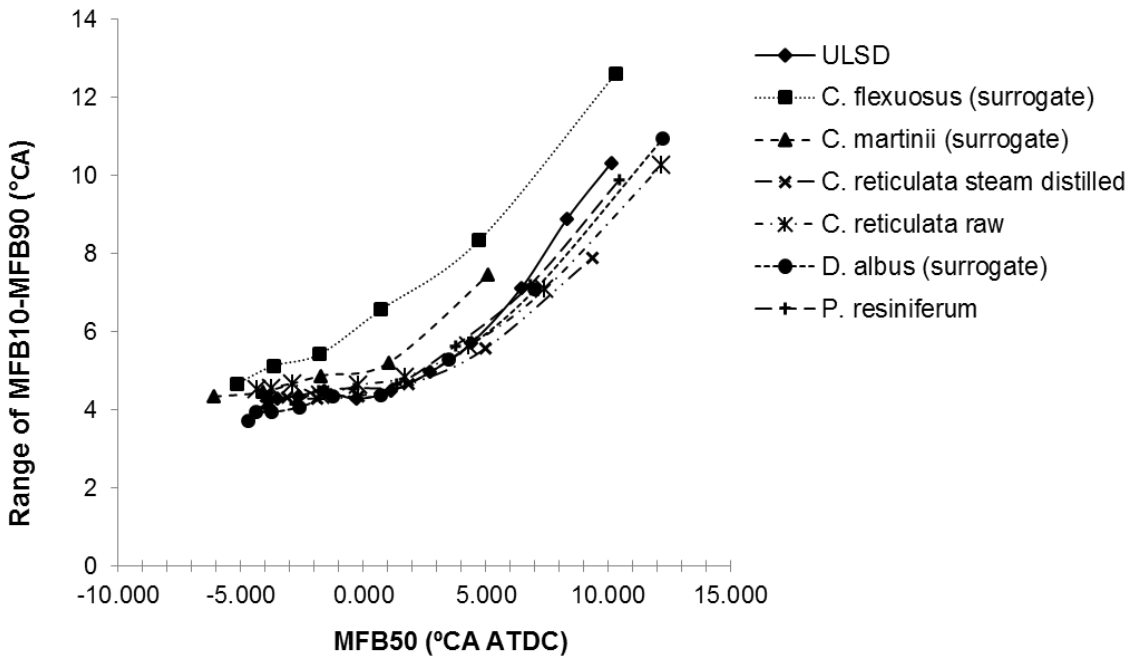
14
12
10
8
6
4
2
0
-10.000
-5.000
0.000
5.000
10.000
15.000
Range of MFB10-MFB90 (°CA)
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 11.

Figure 12. Unburned hydrocarbon emissions from B20 oil blends in relation to combustion phasing.

All B20 blends had higher hydrocarbon emissions than ULSD except for *Copaifera reticulata* steam distilled and *C. reticulata* raw which were equivalent to ULSD control or had lower hydrocarbon emissions. Unburned hydrocarbon emissions suggest that the all the B20 blends except for *C. reticulata* did not combust as efficiently as ULSD at the engine settings.

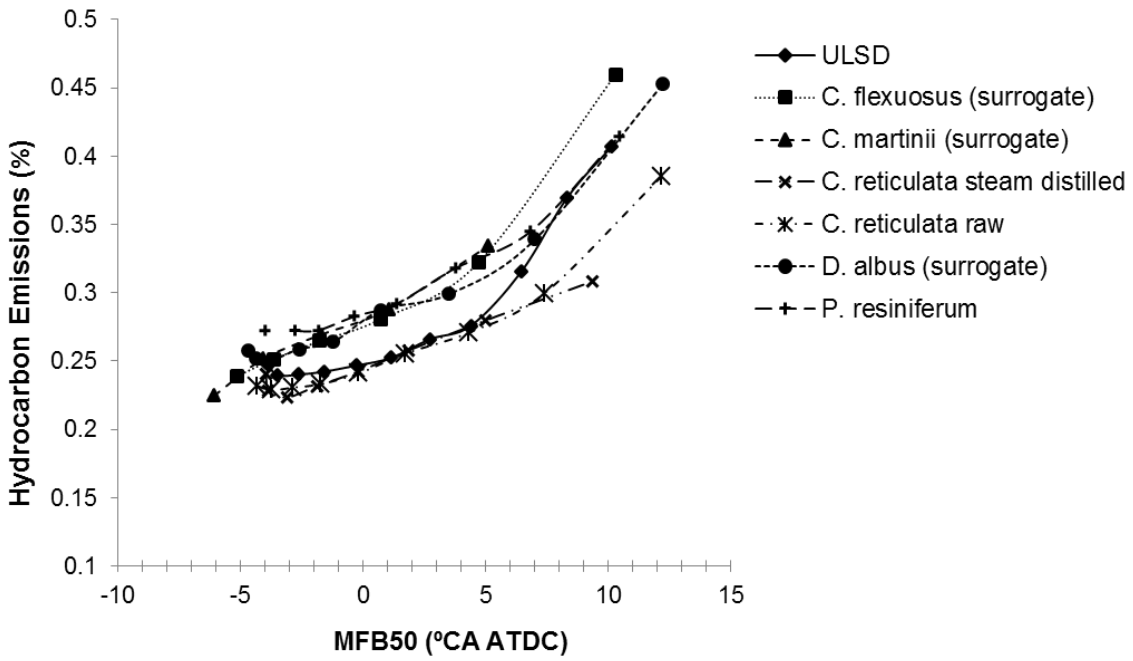
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
Hydrocarbon Emissions (%)
-10
-5
0
5
10
15
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 12.

Figure 13. Nitrous oxide (NOx) emissions in relation to MFB50 (A) and air intake temperature (B) of B20 oil blends.

A) All B20 blends produced similar amounts of NOx emissions to ULSD control except for *Cymbopogon flexuosus* and *C. martinii* which produced less NOx emissions. B) The rate of NOx emissions increased at air intake temperature 260 °C for all B20 blends except for *P. resiniferum* which increased after 270 °C. *Copaifera reticulata* steam distilled and raw produced higher NOx emissions across all temperatures after 260 °C than ULSD while *D. albus* had similar NOx emissions. *Cymbopogon flexuosus* and *C. martinii* produced lower NOx emissions than all other B20 blends because their range of intake temperatures was lower than all other B20 blends.

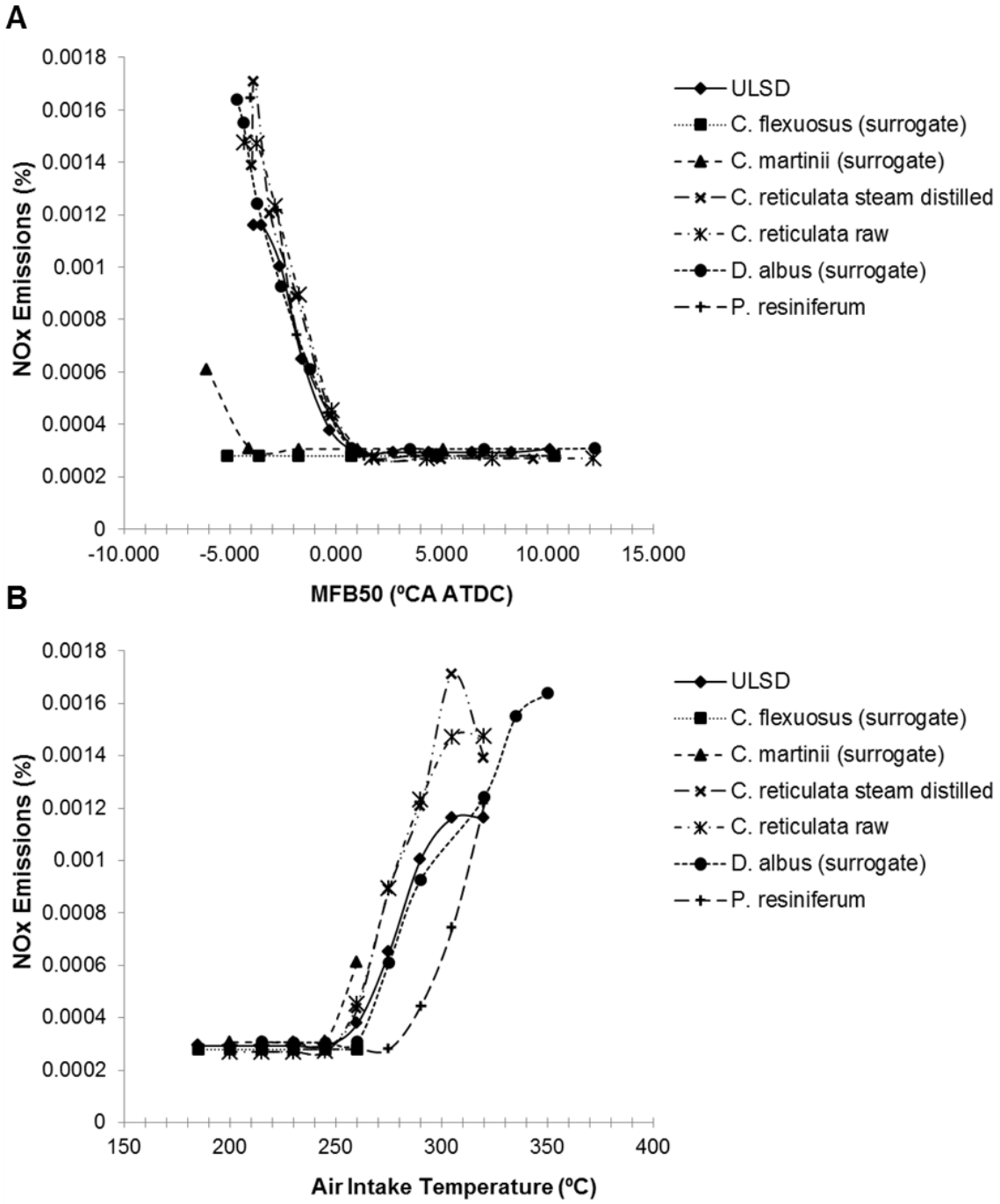
A
0.0018
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
NOx Emissions (%)
-10.000
-5.000
0.000
5.000
10.000
15.000
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum
B
0.0018
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
NOx Emissions (%)
150
200
250
300
350
400
Air Intake Temperature (ºC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 13.

Figure 14. Carbon monoxide emissions in relation to combustion phasing of B20 oil blends and ULSD control.

CO emissions across the B20 blends were variable. *Copaifera reticulata steam* distilled, *C. reticulata* raw, and *D. albus* B20 blends had lower CO emissions than ULSD controls while *Cymbopogon flexuosus* had a higher CO emission than ULSD.

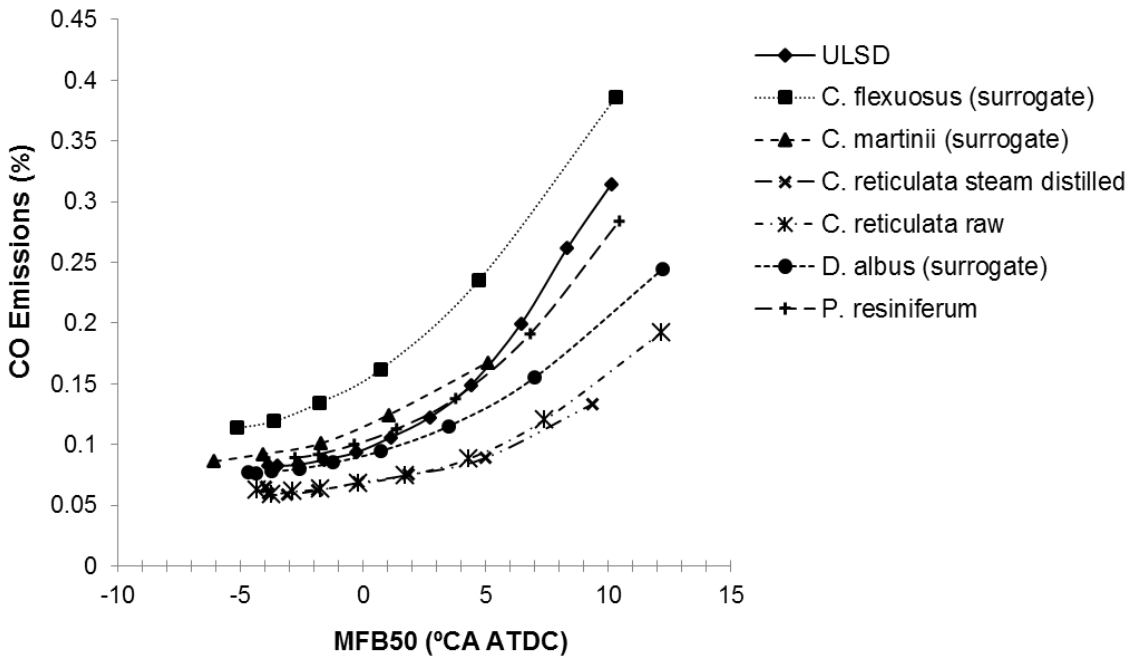
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
CO Emissions (%)
-10
-5
0
5
10
15
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 14.

Figure 15. Soot emissions (filter smoke number or mg m^{-3}) in relation to combustion phasing of B20 oil blends.

Soot emissions were sporadic, but formed two groups overall. *Cymbopogon martinii*, *D. albus*, and *P. resiniferum* had higher soot emissions than ULSD while *Cymbopogon flexuosus*, *Copaifera reticulata* steam distilled, and *C. reticulata* raw B20 blends had a similar soot emission to ULSD control.

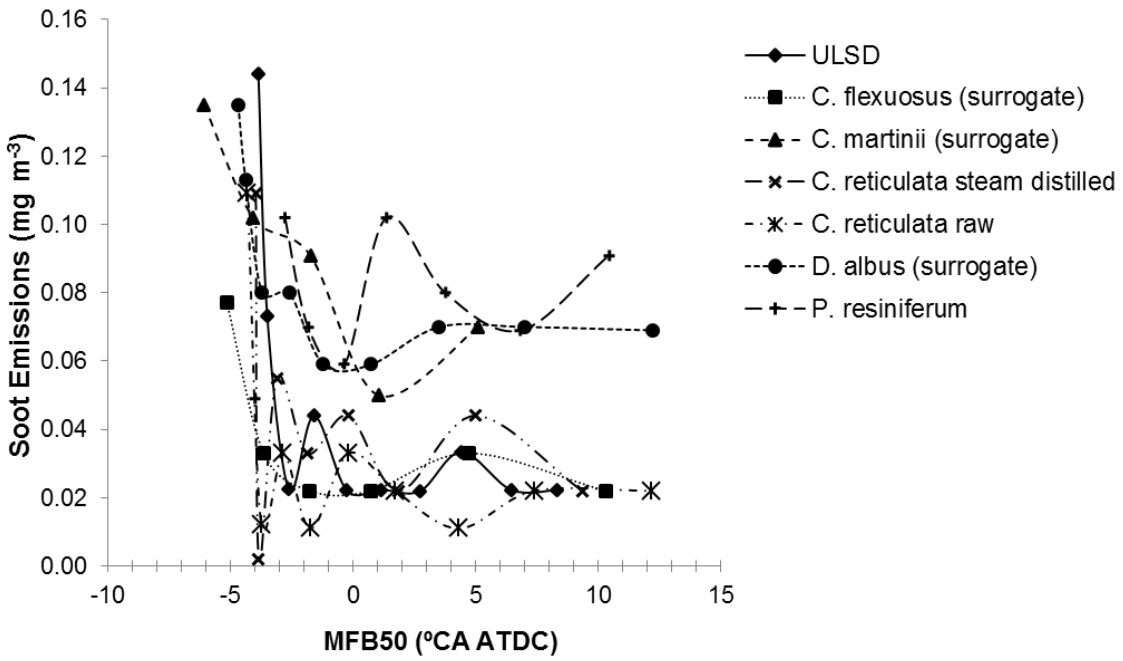
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Soot Emissions (mg m-3)
-10
-5
0
5
10
15
MFB50 (ºCA ATDC)
ULSD
C. flexuosus (surrogate)
C. martinii (surrogate)
C. reticulata steam distilled
C. reticulata raw
D. albus (surrogate)
P. resiniferum

Figure 15.

Additional Figures

Figure 16. GC/MS chromatograms of plant-derived oils used in B20 blends and the ULSD control.

In most cases the chromatograms were too complex to label directly. Identified biochemicals are presented in Table 11.

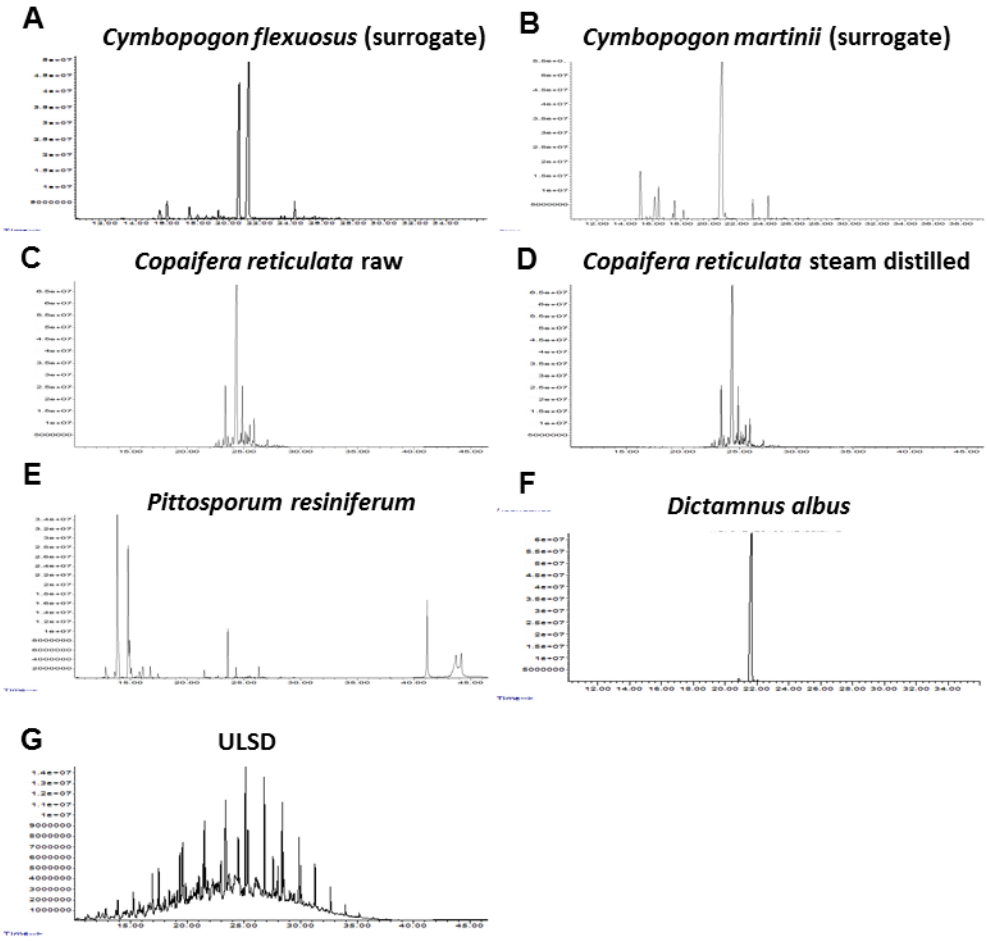
A
Cymbopogon flexuosus (surrogate)
B
Cymbopogon martinii (surrogate)
C
Copaifera reticulata raw
D
Copaifera reticulata steam distilled
E
Pittosporum resiniferum
F
Dictamnus albus
G
ULSD

Figure 16.

Additional Tables

Table 11. Chemicals identified in plant-derived oils used in blending stocks and ULSD.

***Cymbopogon flexuosus* (surrogate)**	**Peak Area (% total)**	**Petroleum fuel class**
1-methyl-2-(1-methylethyl)-Benzene	4.35	olefin
(Z)-Citral	33.14	olefin
Citral	52.84	olefin
***Cymbopogon martinii* (surrogate)**	**Peak Area (% total)**	**Petroleum fuel class**
linalool	2.85	olefin
alloocimene	1.33	olefin
nerol	65.96	olefin
citral	1.05	olefin
ocimene	5.53	olefin
beta-myrcene	11.63	olefin
3-carene	5.39	bicyclic naphthene
4-carene	0.89	bicyclic naphthene
***Copaifera reticulata* raw**	**Peak Area (% total)**	**Petroleum fuel class**
copaene	12.21	bicyclic naphthene
clovene	0.92	bicyclic naphthene
caryophyllene	30.05	bicyclic naphthene
alpha-selinene	9.09	bicyclic naphthene
aromadendrene	0.9	bicyclic naphthene
gamma-muurolene	3.04	bicyclic naphthene
alpha-amorphene	1.07	bicyclic naphthene
delta-cadinene	6.82	bicyclic naphthene
delta-selinene	0.45	bicyclic naphthene
delta-elemene	0.38	monocyclic naphthene
beta-elemene	3.52	monocyclic naphthene
alpha-caryophyllene	5.27	monocyclic naphthene
alpha-bisabolene	1.53	monocyclic naphthene
beta-bisabolene	14.06	monocyclic naphthene
farnesene	0.78	isoparaffin
alpha-calacorene	0.58	aromatic

Table 11. Continued.

***Copaifera reticulata* steam distilled**	**Peak Area (% total)**	**Petroleum fuel class**
copaene	9.07	bicyclic naphthene
beta-caryophyllene	58.04	bicyclic naphthene
delta-cadinene	4.49	bicyclic naphthene
caryophyllene	2.26	bicyclic naphthene
beta-cubebene	1.52	bicyclic naphthene
beta-selinene	1.36	bicyclic naphthene
isocaryophyllene	3.34	bicyclic naphthene
gamma-cadinene	0.83	bicyclic naphthene
caryophyllene oxide	0.98	bicyclic naphthene
clovene	1.52	bicyclic naphthene
alpha-amorphene	2.22	bicyclic naphthene
gamma-muurolene	0.64	bicyclic naphthene
alpha-cubebene	0.92	bicyclic naphthene
alpha-caryophyllene	8.82	monocyclic naphthene
delta-elemene	0.65	monocyclic naphthene
***Dictamnus albus* (surrogate)**	**Peak Area (% total)**	**Petroleum fuel class**
Anethole	96.51	aromatic
Pittosporum resiniferum	**Peak Area (% total)**	**Petroleum fuel class**
nonane	2.12	paraffin
alpha-pinene	38.95	bicyclic naphthene
beta-pinene	5.86	bicyclic naphthene
beta-phellandrene	22.41	monocylic naphthene
limonene	2.14	monocylic naphthene
gamma-terpinene	1.27	monocylic naphthene
beta-elemene	4.9	monocylic naphthene
ULSD	**Peak Area (% total)**	**Petroleum fuel class**
nonane	1.17	paraffin
decane	1.8	paraffin
2-methyl-nonane	0.65	paraffin
decane	2.23	paraffin
4-methyl decane	1.27	paraffin
undecane	2.73	paraffin
undecane	2.64	paraffin
decane	0.74	paraffin
dodecane	2.58	paraffin
dodecane	2.91	paraffin
Tridecane, 7-methyl	1.74	paraffin

Table 11. Continued.

ULSD	Peak Area (% total)	Petroleum fuel class
Tridecane	3.58	paraffin
Tridecane	3.47	paraffin
tetradecane	3.69	paraffin
pentadecane	8.12	paraffin
Hexadecane	6.19	paraffin
heptadecane	4.94	paraffin
2-methyl-tridecane	0.65	paraffin
octadecane	3.37	paraffin
eicosane	1.41	paraffin
heneicosane	0.64	paraffin
nonadecane	2.22	paraffin
Undecane, 2,6-dimethyl	1.29	isoparaffin
Decane, 3,7-dimethyl	0.41	isoparaffin
Hexadecane, 2,6,10,14-tetramethyl	0.71	isoparaffin
Heptadecane, 2,6,10,15-tetramethyl	3.00	isoparaffin
Pentadecane, 2,6,10-trimethyl	2.26	isoparaffin
Pentadecane, 2,6,10,14-tetramethyl	3.96	isoparaffin
Hexadecane, 2,6,10,14-tetramethyl	2.23	isoparaffin
2-methyl-propyl-cyclohexane	1.00	monocyclic naphthene
butyl-cyclohexane	1.03	monocyclic naphthene
octyl-cyclohexane	0.60	monocyclic naphthene
3-ethylcyclopentanone	0.66	monocyclic naphthene
xylene	0.93	aromatic
1-ethyl-4-methyl-benzene	0.99	aromatic
1,2,3-trimethyl-benzene	0.9	aromatic
1-methyl-3-propyl-benzene	0.84	aromatic
Naphthalene, 1,2,3,4-tetrahydro-6-methyl	0.73	aromatic
Naphthalene, 1,2,3,4-tetrahydro-5-methyl	0.6	aromatic
Benzene, 1,1'-ethylidenebis	5.69	aromatic

Table 11. Continued.

ULSD	Peak Area (% total)	Petroleum fuel class
Bacchotricuneatin c	0.77	aromatic
Naphthalene, 1,2,3-trimethyl-4-pro	2.18	aromatic
trans-Decalin, 2-methyl	0.66	bicyclic naphthene
trans-Decalin, 2-methyl-	1.21	bicyclic naphthene
Naphthalene, decahydro-2-methyl	1.46	bicyclic naphthene

CHAPTER III

DE NOVO TRANSCRIPTOME MINING OF COPAIFERA OFFICINALIS AND FUNCTIONAL CHARACTERIZATION OF SESQUITERPENE SYNTHASES IN COPAIFERA OFFICINALIS AND COPAIFERA LANGSDORFFII

This chapter was formatted for submission to BMC Genomics. B.L. Joyce was responsible for annotating transcriptome contigs, phylogenetic analysis of terpene synthases, assembly and motif analysis, isolating and cloning *Copaifera* terpene synthase genes, recombinant expression and functional characterization of terpene synthase genes using *in vitro* assays, and preparing the manuscript. H. Al-Ahmad was involved in designing the project, obtained *Copaifera* germplasm to study, isolated and cloned *Copaifera* terpene synthase genes and revised the manuscript. Y. Peng assembled and annotated the 1000 Plant *Copaifera* transcriptome. P. Ranjan assembled and annotated the Texas A&M transcriptome. S. Liu sequenced the Texas A&M transcriptome. X. Sun sequenced the 1000 Plant transcriptome. J.C. John assisted in cloning and isolating *Copaifera* terpene synthases. G.K. Wong directed the 1000 Plant transcriptome project. F. Chen provided expertise in terpene synthase functional characterization and edited the manuscript. J.S. Yuan was involved in designing the project and edited the manuscript. C.N. Stewart, Jr was involved in designing the project, provided funding, and edited the manuscript.

B.L. Joyce[1], H. Al-Ahmad[2], Y. Peng[1], S. Liu[3], P. Ranjan[1], X. Sun[4], J.C. John[1], G.K. Wong[4 5 6], F. Chen[1], J.S. Yuan[3], and C.N. Stewart, Jr.[1§]

[1]University of Tennessee, Department of Plant Sciences, Knoxville, TN 37996, USA

[2]An-Najah National University, Department of Biology & Biotechnology, Nablus, Palestine

[3]Texas A&M University, Plant Pathology and Microbiology, College Station, TX 77843, USA

[4]Beijing Genomics Institute-Shenzhen, Bei Shan Industrial Zone, Yantian District, Shenzhen 518083, China

[5]University of Alberta, Department of Biological Sciences, Edmonton, AB T6G 2E9 Canada

[6]University of Alberta, Department of Medicine, Edmonton, AB T6G 2E9 Canada

ABSTRACT

Background

Copaifera species, often referred to as the ‘diesel trees,’ are native to the Americas produce a sesquiterpene-rich oleoresin that is collected by tapping the trunks of mature trees. While the oleoresin has been used as traditional medicine in parts of Central and South America, the oleoresin has also been reportedly used in diesel engines as fuel. While production of biofuels from the *Copaifera* genus might currently not be economically feasible, the characterization of their transcriptomes allows for gene discovery related to the unique biosynthetic pathway in *Copaifera* species, which, in turn, should improve our understanding of the copious production of sesquiterpene oleoresins present in these species.

Results

Here we describe the *de novo* assembled transcriptomes and functional characterization of the *Copaifera officinalis* sesquiterpene biosynthetic pathway. Terpene synthases of *C. officinalis* and *Copaifera langsdorffii* contained class I and class II terpene synthase motifs previously only found in bifunctional class I/II gymnosperm diterpene synthases. The majority of terpene synthases characterized had both mono- and sesquiterpene synthase activity. The sesquiterpene synthases responsible for biosynthesis of the largest percentage of *C. officinalis* oleoresins were functionally characterized.

Conclusions

The *de novo* transcriptome of the northernmost *Copaifera* New World species, *C. officinalis*, was robust enough to isolate functional sesquiterpene synthase genes in *C. langsdorffii*, a species native to the southernmost range of the *Copaifera* genus in the New World. *C. officinalis* sesquiterpene synthase gene motifs suggest a link between ancestral monoterpene and diterpene oleoresin production in gymnosperms and sesquiterpene oleoresins present in angiosperm tree species. Further sequencing and functional characterization could lead to determining the evolutionary shift from diterpene-based oleoresin present in African *Copaifera* species to the sesquiterpene-based oleoresin present in New World *Copaifera* species. It could also aid practical use of this metabolic pathway for biofuel.

Background

The focus of terpenoid biosynthesis biochemistry research in plants has been primarily restricted to three plant systems: 1) gymnosperms such as *Abies* (spruce), *Taxus* (yew), and *Pinus* genera, 2) herbaceous dicot genera such as *Artemisia* (Caretto et al., 2011; Kirby & Keasling, 2009), members of the Lamiaceae (mint) family (specifically *Salvia, Perilla, Lavandula, Ocimum*, and *Mentha*), and Arabidopsis (Degenhardt et al., 2009; Lange et al., 2011); and 3) crop species such as *Zea mays,* cotton (*Gossypium spp*.), tomato (*Lycopersicon* spp.), and rice (*Oryza sativa*). The unifying theme across all of these studies and plant species is economic importance of terpene biosynthesis. *Pinus* and *Abies* are economically-important genera for lumber production, and were once important for production of turpentine and rosin which are comprised of mono- and diterpenes (Mirov, 1952). The *Taxus baccata* diterpene biosynthesis pathway has been studied to produce the economically important anticancer drug taxol in recombinant microbial systems or *T. baccata* suspension cell cultures (Malik et al., 2011). Likewise, *Artemisia annua* has been characterized to produce the antimalarial drug artmesinin (Weathers et al., 2011). Finally, many characterized terpene synthesis genes are found in either crop species such as corn, rice, and cotton, or in small-production specialty crops such as basil, spearmint, and lavender. Looking across all green land plants, functional characterization of terpenoid biosynthesis genes comes from a narrow

phylogenetic range comprising only 9 families contained solely within two out of five divisions of seed plants (Pinophyta and Magnoliophyta) (Chen et al., 2011).

Recent genome and transcriptome sequencing projects of species outside this narrow range have provided insight into the origin and evolution of terpene biosynthesis in land plants, including byrophytes, e.g., *Physcomitrella patens* (Rensing et al., 2008) and the spikemoss *Selaginella moellendorffii* (Banks et al., 2011). In addition to *P. patens* and *S. moellendorfii*, the *Vitis vinifera* (Zharkikh et al., 2008), *Populus trichocarpa* (Tuskan et al., 2006) and *Sorghum bicolor* (Chen et al., 2011; Paterson et al., 2009) terpene synthase (TPS) amino acid sequences derived from genomes have been compared which resulted in new terpene synthase subfamily definitions and relations between subfamilies.

The genus *Copaifera*, which includes species that have been commonly referred to as 'diesel trees,' produces a sesquiterpene-rich oleoresin that can be collected when the trunk is drilled into and “tapped” (Calvin, 1980). This oleoresin resin has been used in traditionally medicines and anecdotally as fuel (Calvin, 1980). This led to interest in *Copaifera* trees to produce biofuels but economic botany studies have found trees to produce only 0.459 L yr^{-1} to 0.80 L yr^{-1} on average and many trees produced little or no oleoresin on a second tapping (Plowden, 2004; Plowden, 2003). However, terpenes (isoprenoids) still remain an important target for the second generation of biofuels from biomass-derived

metabolites (Kirby & Keasling, 2009; Lee et al., 2008; Peralta-Yahya & Keasling, 2010).

New World Copaifera species are one of many angiosperm tree genera known to produce terpenoid oleoresins that are used as economically-important natural products around the world (Langenheim, 2003). *Copaifera* oleoresin production has some features in common with that of the gymnosperm genera *Pinus*, *Picea*, and *Taxus*, which produce and store oleoresin in specialized resin canals (Calvin, 1980). However, *Copaifera* oleoresins are composed of sesquiterpenes, whereas gymnosperm oleoresins are almost universally composed of mono- and diterpenoids. *Pinus* and *Picea* mono- and diterpenoid biosynthesis has been thoroughly studied with over 50 terpene synthases functionally characterized (Keeling & Bohlmann, 2006a; Keeling et al., 2011). *Copaifera* species native to the New World produce oleoresins comprised primarily of sesquiterpenoids, whereas African species produce diterpene-rich oleoresins (Langenheim, 1973). The mechanisms responsible for this geographic variation could be as simple as loss of transcript signal peptides or as complex as gene duplication and neofunctionalization events (Benderoth et al., 2006; Chen et al., 2011). Investigating the evolution of oleoresin biosynthesis across gymnosperm and angiosperm tree species could lead to a better understanding of plant-defense evolution.

Additionally, the copious volume of sesquiterpenes produced in *Copaifera* is a unique phenotype. Typically there are only trace amounts of mono-, sesqui-, or diterpenes in plant tissues. Even species that have been used for production of terpenoids have low individual yields with an average of 12 mL yr^{-1} in *Pinus taeda* (Hodges et al., 1977) and an average of 69 mL yr^{-1} for *Pinus elloiti* (Rodrigues & Fett-Neto, 2009). However, the features of terpenoid biosynthesis is a complex phenotype and most likely results from several factors that could include the basic physiology of the trees, transportation and storage of metabolites, and metabolic pathway characteristics, such as expression of enzymes, greater carbon flux into the pathway, and/or increased enzymatic efficiency of terpene production.

The main purpose of this project was to functionally investigate the novel terpenoid biosynthesis pathway of *C. officinalis* using a *de novo* transcriptome analysis to identify transcripts involved in the sesquiterpene biosynthesis pathway of *C. officinalis*. This initial characterization is meant to support future work to investigate the details of the *Copaifera* terpene biosynthesis pathway to assess if it could be used to increase sesquiterpene production in other plant species. Additionally, the *C. officinalis* transcriptome was also used to investigate terpene synthases in *C. langsdorffii* which is native to the southernmost range of the *Copaifera* genus in southern Brazil (Veiga Junior & Pinto, 2002). Characterization of *C. langsdorffii* terpene synthases was carried out to

determine whether species across the *Copaifera* genus could be characterized from only a few *de novo* transcriptomes. This will support future work to characterize the geographical variation observed in New World and African *Copaifera* species.

Results

C. officinalis transcriptome sequencing, assembly, and annotation

A total of 25 626 055 reads containing a total of 1.875 Gbp of sequence was generated by the 1000 Plants (1KP) initiative that passed trimming and quality score filtering (Table 11). Of these contigs, 9% did not assemble. Roughly half of the paired-end reads that were assembled remained paired-end and the other half were broken and only one side of the paired-end sequencing were used. This could also result from poor sequencing quality or insufficient sequencing coverage. A total of 35 940 contigs were generated from the 1KP assembly and 11 417 were generated from at Texas A&M University (TAMU) sequencing for annotation. The contigs from 1KP were longer on average and across all the contigs available were longer than TAMU transcripts represented by their N75, N50, and N25 scores.

Both assemblies were annotated using the Kyoto Encylopedia of Genes and Genomes (KEGG) but only 22% of the 1KP contigs and 30% of the TAMU contigs were annotated. The 1KP transcriptome returned twice as many total

annotated contigs than the TAMU transcriptome (Table 11). The 1KP contig annotation had more pathway enzymatic steps and unique contigs (unigenes) for every metabolic group except for glycan, cofactor biosynthesis, and secondary metabolism (Figure 16). The TAMU transcriptome contained more unigenes for secondary metabolism overall, but less annotated contigs than the 1KP transcriptome specifically for terpenoid metabolism (Table 12).

Annotated mevalonic acid, methylerythritol phosphate, and terpene synthase pathway transcripts

Transcripts in both the mevalonic acid (MVA) and methyleryithritol phosphate (MEP) pathways were annotated in the *C. officinalis* transcriptome from the 1KP and TAMU datasets (Table 12). The annotated transcripts in the MVA and MEP pathways contained every known pathway enzyme (Table 12). The MVA and MEP pathways are responsible for synthesizing the precursors for terpene biosynthesis geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranyl geranyl pyrophosphate (GGPP) (Tholl, 2006).

A total of three contigs were identified from each transcriptome as putative monoterpene synthase genes. Contigs 5663 and 6446 were annotated as putative neo-menthol synthase genes and contig 8809 was identified as a putative β-ocimene/myrcene synthase. These three contigs were not cloned in the current work as the major constituents of *C. officinalis* oleoresins are sesquiterpenes.

Six putative sesquiterpene synthases (sesquiTPSs) were identified using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) annotation. Three of these putative terpene synthases were full length contigs which was determined by rapid amplification of cDNA ends (RACE) PCR. Two contigs were found to be partial contigs (Table 13). An additional 517 bp of *C. officinalis* TPS2 (CoTPS2) and another 365 bp of CoTPS5 was found during 5' RACE. Cloning and sequencing revealed a cryptic stop codon in the coding sequence of *C. officinalis* TPS6 (*CoTPS6*) and so it was ruled out as a putative TPS. Ultimately after RACE PCR, four contigs (CoTPS1-CoTPS4) were annotated by nucleotide sequence similarity as putative sesquiTPSs while CoTPS5 was identified as a putative diTPS.

The signal peptides of the five remaining putative CoTPS genes were predicted to further determine whether the genes were sesquiterpene synthases. WolfPsort, Multiloc2, iPSORT, SIG-Pred, and ESLPred algorithms were used for signal peptide prediction however the results were not conclusive (Table 13). The four putative sesquiTPS, CoTPS1-4, were predicted to by all algorithms to either have a chloroplastic signal peptide or no signal peptide (localization in cytoplasm). The putative diTPS *CoTPS5* was the only of the contigs to have uniform prediction of no signal peptide (Table 13).

Isolation of C. officinalis and C. langsdorffii complementary DNA

Primers specific to *CoTPS1, 2, 3, 4* and *5* were used to isolate the putative TPS from both *C. officinalis* and *C. langsdorffii* complementary DNA (cDNA). *CoTPS1, 2, 3,* and *4* cDNAs were exact nucleic acid sequence matches to contigs in the *C. officinalis* transcriptome. *C. langsdorffii* orthologous TPS cDNAs (ClTPS) isolated from cDNA did not have identical sequence to isolated *CoTPS*. *ClTPS* cDNAs were isolated again and resequenced to differentiate polymerase chain reaction (PCR) mutations from allelic differences naturally present in *C. langsdorffii* populations. Only *ClTPS4* had two reoccurring variants which are referred to hereafter as *ClTPS4-1* and *ClTPS4-2*. Both *CoTPS2* and *ClTPS2* were not isolated from cDNA of stem, leaf, or root of three-year old *C. officinalis* and *C. langsdorffii*.

Sequence similarity of isolated cDNAs with known TPS and motifs

A phylogenetic tree of previously characterized TPS protein sequences ranging from the bryophyte *Physcomitrella patens* to legumes such as *Medicago truncatula* was constructed to further characterize the eight putative TPS identified from both *Copaifera* species (Figure 17). As previously predicted from annotations, almost all *CoTPS* and *ClTPS* were most closely related to δ-cadinene and β-caryophyllene/α-humulene synthases from *Medicago truncatula*. *CoTPS5* was most closely related to *Solanum lycospersicum*, *A. thaliana*, *Zea*

mays, and *Selaginella moellendorffii* copalyl diphosphate synthases (not shown in Figure 17).

Deduced amino acid sequences from *CoTPS1, 2, 3*, and *4* were aligned with the *M. truncatula* δ-cadinene synthase amino acid sequence that shared the closest homology on the phylogenetic tree to investigate conserved domains/motifs (Figure 17). *CoTPS1, 2, 3*, and *4* contained the DDXXD motif. *CoTPS1, 2*, and *4* also contained the EDXXD or a substituted EDXXD, i.e. EEXXD, while CoTPS3 and *M. truncatula* δ-cadinene synthase had remnants of the domain (Figure 17). *CoTPS1, 2, 3*, and *4* also contain the R(R)X8W motif on their N-terminus; however only *CoTPS4* had the RR portion of the motif. Another 'RR' motif is located downstream, but only *CoTPS3* has this motif whereas the other *C. officinalis* contigs have remnants of the motif with a lysine (K) substation and *M. truncatula* has a double replacement to 'KK'. The last motif common to *CoTPS1, 2, 3* and *4* was the RXR motif.

A partial NSE/DTE motif is present in CoTPS1 and 2 while a full motif with a few insertions is present in CoTPS3, CoTPS4, and the *M. truncatula* δ-cadinene synthase. *CoTPS1, 2*, and *4* contained a duplicate and substituted DXDD motif: EDXDD. While the *CoTPS1, 2*, and *4* EDXDD motifs were partially substituted, only CoTPS3 lacked the domain entirely containing only DDXXX.

CoTPS5 was aligned with other diTPS sythases to further determine whether it was a sesquiTPS or a diTPS. CoTPS5 contained the R(R)X8W motif

in the same position as *CoTPS1, 2, 3,* and *4* (Figure 19). CoTPS5 also has both the 'QXXDGGWG' and 'EXDD' motifs. The DXDDTAM motif is part of the β-subunit of diterpene synthases, however, no α-domain motifs, specifically the DDXXD motif, were found in the sequence.

Functional terpene synthase assays of C. officinalis and C. langsdorffii TPS cDNAs

Mono- and sesquiterpene synthase activity of cloned and recombinant synthesized proteins of *C. officinalis* and *C. langsdorffii* contigs were assayed (Table 14). In all instances, *C. officinalis* and *C. langsdorffii* orthologous contigs yielded the same sesquiterpene products (Figure 19 and 19). All major sesquiterpene products were produced by the characterized terpene synthases in this work except for δ-cadinene. CoTPS1 and ClTPS1 yielded β-caryophyllene and α-humulene (Figure 20A, B). ClTPS3 yielded α-bisabolene and farnesol (Figure 20C). CoTPS4, ClTPS4-1, and ClTPS4-2 yielded germacrene D and γ-elemene (Figure 20D, E, G). ClTPS4-1 produced aromadendrene, β-elemene, and muurolene while ClTPS4-2 did not (Table 14). No sesquiterpene products were produced in assays without FPP (Figure 20F).

In TPS *in vitro* assays containing GPP CoTPS1 and ClTPS1, 3, 4-1, and 4-2 produced geraniol and citral products. However, CoTPS4 did not yield any monoterpene products while ClTPS4 was capable of producing citral and

geraniol (Figure 21). The linear products geraniol and citral could result from initial cyclization of GPP by the monoTPS the motifs in CoTPSs and ClTPSs.

Discussion

C. officinalis transcriptome assembly and annotation

De novo assembly and annotation of the *C. officinalis* transcriptome identified 6 putative terpene synthase transcripts from 35 940 total contigs. Several legume transcriptomes and genomes have been published, but the majority of these are herbaceous legumes such as *Glycine max*. The closest phylogenetic comparisons for the assembled *Copaifera* transcriptomes other than *Glycine max* is *Cicer arietinum* (chickpea) which had 34 760 contigs with an average length of 1 020 bp (Garg et al., 2011). The transcriptome of *Cajanus cajan* (pigeon pea) has also been sequenced yielding 21 434 contigs with an average length of 1 510 bp (Kudapa et al., 2012). Several other legume transcriptomes have been assembled, but all are contained in the Fabaceae subfamily Faboideae, whereas *Copaifera* is part of the subfamily Caesalpinioideae. While none of these species belong to the same subfamily as *Copaifera* within the family Fabaceae, the number of contigs and the average length of contigs in these transcriptomes were comparable to the 1KP assembly.

Annotation was partly confounded by the lack of characterized legume sesquiterpene synthases. Surprisingly, no sesquiterpene synthases from *Glycine max* have been functionally characterized though several are predicted from annotated genomes and transcriptomes, e.g. GenBank XM_003541823. The majority of functionally characterized legume terpene synthases come from *Medicago truncatula*. Both the 1KP and TAMU transcriptomes contained three putative monoterpene synthases and five putative sesquiterpene synthase transcripts of interest. Although no functional characterization of the pathway was reported, sequencing allowed for the identification of twelve transcripts involved in triterpene biosynthesis in *S. grosvenorii* from a library of 43,891 contigs which is comparable with this work (Tang et al., 2011). *CoTPS6* was annotated only in the TAMU transcriptome and had the lowest average coverage (16.1 reads bp^{-1}) of all the putative terpene synthases identified in this work. Additionally, two of the five putative terpene synthase genes were not full length contigs. Low coverage and incomplete contigs suggest that more sequencing will be required to complete the *C. officinalis* transcriptome and may also yield more putative terpene synthase transcripts to characterize in future work.

In silico and functional characterization of C. officinalis and C. langsdorffii terpene synthase activity

Phylogenetic analysis and motif identification of deduced amino acid sequences from the five putative *C. officinalis* TPS transcripts and the four isolated *C. langsdorffii* TPS cDNAs. The unrooted phylogenetic tree was in line with previous report of TPS gene families (Chen et al., 2011) and suggested that *CoTPS1, 2, 3, 4* and *ClTPS1, 3, 4-1, 4-2* were most closely related to sesquiterpene synthases in the subfamily A-1. TPSs are divided into class I and class II according to conserved domains. Class I TPSs have the conserved domain DDXXD that coordinates with Mg^{2+} on their C-terminal ends whereas class II TPSs have DXDD on their N-terminal ends (Cao et al., 2010). These motifs determine how the final terpenoid product is formed as class I TPSs initiate product formation by a Mg^{2+}-dependent ionization of the substrate while class II TPSs initiate by protonation of substrate (Cao et al. 2010). *CoTPS1, 2, 3*, and *4* contained the DDXXD motif in the C-terminus suggesting that these are class I TPSs that require Mg^{2+} to catalyze reactions. The RXR motif is highly conserved in TPSs and is involved in stabilizing carbocation intermediates after shifts in the substrate pyrophosphate bonds (Starks et al., 1997). This motif is typically found just upstream of the DDXXD motif (Degenhardt et al., 2009), and is conserved across *CoTPS1, 2, 3*, and *4* (Figure 18). Additionally, a partial NSE/DTE motif is present in CoTPS1 and 2 while a full motif with a few insertions

is present in *CoTPS3, CoTPS4*, and the *M. truncatula* δ-cadinene synthase. This motif is involved in binding Mg2+ clusters in conjunction with the DDXXD motif which in turn binds the pyrophosphate substrates (Degenhardt et al., 2009). Presence of these motifs suggested that the annotated TPS from the *C. officinalis* transcriptome would have TPS when cloned and expressed.

However, *CoTPS1, 2*, and *4* also contained the EDXXD or a substituted EDXXD, i.e. EEXXD, at the N-terminus that is typically associated with class II or bifunctional class I/II diterpene synthases (Cao et al., 2010; Zerbe, Chiang, and Bohlmann, 2012). Bifunctional class I/II diterpene synthases have previously been characterized in gymnosperm species (Cao et al., 2010; Keeling et al. 2011), however to our knowledge no angiosperm bifunctional class I/II sesquiterpene synthases have been isolated previously. This suggests that *CoTPS1, 2*, and *4* TPS genes could have developed through domain loss of a bifunctional diTPS involved in diterpene resin acid biosynthesis.

Several other factors support this hypothesis. *CoTPS3* and *4* contain a substituted second DXDD motif: (E/D)XDD. This alternate motif was previously identified in an *Oryza sativa* ent-copalyl diphosphate synthase involved specifically in primary metabolism of gibberellic acid from diterpene precursors (Prisic et al., 2004). However, this motif was found at the C-terminus of the *O. sativa* diTPS while the motif was found at the N-terminus of *CoTPS3* and *4* near the $R(X)_8W$ motif. New World *Copaifera* species produce oleoresins that are

made primarily of sesquiterpenoid constituents whereas African oleoresins are comprised primarily of diterpenes (Langenheim, 2003). *C. officinalis* oleoresin has previously reported to contain a trace amount of the clarodane diterpene hardwickiic acid (Veiga Junior & Pinto, 2002). The hardwickiic acid production in New World *Copaifera* species could result from an ancestral diTPS which are still functional in African *Copaifera* species. Regrettably, no hardwickiic acid diTPSs have been previously described and so we were unable to specifically identify putative hardwickiic acid diTPS contigs. The genus *Hymenaea* which is related to *Copaifera* also shares the trans-Atlantic distribution and chemical variation between New World and African species (Langenheim Jean, 1996). However, *Hymenaea* oleoresins contain diterpenes in both geographic locations and the chemical variation occurs only within diterpenes rather than the diterpene to sesquiterpene shift as in *Copaifera*. Transit peptide predictions of the isolated CoTPS cDNAs suggest that a partial chloroplast transit peptide may exist (Table 13). Loss of the chloroplast transit peptide along with the diTPS domain would have resulted in localization to the cytoplasm where sesquiTPSs are typically functional leading to the observed shift from diterpenes in African *Copaifera* species to the sesquiterpenes present in New World species. This shift could have occurred during radiation of *Copaifera* species from the African continent to the South American continent where new herbivores were present. Sesquiterpene content of *C. langsdorffii* has been linked to reduced lepidopteran

activity (Macedo & Langenheim, 1989a; Macedo & Langenheim, 1989b; Macedo & Langenheim, 1989c). Evolution of a sesquiTPS gene by diTPS domain loss has also been previously described in *Triticum aestivum* (Hillwig et al., 2011).

CoTPS5 was most closely related to *S. lycospersicum*, *A. thaliana*, *Z. mays*, and *S. moellendorffii* copalyl diphosphate synthases which suggests that *CoTPS5* is involved in gibberellic acid biosynthesis rather than sesquiterpene biosynthesis. *CoTPS5* also contained the R(R)X8W motif in the same position as *CoTPS1, 2, 3*, and *4* (Figure 19). Additionally, *CoTPS5* has both the 'QXXDGGWG' and 'EXDD' motifs associated with diTPS γ-domains that *CoTPS1, 2, 3*, and *4* do not have (Chen et al., 2011; Hillwig et al., 2011). The DXDDTAM motif is part of the β-subunit of diterpene synthases that are also not present in *CoTPS1, 2, 3*, and *4* (Cao et al., 2010). As *CoTPS5* is likely involved in gibberellic acid biosynthesis it was not included in functional assays.

All cloned *CoTPS*s and *ClTPS*s produced sesquiterpenes *in vitro*. The two versions of *ClTPS4* cDNA, *ClTPS4-1* and *ClTPS4-2*, yielded two different profiles in *in vitro* assays. *C. langsdorffii* has been determined to be an outcrossing species but populations have a low genetic diversity because of efficient gene flow through bee pollination and wide-spread seed dispersal by birds (Gonela et al., 2013). This suggests ClTPS4-1 and ClTPS4-2 are two different alleles in *C. langsdorffii* populations, but also suggests these allelic versions of TPSs in *C. officinalis* and *C. langsdorffii* species were not found. The characterized CoTPSs

and ClTPSs were responsible for synthesizing β-caryophyllene and germacrene D which were the major sesquiterpenes identified in *C. officinalis* tissues (Chen et al., 2009). This would suggest that these TPSs are sesquiTPS specifically. However, *CoTPS1, 2, 3*, and *4* also contained monoTPS motifs. The R(R)X8W motif and the downstream RR motif is commonly identified as a monoTPS motif (Williams et al., 1998) and even a putative chloroplastic signal peptide (Lee & Chappell, 2008). The RR or R(R)X8W motif has been implicated in the initial cyclization of GPP to 3S-linalyl diphosphate to generate monoterpenes (Williams et al., 1998). Williams *et. al* (1998) observed that the RR motif was replaced with an RA or RP motif in sesquiTPS and diTPS, which is found in *CoTPS1* and *3* (Figure 18). The 'RR' motif has been previously identified in *Magnolia* TPSs which have both mono- and sesquiTPS activity (Lee & Chappell, 2008). *Magnolia* is an early angiosperm genus which links angiosperm and gymnosperm evolutionary history. As with the identified CoTPSs in this work, the sesquiTPSs from *Magnolia* did not belong to the previously defined class I or class II TPSs (Lee & Chappell, 2008). Presence of these motifs led us to test recombinant CoTPS and ClTPS enzyme activity with GPP. As a result, CoTPS1 and ClTPS1, 3, 4-1, and 4-2 had monoterpene activity in *in vitro* assays catalyzing geraniol and citral from GPP. However, to our knowledge no monoterpenes have ever been reported in either *Copaifera* tissues or oleoresins. This likely occurs from loss of the chloroplast transit peptide on the CoTPS and ClTPS genes as

monoterpenes are typically synthesized in chloroplasts from GPP created by the MEP pathway (Tholl, 2006). The mono- and sesquiTPS activity of the CoTPS and ClTPS recombinant enzymes supports the conclusion that sesquiterpenoid oleoresins produced in angiosperm tree species evolved from mono- and diterpene oleoresins that have been characterized in gymnosperms and basal angiosperms (Banks et al., 2011; Mirov, 1952).

Implications for future work

The success of cross-species gene discovery in this work suggests that even a small investment of next-generation sequencing can lead to functional characterization of several genes in non-model species across a genus. This can be leveraged in future investigation of *Copaifera* terpene biosynthesis to determine the evolutionary pattern of TPS genes across African and New World species in the genus.

The majority of IPP precursors are assumed to be derived from the MVA pathway as sesquiterpenes comprise the majority of *C. officinalis* oleoresins (Chen et al., 2009). However, further investigation into carbon flux from MEP into sesquiterpene biosynthesis will be needed to determine the amount of cross-talk between the two pathways as this could influence the unusually large volume of terpenes produced in *Copaifera*. Expression patterns of characterized TPS genes could also clarify which tissues are primarily responsible for sesquiterpene

production to clarify the genetic and morphologic aspects of this unique phenotype.

The *Copaifera* sesquiterpene biosynthesis pathway also has implications in ecology and natural product economic botany (Fung et al., 2010; Newton et al., 2011; Tang et al., 2011). For example, it is unclear whether oleoresin production is responsive to biotic or abiotic stresses. Variation in leaf oils have been observed in *C. langsdorffii* in response to microlepidopteran herbivory (Macedo & Langenheim, 1989b), however trees previously damaged by termites did not store more oleoresin in trunk tissues (Plowden, 2004; Plowden, 2003). Ultimately, uncertainty about this aspect of the terpene biosynthesis pathway has resulted in oleoresin collectors burning or mechanically wounding trees in the hopes of inducing biosynthesis in trees to yield more oleoresin. However, burning trees likely leads to mortality while trees that have only been tapped can be reharvested yearly (Newton et al., 2011). Induction and expression profiling of transcripts of the *Copaifera* terpene biosynthetic pathways could help to definitely answer these questions that are difficult to determine in highly variable field sites.

Conclusions

This work describes *de novo* transcriptome sequencing, assembly, annotation, and functional characterization of the *C. officinalis* sesquiterpene biosynthesis pathway. The *Copaifera* sesquiterpene biosynthesis pathway is unique regarding

the large volumes of terpenoids produced by trees, the types of terpenoids produced based on geography, and its potential evolutionary link to previously characterized gymnosperm oleoresin production systems. Motifs and the dual mono- and sesquiTPS activity in *C. officinalis* suggests a link between ancestral monoterpene and diterpene oleoresin production in gymnosperm tree species to sesquiterpene oleoresins present in *Copaifera* species. Future analysis should expand to both New World and African *Copaifera* species to fully characterize the underlying genetic mechanisms that respond to the evolution pressures between diterpenoid resins to sesquiterpenoid resins exhibited in *Copaifera*. This work also demonstrates the opportunities that next-generation sequencing technologies provide researchers the ability to investigate biosynthesis pathways in previously unstudied plant species. Ultimately, this will allow researchers to formulate and test hypotheses about entire biosynthetic pathways in novel plant species like *Copaifera* and other species.

Methods

Plants

C. officinalis seeds were received from the University of Puerto Rico courtesy of Dr. James Ackerman and *C. langsdorffii* seeds were received from Nova Odessa, Brazil courtesy of Harry Lorenzi. Seeds were soaked overnight in water

and surface sterilized with 70% ethanol and 10% bleach. Sterilized seeds were then plated on solid MSO plates (Murashige & Skoog, 1962) and cultured until germination in a growth room with 16 hour light photoperiod and 85 °C. Germinated plants were moved to potting media and maintained in greenhouse conditions with 18 hour light photoperiods. Plants were maintained for three years before RNA was extracted.

Illumina library preparation and sequencing of C. officinalis

Total RNA was isolated from the leaf and stem tissues of 16 greenhouse grown *C. officinalis* individuals using TRI Reagent (Molecular Research Center Inc, Cincinnati, OH) and additionally purified using the RNeasy kit (Qiagen, Valencia, CA). Two sets of sequencing data were collected for pooled leaf and stem tissues of *C. officinalis*: 1) sequencing of leaf RNA preparations was performed by BGI as part of the Thousand Plant Transcriptomes project (1KP) and are publically available here: www.onekp.com; and 2) leaf and stem tissues were paired-end sequenced on the Illumina platform at Texas A&M University.

Library construction began with isolation of polyA mRNA from 20 µg of total RNA treated by DNase I (NEB) using Dynabeads mRNA Purification Kit (Life Technologies, Carlsbad, CA). Purified polyA RNA was fragmented in fragmentation buffer (Life Technologies) at 70°C for 90 s to 100-300 nt fragment sizes. The first cDNA strand was synthesized with random hexamer primers using the SuperScript II Reverse Transcription Kit (Life Technologies). The

second-strand was synthesized with RNase H (Life Technologies) and DNA polymerase I (Life Technologies). Short double-stranded cDNA fragments were purified using QIAquick PCR Purification Kit (Qiagen), end-repaired with Klenow polymerase, T4 DNA polymerase and T4 polynucleotide kinase (Enzymatics, Beverly, MA). A single 3' adenosine (A base) was added to the double-stranded cDNA using Klenow (3' to 5' exo-) (Enzymatics) and dATP (Enzymatics). The Illumina PE index adapters were ligated onto the A base on repaired cDNA ends and gel-electrophoresis was used to separate library DNA fragments from unligated adapters by selecting cDNA fragments of 200 bp (± 10% deviation) in size. Libraries were amplified by 15 cycles of PCR with Phusion polymerase (New England Biolabs, Ipswich, MA) and "indexed" by paired-end PCR primers. The libraries were denatured with sodium hydroxide and diluted to 2.5 pM in hybridization buffer for loading onto a lane of an Illumina GAIIx or HiSeq flowcell. Samples sequenced by 1KP were multiplexed whereas those sequenced at Texas A&M University were not. The cDNA libraries were not normalized.

De novo assembly of the C. officinalis transcriptome

After data quality checks were performed using the Illumina platform sequencing pipeline, any read containing more than 10% uncalled bases, 50% of bases with a quality score 5 or lower, or that which contained adapter sequence were removed before assembly. Reads from 1KP were 75 bp paired-end reads. Reads generated by the 1KP project were assembled into contigs using SOAPdenovo

(http://soap.genomics.org.cn/soapdenovo.html) and ultimately were assembled into scaffolds using paired-end sequences to expand contigs. Reads generated at TAMU were assembled using CLC Genomics Workbench.

Functional annotation of assembled C. officinalis reads

Assembled contigs were annotated using the Gene Ontology database (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) automatic annotation server (KAAS) using the SBH algorithm and all plant species as comparative datasets. Annotated KO numbers were assigned to metabolic pathways using KEGG mapper and KegHier 1.1.0 software. Annotated contigs associated with the MVA pathway, MEP pathway, as well as mono-, sesqui-, and diterpene synthases were compared to known terpenoid biosynthesis genes in *Arabidopsis thaliana* and *Medicago truncatula* using NCBI BLAST (blastn and blastx) as well as multiple sequence alignment (MAFFT).

To determine whether the sequenced contigs were full length, both 5' and 3' rapid amplification of cDNA ends (RACE) was carried out using the SMARTer kit (Clontech, Mountain View, CA). Sequence obtained from RACE was assembled with putative terpene synthase contigs identified in the *C. officinalis* transcriptome to build full mRNA contigs. Consensus sequence was then put into the NCBI Orf Finder algorithm to identify coding region nucleotides and amino acid sequence for the putative terpene synthases. Finally, the predicted coding regions were further characterized by identifying motifs using the NCBI Conserved Domain

Search (CDS). Prediction of signal peptide in the coding regions of putative CoTPS cDNAs were carried out with MultiLoc2 using the setting “HighRes plant, 10 localizations” (Höglund et al., 2006), WolfPsort, SIG-Pred, ESLPred (Bhasin & Raghava, 2004), and iPSORT (Bannai et al., 2002) were used on either eukaryotic or plant settings.

Phylogenetic analysis of C. officinalis and C. langsdorffii terpene synthases

To investigate the relatedness of *Copaifera* TPS to characterized TPS phylogenetic trees were built. First, only published, functionally characterized mono-, sesqui-, or diTPS synthase amino acid sequences from land plants were selected from NCBI Protein database or Uni-prot. These sequences were downselected by removing duplicates, which yielded 141 TPS synthases. The TPS amino acid sequences were then aligned in MAFFT (Katoh et al., 2002). Alignments were checked by hand to assure quality and then tested in ProtTest v3.0 (Darriba et al., 2011) to determine the appropriate model of amino acid evolution. TPS amino acid library sequence alignment was input into MrBays (http://mrbayes.sourceforge.net/) along with the evolutionary model JTT + G + F as found by ProtTest with Bayesian information criterion (BIC). Metropolis-coupled Markov chain Monte Carlo simulations starting from a random tree was started in the Bayesian inference to run 2 000 000 generations as previously described (Renner & Waters, 2007). The consensus tree from the analysis is

presented unrooted to determine the relation of identified *C. officinalis* and *C. langsdorffii* to previously described TPS gene families.

Isolation and functional characterization of identified terpene synthase contigs

To clone the putative terpene synthase gene cDNA primers were designed for predicted *C. officinalis* TPS coding regions. These primers were used in PCR reactions using *C. officinalis* and *C. langsdorffii* leaf, stem, and root cDNA as templates. Bands of the expected sizes were gel extracted, purified, and cloned into pCR8/GW/TOPO (Invitrogen, Carlsbad, CA) and sequence confirmed. Additionally, TPS coding region-specific primers were used in PCR reactions with genomic DNA from *C. officinalis* and *C. langsdorffii* to determine whether the putative TPS genes contained introns. This was used to distinguish eukaryotic sequences from potential prokaryote endophyte contamination. Contigs were then cloned into pET101/TOPO *Escherichia coli* expression vectors (Invitrogen, Carlsbad, CA). Transformed BL21 *E. coli* colonies were grown to optical density (OD) 0.50 in 250 mL Erlenmeyer flask in 50 mL of lysogeny broth (LB). Recombinant protein expression was induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) and colonies were placed in 18 °C shaking incubators for 18 hours. Total protein was extracted by sonication as previously described (Chen et al., 2003) in enzyme extraction buffer (50 mM MOPSO, pH 7.0, 10% glycerol, 5 mM $MgCl_2$, 5 mM DTT, 5 mM sodium-ascorbate, 0.5 mM

phenylmethanylsulfanyl fluoride). Crude protein was desalted on PD-10 desalting columns (GE Healthcare Life Sciences, Pittsburgh, PA) in assay buffer (10 mM MOPSO, pH 7.0, 10% glycerol, 1.0 mM DTT). Crude protein was then used in terpene synthase enzyme assays as described previously (Chen et al., 2003; O'Maille et al., 2004). After assays were incubated for at least 5 h at 30 °C, a solid phase microextraction (SPME) silica thread was inserted into the assay headspace for 15 min and then loaded onto the GC/MS and analyzed using previously described methods (Chen et al., 2009).

Authors' contributions

B.L. Joyce isolated the sequenced RNA, conducted transcriptome assembly, annotation, and analysis. He conducted terpene synthase assays, ran the GC-MS, analyzed the data, and wrote the corresponding manuscript. Y. Peng was responsible for assembling reads from the 1KP dataset. H. Al-Ahmad initially aided in developing the project, isolating total RNA from tissues, and cloning transcripts from the de novo transcriptome assembly. S. Liu carried out Illumina sequencing of supplied RNA at Texas A&M and provided sequenced reads. P. Ranjan assembled the TAMU transcriptome, carried out GO annotation of Copaifera transcriptomes, and established a website to run BLAST searches of Copaifera contigs and retrieve sequence. X. Sun sequenced the provided RNA and provided the 1KP sequencing. J.C. John helped process RNA into cDNA,

cloning and sequence verficiation of sesquiterpene synthases of interest. G.K. Wong conceptualized and led the 1KP project for sequencing. F. Chen provided equipment and training for terpene assays and GC-MS and edited the manuscript. J.S. Yuan provided expertise and performed RNAseq experiments at TAMU. C.N. Stewart, Jr developed the *Copaifera* project with H. Al-Ahmad. All authors contributed to the final manuscript, take responsibility of its contents and approved it.

Acknowledgements

We wish to thank the University of Tennessee Science Alliance, Ivan Racheff Endowment, and 1KP for funding the project.

Abbreviations

FAME	fatty acid methyl ester
SNP	single nucleotide polymorphisms
SPME	solid phase microextraction
TPS	terpene synthase

APPENDIX

Figures

Figure 17 - Comparison of the 1KP and TAMU assembled transcriptome annotation using the KAAS SBH algorithm.

A) Comparison of the two transcriptomes based on the number of pathway enzymes identified in each metabolic function. B) Comparison of the two transcriptomes based on the number of unigenes annotated in each metabolic function.

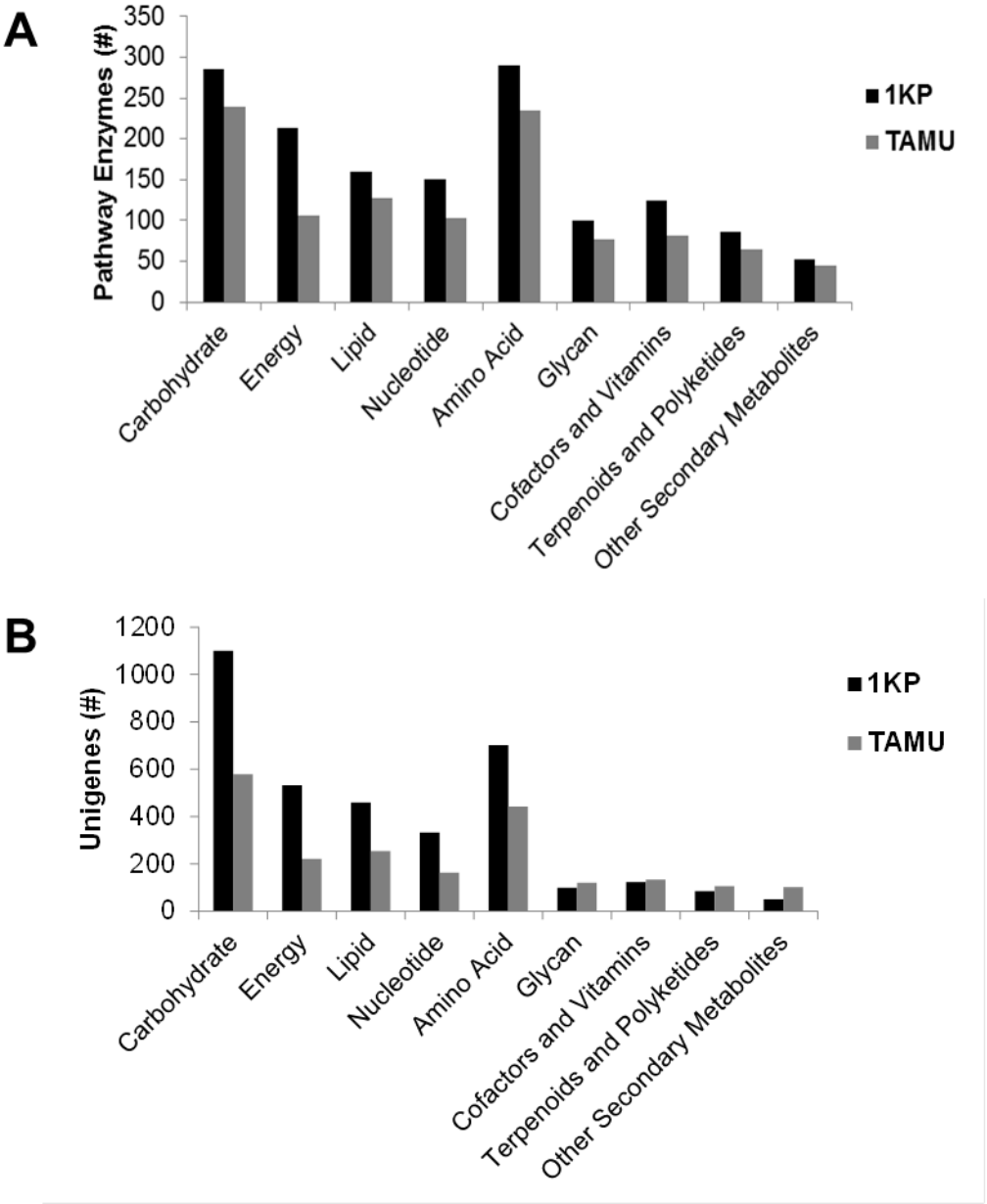
A
Pathway Enzymes (#)
350
300
250
200
150
100
50
0
1KP
TAMU
Carbohydrate
Energy
Lipid
Nucleotide
Amino Acid
Glycan
Cofactors and Vitamins
Terpenoids and Polyketides
Other Secondary Metabolites
B
Unigenes (#)
1200
1000
800
600
400
200
0
1KP
TAMU
Carbohydrate
Energy
Lipid
Nucleotide
Amino Acid
Glycan
Cofactors and Vitamins
Terpenoids and Polyketides
Other Secondary Metabolites

Figure 17.

Figure 18 - Unrooted phylogenetic analysis of identified *C. officinalis* and *C. langsdorffii* terpene synthases and previously functionally characterized mono-, sesqui-, and diterpene synthases.

For clarity, the TPS subfamily A-1 is shown from the analysis and all the isolated TPS genes belonged to this subfamily. Plant species name, GenBank amino acid sequence IDs, and GenBank name is given for each TPS enzyme in the alignment. *CoTPS5* was found to be part of the TPS subfamily-c with other ent-kaurene and ent-copalyl diphosphate synathses and as such is not shown. Only functionally characterized TPS amino acid sequences were used to build the phylogenetic tree.

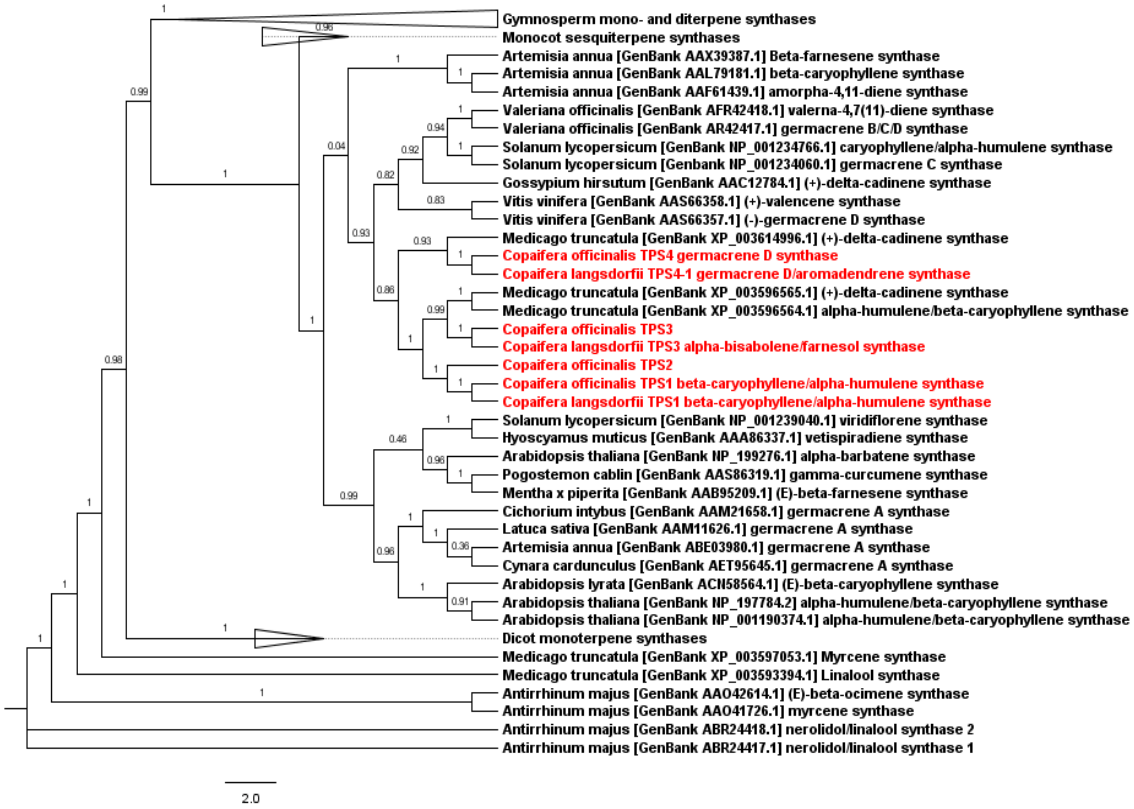
Gymnosperm mono- and diterpene synthases
Monocot sesquiterpene synthases
Artemisia annua [GenBank AAX39387.1] Beta-farnesene synthase
Artemisia annua [GenBank AAL79181.1] beta-caryophyllene synthase
Artemisia annua [GenBank AAF61439.1] amorpha-4,11-diene synthase
Valeriana officinalis [GenBank AFR42418.1] valerna-4,7(11)-diene synthase
Valeriana officinalis [GenBank AR42417.1] germacrene B/C/D synthase
Solanum lycopersicum [GenBank NP_001234766.1] caryophyllene/alpha-humulene synthase
Solanum lycopersicum [Genbank NP_001234060.1] germacrene C synthase
Gossypium hirsutum [GenBank AAC12784.1] (+)-delta-cadinene synthase
Vitis vinifera [GenBank AAS66358.1] (+)-valencene synthase
Vitis vinifera [GenBank AAS66357.1] (-)-germacrene D synthase
Medicago truncatula [GenBank XP_003614996.1] (+)-delta-cadinene synthase
Copaifera officinalis TPS4 germacrene D synthase
Copaifera langsdorfii TPS4-1 germacrene D/aromadendrene synthase
Medicago truncatula [GenBank XP_003596565.1] (+)-delta-cadinene synthase
Medicago truncatula [GenBank XP_003596564.1] alpha-humulene/beta-caryophyllene synthase
Copaifera officinalis TPS3
Copaifera langsdorfii TPS3 alpha-bisabolene/farnesol synthase
Copaifera officinalis TPS2
Copaifera officinalis TPS1 beta-caryophyllene/alpha-humulene synthase
Copaifera langsdorfii TPS1 beta-caryophyllene/alpha-humulene synthase
Solanum lycopersicum [GenBank NP_001239040.1] viridiflorene synthase
Hyoscyamus muticus [GenBank AAA86337.1] vetispiradiene synthase
Arabidopsis thaliana [GenBank NP_199276.1] alpha-barbatene synthase
Pogostemon cablin [GenBank AAS86319.1] gamma-curcumene synthase
Mentha x piperita [GenBank AAB95209.1] (E)-beta-farnesene synthase
Cichorium intybus [GenBank AAM21658.1] germacrene A synthase
Latuca sativa [GenBank AAM11626.1] germacrene A synthase
Artemisia annua [GenBank ABE03980.1] germacrene A synthase
Cynara cardunculus [GenBank AET95645.1] germacrene A synthase
Arabidopsis lyrata [GenBank ACN58564.1] (E)-beta-caryophyllene synthase
Arabidopsis thaliana [GenBank NP_197784.2] alpha-humulene/beta-caryophyllene synthase
Arabidopsis thaliana [GenBank NP_001190374.1] alpha-humulene/beta-caryophyllene synthase
Dicot monoterpene synthases
Medicago truncatula [GenBank XP_003597053.1] Myrcene synthase
Medicago truncatula [GenBank XP_003593394.1] Linalool synthase
Antirrhinum majus [GenBank AAO42614.1] (E)-beta-ocimene synthase
Antirrhinum majus [GenBank AAO41726.1] myrcene synthase
Antirrhinum majus [GenBank ABR24418.1] nerolidol/linalool synthase 2
Antirrhinum majus [GenBank ABR24417.1] nerolidol/linalool synthase 1
2.0

Figure 18.

Figure 19 - Alignment of isolated *C. officinalis* sesquiterpene synthases with *Medicago truncatula* (+)-δ-cadinene synthase (GenBank XP_003596565.1).

Conserved domains/motifs are in bold below alignment. Correct corresponding motif amino acids in alignment are underlined.

The RXR motif is highly conserved in TPSs and is involved in stabilizing carbocation intermediates after shifts in the substrate pyrophosphate bonds (Starks et al., 1997).

CoTPS3 and *4* contain a substituted second DXDD motif: (E/D)XDD. This alternate motif was previously identified in an *Oryza sativa* ent-copalyl diphosphate synthase involved specifically in primary metabolism of gibberellic acid from diterpene precursors (Prisic et al., 2004).

Class I TPSs have the conserved domain DDXXD that coordinates with Mg^{2+} on their C-terminal ends whereas class II TPSs have DXDD on their N-terminal ends (Cao et al., 2010).

Additionally, a partial NSE/DTE motif is present in CoTPS1 and 2 while a full motif with a few insertions is present in *CoTPS3, CoTPS4*, and the *M. truncatula* δ-cadinene synthase.

This motif is involved in binding Mg2+ clusters in conjunction with the DDXXD motif which in turn binds the pyrophosphate substrates (Degenhardt et al., 2009).

However, *CoTPS1, 2*, and *4* also contained the EDXXD or a substituted EDXXD, i.e. EEXXD, at the N-terminus that is typically associated with class II or bifunctional class I/II diterpene synthases (Cao et al., 2010; Zerbe, Chiang, and Bohlmann, 2012). Bifunctional class I/II diterpene synthases have previously been characterized in gymnosperm species (Cao et al., 2010; Keeling et al. 2011).

```
CoTPS1        ---------------------MGRPTANFSSSVWGNQFLSIASGPLLKNKEAEIHQHLQN
CoTPS2        ---------------------MARSTIGYTPGVWGNQFLSVASGPLMKNKKEEIHQHLQN
CoTPS3        -----MATEVSEHVALSSIQNADRPLVKYVPSIWGDFFLQYASEFMEVDD--NMKQKVGV
M.truncatula  ---MSLAPATSVDSTEHAIPDFKRPIVNFSPSIWRNVFLQYDSESVEING--NMKQQVEM
CoTPS4        MSVAALAIATSTPSS-----DVPRRSANYHPSVWGDHFLKYASQPLEVDE--KMEDRIGT
                                     RXXXXXXXXXW            EXDD
CoTPS1        LKEQVRKQLKNGVEEP-SEKLNMIDTIQRLGVSYHFETEIVESLQQLHKNPPSSWDAEDV
CoTPS2        LKQQLGRQLKS-VKEP-CEKLNLIDTMQRLGVSYHFQSEIEESLKHLHKNPPSSWNAKDI
CoTPS3        LKEEVRRMLVSSVNHNFSRKLDFIDSIQRLGVSYHFQHEIDEALKQIHDSFTNNAIITPS
M.truncatula  EKDEVKKMFLFSRNDS-EQNLNFIDSLQRLGISYHFEREIDEALEQIHNTFTNNKEITTK
CoTPS4        LKETVRKMLVPATDKP-LTKVRLIDSIQRLGVDYHFESEIDEVLCQIQNNYVKDGIITLN
                   RR
CoTPS1        DAHLLSISLWFRLLRQQGYYVSCDVFNKFKDDKGVFKTALIDDVEGMLALYEAAYLGIRG
CoTPS2        NSHLLGTALWFRLLRQQGYYVSCDIFNKFKDDKGDFKTILIDDVEGMLALYEAAHLGIRG
CoTPS3        DHDLHSIALLFRLLRQQGYHVSSGIFIQYKDQNGNFNEKLRNDVRGMLSLYEAAQLRIDG
M.truncatula  EGSLHFLALAFRLLRQNRHHLSADIFEKFKNNKGNFNEKLFQDVQEMWSLYEAAQLKING
CoTPS4        E-DLHSLALLFRLLRQQGYHVSPDVFNKFKDEQGKISETIANDVEGMLSLYEAAHLRIHG
                                                 LLNDVXXXXXXXE
CoTPS1        EEILDQVLEFTVFHL-KSRLEGMTPYLQERVDRALYCPINKGLPRIETRYFISTYSKKDS
CoTPS2        EEILDQMLEFTMSYL-KSRLKGMTPYLQERANRALHCPIHKGMLRIETRYYIPIYSKKDS
CoTPS3        DDILAEALDFTSTQL-KLLSSQLGPSLVTEVEHSLRLPLHKTLQRIEARHYMSFYQDDPS
M.truncatula  EDILNEALDFTFSHLNSLITNKLSPFLEKKIRHCLKTPLHKGVPRLETRCYISSYSEEPS
CoTPS4        EDILDEALDFTSTHL-KFLTTQLSDSHAGNVIRSLKRPLRMRLPRLEARHYFFTYQEDPS
              EDXXD
CoTPS1        RNDLLLEFAMLDFNILQQQYQKELSHLTE--WYKKLDFVSKVPYTRDRIVEGYFWPLGAY
CoTPS2        RNDLLLEFAILDFNILQQQYQKELSYITK--WYKKLDFVSKVPYTRDRIVEGYFWPLATY
CoTPS3        HNEILLTFAKLDFDMLQKLHQNEIGNITK--WWKKSDCARRVPYGRDRLVESYFWPLSIS
M.truncatula  HSKILLNFAKLDFNMLQKMHKKELGSITKDMWWKKTDFATEVPYVRDRVVEAYFWPLCMS
CoTPS4        HIETLLAFAKLDFNGLQKLHQREIGNLSK--WWKDLDFATKLPFARNRLVEAYFWILGVY
                                                        RXRXXXXXXW
CoTPS1        FENQYSKGRIIVSKLISVLTALDDTYDAYGTVDELKLFTEAIKRWDINMVASLPECMKVV
CoTPS2        FEKQCSRGRRILGKMIGVFTCLDDTYDNFGTVDELNVLTEAIMRWDINLVASLPECMKVV
CoTPS3        HEPQYSIARSITGKLIAVLALLDDTYDAYGTVQELELFTEAVRRWDASLINFHPEGMKAV
M.truncatula  YEPKYTTSRKIVGKLVACISLLDDTYDAYGTVEELELFTQAIQRWDFSLIQSLPKCMKVV
CoTPS4        FEPCYSLARRIMTKVISLTSIIDDIYDVYGTLEELQLFTEAIDKWDISCMDFLPEYMKLI
                                     DDXXD
```

Figure 19.

```
CoTPS1        FQAILDLLSEMELLTEEDGIS-SFVEYVKPALQDLAKSYLLEAEWRDKSYIPTYEEYMAN
CoTPS2        FETTLDFLIEIELLTEEDGIS-FVVEYVKQGIQGLAKGYMVEAEWRAKGYIPTYDEYIEN
CoTPS3        FGIITELCNEIESVIANEGKLNFIIEHVKHAIYNLAQAYLTETKWGNEGYIPTYSEYKSN
M.truncatula  FNTIVELWDEIVMILVETGKSNLVLQYIKEEFYKLAQSYLVETKWCNEGFIPTYDEYKAN
CoTPS4        YQPLLDVYDEIERETAKEGRA-FCVNYGKEQMRKVVRAYLAEAKWFHNNYTPTLEEYMEV

CoTPS1        GVFSCGYPAVETTSLLGLGKTATKEVFDWISNVPKIVRASSIMCRLTDDLASHKFEQNRE
CoTPS2        GIWTAGYPALEITSLLALGNIATKEVFDWISSMPKIVRASGIVGRIGNDLGSHKREKNIG
CoTPS3        GVATSTYP-LEIISFVSLTTLATEEVLNWISSDPEILKATSIIGRLLDDMASHKFEQERV
M.truncatula  GIISSTLP-LQILSFLGFGEFSNKELFDWIFSDPKIIEAVSAIGRLADDISSHKFEQQRV
CoTPS4        AQVSSAYSMLTTVSFIGVGSIATEEAFKWVTKDPKIVKASLIICRLMDDIVSSKFEQERG

CoTPS1        HVGSAIECCMRQYEVSEEEAYKILLKEIENAWKDLNEEYMKPNG-VPKVVLKCVLNFSRV
CoTPS2        HVATSVECYMKQYGVPEDEAYKLLLKEMENAWKDLNEEYMKPSS-IPKVVLDRVRNYMRA
CoTPS3        HVASSVECCMKQYGISEEEAYKVLHDDITHYWNVLNEESLKLMNVIPKAVLEFLVNLARV
M.truncatula  HVAS-----------SREEAYKLIQIEIEDYWIIMNEECLKIEN-IPRSVLEIILNVARI
CoTPS4        HVVSALECYMKQHGATEEETIVEFRRRVENAWKDINEACLQPFE-VAKPLLMRSLNLSRV

CoTPS1        IEFLYGHFVDKYTNAE-MLKDQIASLFVDPIAIELNK
CoTPS2        NEFYYDRFVDNYTIGE-GMKDDVAAVFLDPIDIDHNK
CoTPS3        SEIAYEKYQDGYTKGE-FLKKYVDAVIVNPIP-----
M.truncatula  TEFTYENFEDKYTKAE-LMKDYIVALLIDPIRIEQCK
CoTPS4        ISLLYTD-DDCYTRSAGNTKKNIEALLINPVA-----
```

Figure 19. Continued.

Figure 20 - Alignment of *C. officinalis* contig 5 with *A. thaliana* (GenBank NP_192187.1) and *Solanum lycopersicum* (GenBank NP_001234008.1) copalyl-diphosphate synthases.

The motifs present in CoTPS5 are primarily associated with diTPS. Conserved domains/motifs are in bold below alignment. Correct corresponding motif amino acids in alignment are underlined. *CoTPS5* also contained the R(R)X8W motif in the same position as *CoTPS1, 2, 3*, and *4* (Figure 18). Additionally, *CoTPS5* has both the 'QXXDGGWG' and 'EXDD' motifs associated with diTPS γ-domains that *CoTPS1, 2, 3*, and *4* do not have (Chen et al., 2011; Hillwig et al., 2011). The DXDDTAM motif is part of the β-subunit of diterpene synthases that are also not present in *CoTPS1, 2, 3*, and *4* (Cao et al., 2010).

```
CoTPS5       MS--SANLSTFCNNNAIGPLPSSMHPSFL---SSPLS----FPNCYRPSKSFNGVSLFKS
S.lycopersic MSI-SASFLRFSLTAHYQPSPS----------SSPPNQPFKFLKSNREHVEFNRI-----
A.thaliana   MSLQYHVLNSIPSTTFLSSTKTTISSSFLTISGSPLNVARDKSRSGSIH-----------

CoTPS5       RAIPVNSSFRIKCSATTGLSAPPSPDI---LSDKSGLPPFARIEILPNRKVGEVSKVSVM
S.lycopersic ----------LQCHA---VSRRRTKDY---KEVQSGSLPVIKWDDIAEEVDEETHTLEVY
A.thaliana   ------------------CSKLRTQEYINSQEVQHDLPLIHEWQQLQGEDAPQISVGS--

CoTPS5       N-YEIQMRVDAVKAMWESIEDGWLNISAYDTAWVALVEDINGSGSPQFPSCLQWIVENQL
S.lycopersic DPSSNEDHIDAIRSMLGSMGDGEISVSAYDTAWVAMVKDVKGTETPQFPSSLEWIANNQL
A.thaliana   NSNAFKEAVKSVKTILRNLTDGEITISAYDTAWVALID--AGDKTPAFPSAVKWIAENQL
                   RXXXXXXXW                                          QX

CoTPS5       PDGSWGDRAVFLSYDRLLSTLACVVALRHWNVHPEKSKRGIEFFKENLERLAKEDPANMX
S.lycopersic ADGSWGDNSIFLVYDRVINTLACVIALKSWNLHPDKILLGMSFMRENLSRIGDENAEHMP
A.thaliana   SDGSWGDAYLFSYHDRLINTLACVVALRSWNLFPHQCNKGITFFRENIGKLEDENDEHMP
             XDGGWG                                              EDENPE

CoTPS5       VGFEMIFPSLIEMARDLNIEVPDPNTHPILKQIYAMKNEKLKRIPMEVVHKMPTSLLFSL
S.lycopersic IGFEVAFPSLIEIAKKLGLDF--PYDSPVLQDIYASRQLKLTRIPKDIMHKVPTTLLHSL
A.thaliana   IGFEVAFPSLLEIARGINIDV--PYDSPVLKDIYAKKELKLTRIPKEIMHKIPTTLLHSL

CoTPS5       EAMPGLQWDKLLKLQSENGSFLSSPASTAFALMQTKDKNCLRYLNDVVQKFSGAVPNFYS
S.lycopersic EGMTDLDWQKLLQFQCTDGSFLFSPSSTAYALMQTQDHNCLNYLKNAVHKFNGGVPNVYP
A.thaliana   EGMRDLDWEKLLKLQSQDGSFLFSPSSTAFAFMQTRDSNCLEYLRNAVKRFNGGVPNVFP


CoTPS5       IEFFEQSWAIDRLTRLGISRYFGEKIKESMNFFYKNWKNTGLGWNRYTCDVPDLDDTIMA
S.lycopersic VDLFEHIWTVDRLQRLGISRYFELKIKKCIDYFSKYWTNKGICWAR-NSPVQDIDDTAMA
A.thaliana   VDLFEHIWIVDRLQRLGISRYFEEEIKECLDYVHRYWTDNGICWAR-CSHVQDIDDTAMA
                                                              DXDDTAM

CoTPS5       FRLLRLHGYDISCDVLKHFETDGEFFCMVGQSSEAVTAMFNLFRASQVSFPGEKIMEDAK
S.lycopersic FRLLRLHGYAVSADVFKHFESKGEFFCFVGQSNQAVTGMYNLYRASHVMFSGEKILENAK
A.thaliana   FRLLRQHGYQVSADVFKNFEKEGEFFCFVGQSNQAVTGMFNLYRASQLAFPREEILKNAK


CoTPS5       RFSCEFLTEKRAANQLGDKWVIAKDIAGEIGFSLDLPWYGILPRIETRFYLDQYGGANDV
S.lycopersic ISTSNYLREKRAQNQLLDKWIITKDLPGEVGYALDVPWYASLPRLETRFFLEHYGGEDDV
A.thaliana   EFSYNYLLEKREREELIDKWIIMKDLPGEIGFALEIPWYASLPRVETRFYIDQYGGENDV

CoTPS5       WIAKVLYRLLRVNNEIYLELGKLDYNNCQALHRTEWAAVQEWYSESGLDQFGLDRDRLLV
S.lycopersic WIGKTLYRMPLVNNSLYLELAKSDYNNCQALHQFEWRRIRKWYYECGLREFGLSEKRLLV
A.thaliana   WIGKTLYRMPYVNNNGYLELAKQDYNNCQAQHQLEWDIFQKWYEENRLSEWGVRRSELLE

CoTPS5       LFFLASSSVFEPERARERLAWVKTSALMEAITSTYNHQRL-RSAFVHEFTNATATSLRSS
S.lycopersic TYYLGSASIFEAQRSTERMAWVKTAALMDCVRSCFGSPQVSAAAFLCEFAHYSSTALNSR
A.thaliana   CYYLAAATIFESERSHERMVWAKSSVLVKAISSSFGESSDSRRSFSDQFHEYIANARRSD

CoTPS5       -KVNERS-----PG------LVNTLMKTLHDISLSTSTAH----YGTLQKMWKKWLLRWE
S.lycopersic YNTEDR--------------LVGVILGTLNHLSLSALLTHGRDIHHYLRHAWENWLLTVG
A.thaliana   HHFNDRNMRLDRPGSVQASRLAGVLIGTLNQMSFDLFMSHGRDVNNLLYLSWGDWMEKWK
```

Figure 20.

```
CoTPS5        SEGDDCEGGAELLANMININAGYFLSRKLQL-NPEYQRLVQLTNQLCHRLQSLQN--SKE
S.lycopersic  EGEGEGEGGAELIIRTLNLCSVHWISEEILLSHPTYQKLLEITNRVSHRLRLYKG--HSE
A.thaliana    LYGDEGEG--ELMVKMIILMKNNDLTN--FFTHTHFVRLAEIINRICLPRQYLKARRNDE

CoTPS5        PASSNNSNKTGLSDPEIESKMQELVQLVLLNS-SNGIDSNIKKTFLALTKTFYYAAYCDS
S.lycopersic  KQVGMLTFSE-----EIEGDMQQLAELVLSHSDASELDANIKDTFLTVAKSFYYSAYCDD
A.thaliana    KEKTI----K-----SMEKEMGKMVELALSESDTFR---DVSITFLDVAKAFYYFALCGD

CoTPS5        KTIDTHIAKVLFERVN
S.lycopersic  RTINFHIAKVLFERVV
A.thaliana    H-LQTHISKVLFQKV-
```

Figure 20. Continued.

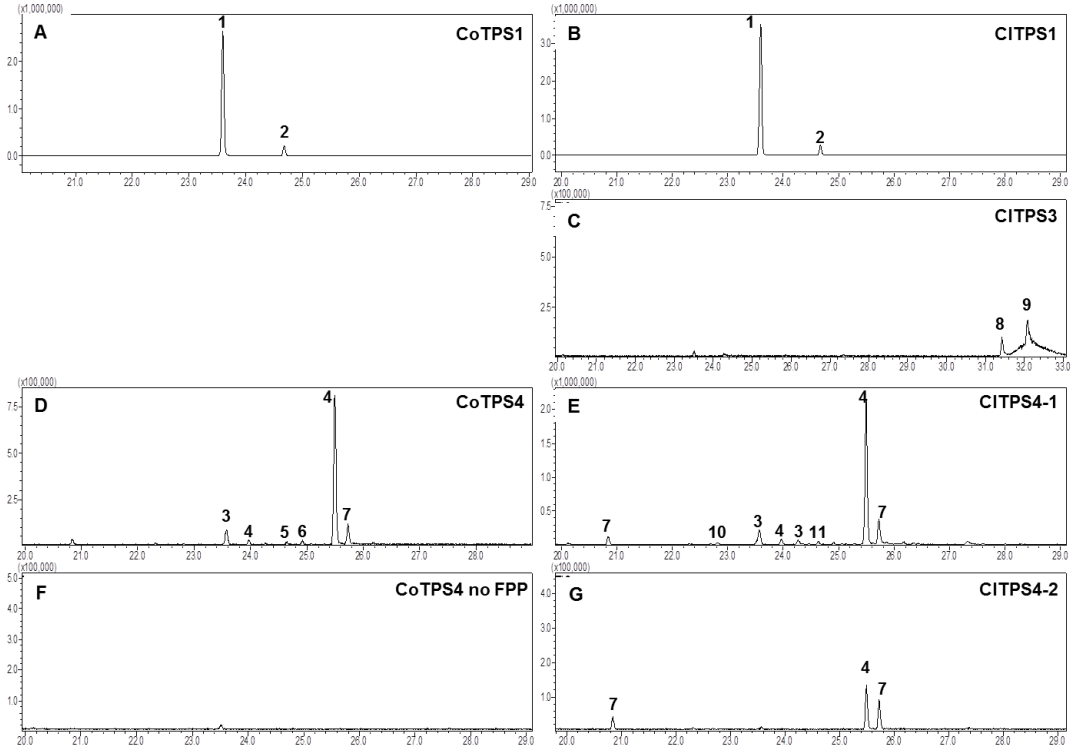

Figure 21 - GC-MS chromatograms of *C. officinalis* (Co) and *C. langsdorffii* (Cl) terpene synthase enzyme assay using FPP.

1) β-caryophyllene, 2) α-caryophyllene, 3) aromadendrene, 4) germacrene D, 5) α-bourbonene, 6) α-farnesene, 7) γ-elemene, 8) α-bisabolene, 9) farnesol, 10) β-elemene, 11) muurolene.

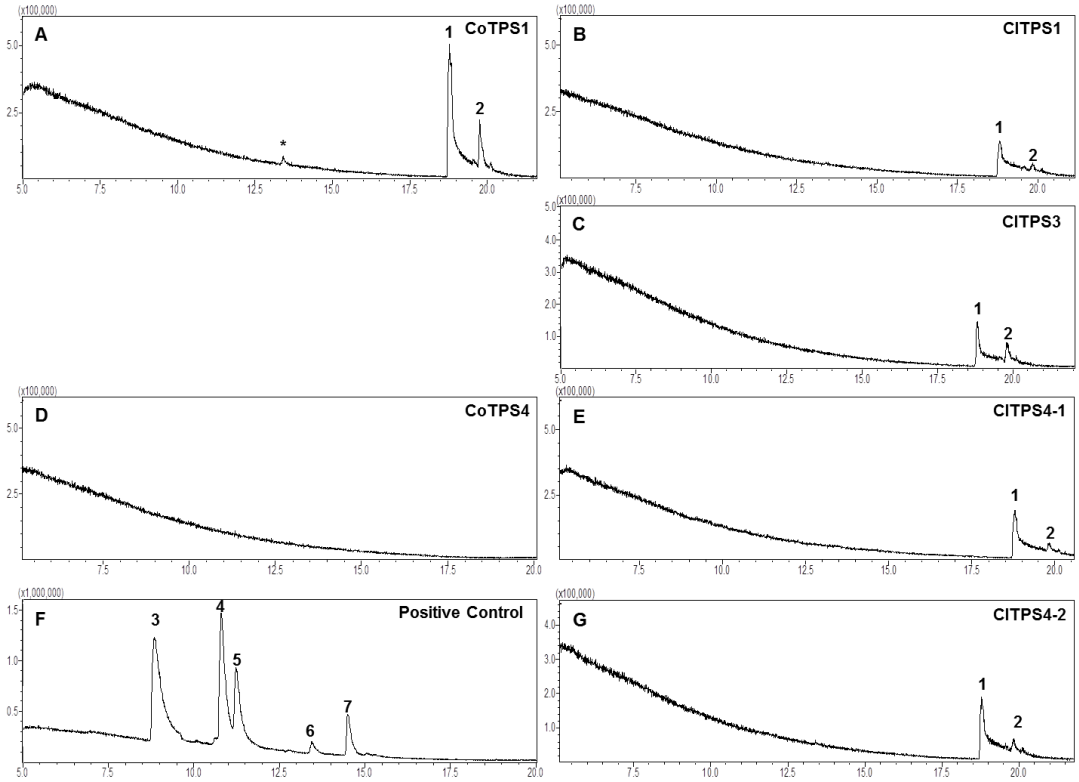

Figure 22 - GC-MS chromatograms of *C. officinalis* (Co) and *C. langsdorffii* (Cl) terpene synthase enzyme assays using GPP as a substrate.

1) geraniol, 2) citral, 3) β-myrcene, 4) α-pinene, 5) β-ocimene, 6) unknown, 7) unknown.

Tables

Table 12. *De novo* assembly statistics for *C. officinalis* transcriptome.

Assembly Statistics[†]	**1KP**	**TAMU**
All reads (total count)	25,626,055	
All reads (total bp)	1,875,534,405	
Average read length (bp)	73	
Matched (count)	23,417,908 (91%)	
Paired reads matched	11,604,984	
Broken paired reads matched	11,809,512	
Not matched (count)	2,208,147 (9%)	
Total contigs (count)	35,940	92147[¥] 11417[€]
Contigs average Length	984	521
Total bp of contigs	35,369,123	47,973,086
N75	807	330
N50	1,557	692
N25	2,422	1460
Shortest contig length (bp)	172	93
Longest contig length (bp)	13,853	12419
Contigs annotated (KAAS SBH)	7,771 (22%)	3,448 (30%)

[†] CLC Genomics Workbench

[¥] 1st assembly using CLC standard settings

[€] 2nd assembly using contigs from 1st assembly; remaining statistics are for this assembly

Table 13. KEGG annotated transcripts in terpenoid biosynthetic pathways for the 1KP and TAMU *C. officinalis* transcriptome.

Pathway	Gene	EC Number	Number of Unigenes	
			1KP	TAMU
MVA	Acetyl-CoA C-acetyltransferase	2.3.1.9	2	
	Hydroxymethylglutaryl-CoA synthase	2.3.3.10	5	1
	Hydroxymethylglutaryl-CoA reductase	1.1.1.34	2	2
	Mevalonate kinase	2.7.1.36	3	2
	Phosphomevalonate kinase	2.7.4.2	1	1
	Diphosphomevalonate decarboxylase	4.1.1.33	1	1
MEP	1-deoxy-D-xylulose-5-phosphate synthase	2.2.1.7	5	3
	1-deoxy-D-xylulose-5-phosphate reductoisomerase	1.1.1.267	3	1
	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	2.7.7.60	1	1
	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	2.7.1.148	1	1
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	4.6.1.12	3	
	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	1.17.7.1	4	
	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	1.17.1.2	2	
	Isopentenyl-diphosphate delta-isomerase	5.3.3.2	2	
	Geranyl pyrophosphate synthase	2.5.1.1	5	2
	Farnesyl pyrophosphate synthase	2.5.1.10	3	1
	Geranylgeranyl pyrophosphate synthase	2.5.1.29	4	1
Monoterpene	Myrcene/ocimene synthase	4.2.3.15	4	1
	Linalool synthase	4.2.3.25	1	
	1,8 cineole synthase	4.2.3.108	1	
Sesquiterpene	α-humulene/β-caryophyllene	4.2.3.-4.2.3.57	10	4
Diterpene	Ent-copalyl pyrophosphate synthase	5.5.1.13	2	2
	Ent-kaurene synthase	4.2.3.19	1	

Table 14. Contig length, annotation, and signal peptide prediction of isolated terpene synthases.

Identified Terpene Synthase	Length of Contig in bp (1KP, TAMU)	RACE Length (bp)	Average Coverage (Reads bp^{-1})	KAAS Annotation	Signal Peptide Prediction†
CoTPS1	1650, 1650	Full length	66.98	B-caryophyllene α-humulene synthase	Chloroplast (11) Cytoplasm (0.36) Chloroplast Cytoplasm Cytoplasm (94%)
CoTPS2	1218, 1218	5' - 517	51.06	B-caryophyllene α-humulene synthase	Cytoplasm (10) Cytoplasm (0.78) Chloroplast Cytoplasm Cytoplasm (75%)
CoTPS3	1683, 1680	Full length	173.17	B-caryophyllene α-humulene synthase	Cytoplasm (11) Peroxisome (0.73) Cytoplasm Cytoplasm Nuclear (75%)
CoTPS4-1 CoTPS4-2	1674, 1674 1482	Full length	162.33	B-caryophyllene α-humulene synthase	Chloroplast (9.5) Cytoplasm (0.57) Chloroplast Cytoplasm Cytoplasm (53%)
CoTPS5	N/A, 2148	5' - 365	31.9	Ent-copalyl diphosphate synthase	Cytoplasm (10) Cytoplasm (0.49) Cytoplasm Cytoplasm Cytoplasm (53%)
CoTPS6	1833, N/A	Full length	16.1	Isolated coding region contained cryptic stop codons	

† Prediction was carried out by WolfPsort, Multiloc, iPSORT, SIG-Pred, and ESLPred followed by prediction scores

Table 15. List of volatile terpenes found during *C. officinalis* and *C. langsdorffii* enzyme assays with GPP and FPP.

Terpene Synthase Activity	Monoterpenes (GPP activity)	Peak Area (%)	Sesquiterpenes (FPP activity)	Peak Area (%)
CoTPS1	(Z)-geraniol Citral	77.91 22.09	β-Caryophyllene α-Caryophyllene	92.72 7.28
ClTPS1	(Z)-geraniol Citral		β-Caryophyllene α-Caryophyllene	92.72 7.28
ClTPS3	(Z)-geraniol Citral	69.45 30.55	Farnesol α-Bisabolene	26.81 73.19
CoTPS4	No monoterpene products		Germacrene D Aromadendrene α-Bourbonene α-Muurolene γ-Elemene	80.35 8.08 1.25 1.67 9.27
ClTPS4-1	(Z)-geraniol (2E)-geraniol	91.97 8.03	Germacrene D Aromadendrene γ-Elemene β-Elemene Muurolene	60.67 7.74 3.54 0.79 1.39
ClTPS4-2	(Z)-geraniol (2E)-geraniol	85.31 14.69	Germacrene D γ-Elemene	53.99 32.65

Additional files

Table 16. Primers used to isolate full length transcripts and cloning terpene synthase into vectors.

Primer Name	Use	Sequence	Tm
Co TPS1 1F	CDS Isolation	5'-ATG GGG CGA CCC ACG GCA AAC-3'	65.5
Co TPS1 1R	CDS Isolation	5'-CTA TTT ATT GAG TTC AAT GGC AAT GGG-3'	54.8
Co TPS2 1F	CDS Isolation	5'-ATG CTA GCC TTA TAT GAA GCT GCA-3'	56.4
Co TPS2 1R	CDS Isolation	5'-CTA CTT ATT GTG ATC AAT ATC TAT GGG-3'	51.2
Co TPS3 1F	CDS Isolation	5'-ATG GCT ACT GAA GTT TCA GAA CAT-3'	54.5
Co TPS3 1R	CDS Isolation	5'-TTA TGG GAT TGG ATT CAC AAT CAC-3'	53.3
Co TPS4 1F	CDS Isolation	5'-ATG TCG GTT GCA GCG TTA GCA AT-3'	59.8
Co TPS4 1R	CDS Isolation	5'-TCA TGC GAC AGG ATT TAT GAG CAA G-3'	57.4
Co TPS5 1F	CDS Isolation	5'-ATG AAC TAT GAG ATC CAG ATG AGG G-3'	55.7
Co TPS5 1R	CDS Isolation	5'-TCA GTT GAC CCT TTC GAA TAG AAC C-3'	56.7
Co TPS6 1F	CDS Isolation	5'-ATG GCT CTT CTT TTT ATG TCT TCT C-3'	53.2
Co TPS6 1R	CDS Isolation	5'-CTA TAA ATC ACA GGG ACG AAT CTT-3'	53.4
CoTPS1 Pet F	pET101 cloning	5'-CAT CAT GGG GCG ACC CAC GGC AAA C-3'	68.4
CoTPS1 PET 1R	pET101 cloning	5'-TTT ATT GAG TTC AAT GGC AAT GGG-3'	54.2
Co TPS2 Pet F	pET101 cloning	5'-CAC CAT GCT AGC CTT ATA TGA AGC TGCAT-3'	60.4
CoTPS2 PET 1R	pET101 cloning	5'-CTTATT GTGATC AAT ATC AAT ATC TAT GGG-3'	49
CoTPS3 PETF	pET101 cloning	5'-CAC CAT GGC TAC TGA AGT TTC AGA ACA T-3'	58.9
CoTPS3 PET1R	pET101 cloning	5'-TGG GAT TGG ATT CAC AAT CAC-3'	53
CoTPS4 PETF	pET101 cloning	5'-CAC CAT GTC GGT TGC AGC GTT AGC AAT-3'	63.4
CoTPS4 PET1R	pET101 cloning	5'-TGC GAC AGG ATT TAT GAG CAA-3'	55.4

Table 16. Continued.

Primer Name	Use	Sequence	Tm
CoTPS5 PETF	pET101 cloning	5'-CAC CAT GAA CTA TGAGATCCA GAT GAG GG-3'	59.8
CoTPS5 PET1R	pET101 cloning	5'-GTT GAC CCT TTC GAA TAG AAC C-3'	53.5
CoTPS1 3RACE 1F	RACE PCR	5'-AGC CAA ATG GTG TCC CAA AGG TGG T-3'	63.1
CoTPS2 3RACE 1F	RACE PCR	5'-GGT TGA AGC TGA GTG GAG GGC CAA A-3'	63.3
CoTPS2 3RACE 2F	RACE PCR	5'-GAG AGG GGC TGA AGG ATC ACG TTG C-3'	63.3
CoTPS3 3RACE 1F	RACE PCR	5'-TGG CAA CCG AGG AGG TGC TTA ACT G-3'	62.9
CoTPS4 3RACE 1F	RACE PCR	5'-GCC TCT GCT GAT GCG AAG CCT GAA C-3'	64
CoTPS5 3RACE 1F	RACE PCR	5'-TAA CGC CAC GGC TAC ATC GCT CAG A-3'	63.7
CoTPS5 3RACE 2F	RACE PCR	5'-CTG GCC TTT CGG ACC CTG AAA TTG A-3'	61.2
TPS1 5' RACE 1R	5' RACE PCR	5'-AGC CCT CGA CAA TTC TGT CTC TGG TGT-3'	63
TPS1 5'RACE 1NR	Nested 5' RACE	5'-CCA GGA AGA AGG AGG ATT CTT GTG CAG-3'	60.8
TPS2 5' RACE 1R	5' RACE PCR	5'-GCC CTC CAC TCA GCT TCA ACC ATG TAT-3'	62.3
TPS2 5'RACE 1NR	Nested 5' RACE	5'-GAT TCG GGG AGA GAA GCC ACC AGA TTA-3'	61.5
TPS3 5'RACE 1R	5' RACE PCR	5'-CCG TAT GGA ACT CGT CTT GCA CAG TCT-3'	61.8
TPS3 5'RACE 1NR	Nested 5' RACE	5'-CCG TGC CTC TAT CCT TTG CAG TGT CTT-3'	62.2
TPS4 5'RACE 1R	5' RACE PCR	5'-CAG AGG CCG CTT TAA GCT TCG AAT GAC-3'	61.5
TPS4 5' RACE 1 NR	Nested 5' RACE	5'-TCG ACC TCC AAA GGC TGA GAA GCA TA-3'	61.4
TPS5 5'RACE 1R	5' RACE PCR	5'-GTG GAG GCT GGA GAG GAC AAG AAG GAT-3'	63.2
TPS5 5'RACE 1NR	Nested 5' RACE	5'-GGA TGA GTG TTG GGA TCA GGA ACT TCG-3'	60.7

Table 17. Functionally characterized terpene synthase protein sequences used in the phylogenetic analysis.

Plant species	Database	ID Number	Enzyme Name
Abies grandis	GenBank	AAB05407.1	Abietadiene cyclase
Abies grandis	Uniprot	O24474.1	Myrcene synthase
Abies grandis	GenBank	AAB71085.1	Pinene synthase
Abies grandis	GenBank	AAK83561.1	Delta-selinene synthase
Abies grandis	GenBank	AAC05728.1	Gamma-humulene synthase
Abies grandis	GenBank	AAF61453.1	Beta-phellandrene synthase
Abies grandis	Uniprot	Q9M7D0.1	Terpinolene synthase
Abies grandis	GenBank	AAF61455.1	Limonene/alpha-pinene synthase
Abies grandis	GenBank	AAB70907.1	4S-limonene synthase
Abies grandis	GenBank	AAK83562.1	E-alpha bisabolene synthase
Abies grandis	Uniprot	Q948Z0.1	Camphene synthase
Adiantum capillus-veneris	GenBank	BAF93208.1	Cycloartenol synthase
Antirrhinum majus	GenBank	AAO41726.1	Myrcene synthase
Antirrhinum majus	GenBank	AAO42614.1	(E)-beta-ocimene synthase
Antirrhinum majus	GenBank	ABR24417.1	Nerolidol/linalool synthase 1
Antirrhinum majus	GenBank	ABR24418.1	Nerolidol/linalool synthase 2
Arabidopsis lyrata	GenBank	ACN58564.1	(E)-beta-caryophyllene synthase
Arabidopsis thaliana	GenBank	NP 001031651.1	(E)-beta-ocimene synthase
Arabidopsis thaliana	GenBank	NP 001190374.1	Alpha-humulene/beta-caryophyllene synthase
Arabidopsis thaliana	GenBank	NP 176868.1	Lupeol synthase
Arabidopsis thaliana	GenBank	NP 178064.1	Ent-kaur-16-ene synthase
Arabidopsis thaliana	GenBank	NP 179998.1	Myrcene/ocimene synthase
Arabidopsis thaliana	GenBank	NP 189212.1	1,8-cineole synthase
Arabidopsis thaliana	GenBank	NP 192187.1	Ent-copalyl diphosphate synthase
Arabidopsis thaliana	GenBank	NP 193272.1	Baruol synthase
Arabidopsis thaliana	GenBank	NP 197784.2	Alpha-humulene/beta-caryophyllene synthase
Arabidopsis thaliana	GenBank	NP 199276.1	Alpha-barbatene synthase
Artemisia annua	GenBank	AAF13356.1	Linalool synthase

Table 17. Continued.

Plant species	Database	ID Number	Enzyme Name
Artemisia annua	GenBank	AAF61439.1	Amorpha-4,11-diene synthase
Artemisia annua	GenBank	AAK58723.1	Beta-pinene synthase
Artemisia annua	GenBank	AAL79181.1	Beta-caryophyllene synthase
Artemisia annua	GenBank	AAX39387.1	Beta-farnesene synthase
Artemisia annua	GenBank	ABE03980.1	Germacrene A synthase
Chamaecyparis formosensis	GenBank	AFJ23663.1	Beta-cadinene synthase
Cichorium intybus	GenBank	AAM21658.1	Germacrene A synthase
Citrus limon	GenBank	AAM53944.1	(+)-limonene synthase
Citrus limon	GenBank	AAM53945.1	Beta-pinene synthase
Citrus limon	Uniprot	Q8L5K4.1	Gamma-terpinene synthase
Copaifera langsdorffii TPS1	GenBank	KF218242	Beta-caryophyllene/alpha-humulene synthase
Copaifera langsdorffii TPS3	GenBank	KF218243	
Copaifera langsdorffii TPS4-1	GenBank	KF218244	
Copaifera officinalis TPS1	GenBank	KF218237	Beta-caryophyllene/alpha-humulene synthase
Copaifera officinalis TPS2	GenBank	KF218238	
Copaifera officinalis TPS3	GenBank	KF218239	
Copaifera officinalis TPS4	GenBank	KF218240	Germacrene D synthase
Copaifera officinalis TPS5	GenBank	KF218241	
Cynara cardunculus	GenBank	AET95645.1	Germacrene A synthase
Ginkgo biloba	GenBank	AAS89668.1	Levopimaradiene synthase
Gossypium hirsutum	GenBank	AAC12784.1	(+)-delta cadinenesynthase
Hyoscyamus muticus	GenBank	AAA86337.1	Vetispiradiene synthase
Jungermannia subulata	GenBank	BAJ39816.1	Ent-kaurene synthase
Lactuca sativa	GenBank	AAM11626.1	Germacrene A synthase
Lavandula angustifolia	GenBank	ABB73044.1	Limonene synthase
Lavandula angustifolia	GenBank	ABB73045.1	Linalool synthase

Table 17. Continued.

Plant species	Database	ID Number	Enzyme Name
Lavandula angustifolia	GenBank	ABB73046.1	Trans-alpha-bergamotene synthase
Lavandula angustifolia	GenBank	ADQ73631.1	Beta-phellandrene synthase
Lavandula angustifolia	GenBank	AFL03423.1	1,8-cineole synthase
Medicago truncatula	GenBank	XP 003593394.1	Linalool synthase
Medicago truncatula	GenBank	XP 003596036.1	(R)-limonene synthase
Medicago truncatula	GenBank	XP 003596564.1	Alpha-humulene/beta-caryophyllene synthase
Medicago truncatula	GenBank	XP 003596565.1	(+)-delta cadinene synthase
Medicago truncatula	GenBank	XP 003597053.1	Myrcene synthase
Medicago truncatula	GenBank	XP 003600343.1	Ent-kaur-16-ene synthase
Medicago truncatula	GenBank	XP 003614996.1	(+)-delta-cadinene synthase
Medicago truncatula	GenBank	XP 003621227.1	Alpha-farnesene synthase
Mentha x piperita	GenBank	AAB95209.1	(E)-beta-farnesene synthase
Oryza sativa	GenBank	ABJ16553.1	(E)-beta-caryophyllene/beta-elemene synthase
Oryza sativa	GenBank	ABJ16554.1	Farnesol synthase
Physcomitrella patens	GenBank	BAF61135.1	Ent-kaurene synthase
Picea abies	GenBank	AAO73863.1	(+)-3-carene synthase
Picea abies	GenBank	AAS47697.1	E,E-alpha-farnesene synthase
Picea abies	GenBank	BAG68223.1	Gamma-humulene synthase
Picea abies	Uniprot	Q675L0.1	Longifolene synthase
Picea abies	Uniprot	Q675L4.1	Levopimaradiene synthase
Picea abies	Uniprot	Q675L5.2	Isopimaradiene synthase
Picea engelmannii x Picea glauca	GenBank	ADZ45497.1	1,8-cineole synthase
Picea engelmannii x Picea glauca	GenBank	ADZ45510.1	(+)-3-carene synthase
Picea engelmannii x Picea glauca	GenBank	ADZ45514.1	Alpha-farnesene synthase/beta-ocimene synthase
Picea engelmannii x Picea glauca	GenBank	ADZ45515.1	Longifolene synthase
Picea glauca	GenBank	ACM04452.2	3-carene synthase
Picea glauca	GenBank	ACY25275.1	Ent-kaurene synthase
Picea glauca	GenBank	ADZ45498.1	1,8-cineole synthase
Picea glauca	GenBank	ADZ45500.1	Linalool synthase
Picea glauca	GenBank	ADZ45508.1	Alpha/beta-pinene synthase

Table 17. Continued.

Plant species	Database	ID Number	Enzyme Name
Picea glauca	GenBank	ADZ45513.1	Alpha-humulene synthase
Picea glauca	Uniprot	C7ASI9.2	Carene synthase
Picea sitchensis	GenBank	AAP72020.1	Pinene synthase
Picea sitchensis	GenBank	ABA86247.1	Linalool-like synthase
Picea sitchensis	GenBank	ABA86248.1	Limonene synthase
Picea sitchensis	GenBank	ADB55710.1	Ent-kaurene synthase
Picea sitchensis	GenBank	ADU85930.1	(+)-sabinene synthase
Picea sitchensis	GenBank	ADZ45499.1	1,8-cineole synthase
Picea sitchensis	GenBank	ADZ45503.1	Beta-phellandrene synthase
Picea sitchensis	GenBank	ADZ45509.1	Alpha/beta-pinene synthase
Picea sitchensis	GenBank	ADZ45511.1	(+)-3-carene synthase
Picea sitchensis	GenBank	ADZ45512.1	Isopimaradiene synthase
Picea sitchensis	GenBank	ADZ45517.1	Levopimaradiene/abietadiene synthase
Pinus attenuata	GenBank	AFJ73535.1	2-methyl-3-buten-2-ol synthase
Pinus sylvestris	GenBank	ABV44452.1	Caryophyllene/humulene synthase
Pinus sylvestris	GenBank	ABV44453.1	1(10),5-germacradien-4-ol synthase
Pinus sylvestris	GenBank	ABV44454.1	Longifolene synthase
Pinus sylvestris	GenBank	ADH29869.1	E-beta-farnesene synthase
Pinus tabuliformis	GenBank	ABY65904.1	Alpha-pinene synthase
Pinus taeda	GenBank	AAO61226.1	Alpha-farnesene synthase
Pinus taeda	GenBank	AAO61227.1	Alpha-terpineol synthase
Pinus taeda	GenBank	AAO61225.1	Alpha-pinene synthase
Pinus taeda	GenBank	AAO61228.1	(+)-alpha-pinene synthase
Pinus taeda	Uniprot	Q50EK2.1	Levopimaradiene synthase
Pogostemon cablin	GenBank	AAS86319.1	Gamma-curcumene synthase
Populus trichocarpa	GenBank	XP 002311286.1	Ent-kaurene synthase
Pseudotsuga menziesii	GenBank	AAX07264.1	Terpinolene synthase
Pseudotsuga menziesii	GenBank	AAX07265.1	(E)-beta-farnesene synthase

Table 17. Continued.

Plant species	Database	ID Number	Enzyme Name
Pseudotsuga menziesii	GenBank	AAX07266.1	(E)-gamma-bisabolene synthase
Pseudotsuga menziesii	GenBank	AAX07267.1	Alpha-pinene/camphene synthase
Pseudotsuga menziesii	GenBank	ADX42737.1	(E)-beta-farnesene synthase 1
Pseudotsuga menziesii	GenBank	ADX42738.1	(E)-beta-farnesene synthase 2
Santalum album	GenBank	ADO87000.1	Santalene synthase
Selaginella moellendorffii	Uniprot	EFJ31965.1	Copalyl-diphosphate synthase
Selaginella moellendorffii	Uniprot	EFJ33584.1	Ent-kaurene synthase
Selaginella moellendorffii	Uniprot	EFJ37889.1	Kaurene synthase
Solanum lycopersicum	GenBank	NP 001234008.1	Copalyl-diphosphate synthase
Solanum lycopersicum	GenBank	NP 001234060.1	Germacrene C synthase
Solanum lycopersicum	GenBank	NP 001234766.1	Caryophyllene/alpha-humulene synthase
Sorghum bicolor	Uniprot	C5YHH7.2	Zingiberene synthase
Taxus baccata	Uniprot	Q93YA3.1	Taxadiene synthase
Valeriana officinalis	GenBank	AFR42417.1	Germacrene B/C/D synthase
Valeriana officinalis	GenBank	AFR42418.1	Valerena-4,7(11)-diene synthase
Vitis vinifera	GenBank	AAS66357.1	Germacrene D synthase
Vitis vinifera	GenBank	AAS66358.1	(+)-valencene synthase
Zea diploperennis	Uniprot	C7E5V9.1	(E)-beta-farnesene synthase
Zea mays	GenBank	NP 001105257.1	Kaurene synthase 2
Zea mays	GenBank	NP 001148059.1	Ent-kaurene synthase B
Zea mays	GenBank	NP 001105850.1	(E)-beta-farnesene synthase
Zea mays	GenBank	NP 001105950.1	(S)-beta-macrocarpene synthase
Zea mays	Uniprot	Q6JD70.1	Sesquithujene synthase A
Zingiber zerumbet	GenBank	BAG12022.1	Beta-eudesmol synthase

CHAPTER IV

ETHANOL AND HIGH-VALUE TERPENE CO-PRODUCTION FROM LIGNOCELLULOSIC BIOMASS OF *CYMBOPOGON FLEXUOSUS* AND *CYMBOPOGON MARTINII*

This chapter is formatted for submission to Biomass & Bioenergy. B.L. Joyce carried out ethanol fermentation experiments and wrote the manuscript. Valtcho D. Zheljazkov was involved in designing the project, conducted field trials of *Cymbopogon flexuosus* and *C. martinii*, carried out chemical analysis of essential oils, and assisted in writing the manuscript. Robert Sykes determined *C. flexuosus* and *C. martinii* biomass composition. Charles L. Cantrell conducted chemical analysis of essential oils and edited the manuscript. Choo Hamilton conducted HPLC analysis on ethanol fermentation experiments. David G. Mann assisted with ethanol fermentation studies and edited the manuscript. Miguel Rodriguez assisted with ethanol fermentation studies and edited the manuscript. Jonathan Mielenz provided simultaneous saccharification and fermentation expertise, provided training and equipment, and edited the manuscript. Tess Astatkie performed statistical analysis and assisted in writing the manuscript. C.N. Stewart, Jr. was involved in designing the project, provided funding, and edited the manuscript.

Blake L. Joyce [a], Valtcho D. Zheljazkov [b,*], Robert Sykes [c,d], Charles L. Cantrell [e], Choo Hamilton [d,f], David G. Mann [a,d], Miguel Rodriguez [d,f], Jonathan Mielenz [d,f], Tess Astatkie [g], and C. Neal Stewart, Jr.[a]

[a] The University of Tennessee, Department of Plant Sciences, Knoxville, TN 37996, USA

[b] University of Wyoming, Sheridan Research and Extension Center, 663 Wyarno Road, Sheridan WY 82801, USA

[c] National Renewable Energy Laboratory, Golden, CO 80401, USA

[d] BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[e] United States Department of Agriculture, Agricultural Research Service, NPURU, University, MS 38677, USA

[f] Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[g] Dalhousie University, Faculty of Agriculture, 50 Pictou Road, P.O. Box 550, Truro, NS, B2N 5E3, Canada

* *Corresponding author.* Address: Tel.: +1 307 7372415; E-mail address: vjeliazk@uwyo.edu (V.D. Zheljazkov).

ABSTRACT

Cymbopogon flexuosus and *C. martinii* are perennial grasses grown to produce essential oils for the fragrance industry. The objectives of this study were (1) to evaluate biomass and oil yields as a function of nitrogen and sulfur fertilization, and (2) to characterize their utility for lignocellulosic ethanol compared to *Panicum virgatum* (switchgrass). Mean biomass yields were 12.83 Mg lemongrass ha^{-1} and 15.11 Mg palmarosa ha^{-1} during the second harvest year resulting in theoretical biofuel yields of 2541 and 2569 L ethanol ha^{-1} respectively compared to reported 1749-3691 L ethanol ha^{-1} for switchgrass. Pretreated lemongrass yielded 198 mL ethanol (g biomass)$^{-1}$ and pretreated palmarosa yielded 170 mL. Additionally, lemongrass yielded 85.7 kg essential oil ha^{-1} and palmarosa yielded 67.0 kg ha^{-1} with an estimated value of (USD) 857 and 1005 ha^{-1}. These data suggest that dual use crops such as lemongrass and palmarosa may increase the economic viability of lignocellulosic biofuels.

Keywords

Cymbopogon flexuosus, *Cymbopogon martinii*, Simultaneous saccharification and fermentation, Ethanol, Lignocellulosic biofuel, High-value coproduct, Switchgrass, Perennial grasses

1. Introduction

Replacing petroleum as a natural resource extends beyond producing renewable liquid fuels. Petroleum products have shaped modern life. Plastics, lubricants, asphalt, petrochemicals for reagents in chemical synthesis, synthetic fibers for textiles, cosmetics, flavoring and food additives, surfactants and cleaning chemicals are all engrained . The petroleum industry currently benefits from over 100 years of optimization and infrastructure development whereas modern biobased products, i.e. liquid biofuels and renewable coproducts, are comparatively a nascent industry. As such, the developing bioproducts industry can benefit from utilizing petroleum economic strategies such as production of high-value commodities alongside low-value fuels and take advantage of established infrastructure and transportation. These strategies can be employed in tandem with designing optimized fermentation or thermochemical conversion processes suitable for the unique nature of renewable bioproducts.

Liquid fuels are currently necessary to maintain modern transportation and industrial infrastructure. However, the feasibility of producing renewable biofuels is inherently linked to the economic viability of production. In the United States, roughly 90% of crude oil is converted for use as liquid fuels; however, the 7-8% of crude oil that produces high-value chemical commodities accounts for an estimated 25-35% of annual profits (Bozell, 2008). Coproduction of high-value commodities from lignocellulosic sources has recently become a major focus of

research for these reasons. However, examples of coproduction of high-value commodities with biofuels from lignocellulosic feedstocks and broad-scale economic evaluation of these examples are still lacking.

Lemongrass [*Cymbopogon flexuosus* (Steud.) Wats, (syn. *Andropogon nardus* var*. flexuosus* Hack; *A. flexuosus* Nees)] and palmarosa [*Cymbopogon martini* (Roxb.) Wats. var *martinii* (syn. C*. martini* Sapg var. *motia*)] are subtropical essential oil plants (Weiss, 1997). Both crops are produced on large tracts of land on multiple continents. The countries with significant production are: Guatemala, Brazil, China, India, Indonesia, Haiti, Madagascar, and other Eastern African countries. China and Indonesia are the major producers of lemongrass oil accounting for 40% of the world production, which is estimated to be between 800 and 1300 tons of oil year^{-1}. India and Brazil produce the majority of palmarosa oil, which is estimated at around 100 tons of oil year^{-1} for each country (Weiss, 1997).

Terpenoid hydrocarbons (major constituents of lemongrass and palmarosa oils) have been investigated as potential, drop-in ready advanced biofuels for more than three decades (Calvin, 1980; Peralta-Yahya & Keasling, 2010; Pinzi et al., 2011). Therefore, terpenoids have both economic advantages, i.e. a diversity of markets and high-value, in addition to advantages as biofuels, i.e. fungibility with existing liquid fuels and positive low temperature operability properties. Although terpenoids are interesting for these beneficial economic and

fuel advantages, they also typically have antimicrobial activities that may inhibit microbial fermentation in biorefineries. Both citral and geraniol have been shown to have antimicrobial activities against *Saccharomyces cerevisiae* (Helal et al., 2006; Prashar et al., 2003).

This study seeks to explore the use of dual or multi-use biomass. Specific objectives of this study were (1) to evaluate *C. flexuosus* and *C. martinii* biomass yields, oil yields and composition as a function of N and S fertilization to determine yields of oleoresins and biomass for biofuel and high-value coproduct applications, (2) to characterize the cell wall components of the biomass, and its usefulness for fermentation to produce lignocellulosic ethanol from *Saccharomyces cerevisiae*, and (3) to evaluate this system for potential implications in direct-production of advanced biofuels and high-value coproducts in plant biomass.

2. Materials and Methods

2.1. Field experiments

In 2008 and in 2009 cropping seasons, a field experiment was carried out in Verona, Mississippi, USA (34°43’22” N and -88°43’22” W). Certified *C. flexuosus* and *C. martinii* were purchased from Richters (ON, Canada). Transplants were produced in a double-plastic controlled temperature greenhouse during March-

April. Lemon grass and palmarosa seeds were sown in Metromix 300 (The Scotts Co., Marysville, OH) growth medium, in 48-cell plastic trays to provide one transplant per cell. The production of transplants in the greenhouse continued for 45 days. The temperature was maintained at 22 to 25 °C during the day and 18 °C at night. Nutrients were provided with weekly fertilization with 1.8 g of 20-20-20 N-P_2O_5-K_2O in 300 mL of water. In addition, transplants were top irrigated daily. Lemongrass and palmarosa seedlings (approximately 12 and 15 cm, respectively) were transplanted into the field in May 2008 and again in May 2009. Plot size was 1.4 x 6 m, and 12 lemongrass and 12 palmarosa plants were transplanted in each plot, into two rows, at 60 cm in-row spacing on each bed. Beds were spaced at 180 cm apart.

The soil at the experimental site was composed of Quitman sandy loam (fine-loamy, siliceous, semiactive, thermic, Aquic Paleudult). Glyphosate at 2 kg ha^{-1} was applied prior to land preparation, which included disking two weeks after the herbicide application. Prior to land preparation, soil samples (0-15 cm deep, 3 composite samples made of 24 soil cores) were analyzed for extractable nutrients. Before transplanting, phosphorus (P) and potassium (K) fertilizers were applied to ameliorate deficiencies based on soil test reports. Lemongrass and palmarosa plants were planted in previously prepared raised beds (12 cm high and 77 cm wide across the top). The raised beds were prepared by using a press-pan-type bed shaper machine. The machine also placed plastic mulch on

the top of the bed and a drip tape irrigation tube at 2-3 cm soil depth below the soil surface, in the middle of the bed.

The N fertilizers (as ammonium nitrate) and S fertilizer (as sulfur bentonite, 90% S) were applied in the middle of the bed, depending on the treatments, and approximately 2 weeks after transplanting. Aboveground portions of lemongrass and palmarosa plants were harvested using a hedge trimmer at approximately 20 cm above the soil surface. Two harvests were taken from each of the crops every year. In 2008, lemongrass 1st harvest was on Sept 23rd, and the 2nd harvest on Oct 28th, while palmarosa 1st harvest was on Sept 24th and the 2nd harvest on Oct 27th. In 2009, lemongrass 1st harvest was on Sept 28th and the 2nd harvest on Oct 20th, while palmarosa 1st harvest was on Sept 28th, and the 2nd harvest on Oct 26th. Whole, above ground plant parts were weighed, air dried in a shaded greenhouse (at approximately 40 ºC), reweighed to record dry weight, and plants were steam distilled for extraction of essential oil.

BioEnergy Science Center (BESC) *Panicum virgatum* ‘Alamo’ standard was used as an external control. A lot was obtained from the National Renewable Energy Laboratory, Golden, CO. *P. virgatum* ‘Alamo’ seed was purchased from MBS Seeds LTD (Denton, TX) which was produced in 2006 (Lot #6011A) with an effective germination rate of 30%. The standard was grown at the Samuel Roberts Noble Foundation in Ardmore, Oklahoma and planted on June 11, 2007 with 17.8 cm row spacing and a seeding rate of 30%. The stand was harvested

on November 2, 2007 and baled on November 5, 2007. The 0.2 ha stand was fertilized with 99.8 kg of 46-0-0 NPK fertilizer yr^{-1} for a rate 45.9 kg N 0.2 ha^{-1} (229.6 kg N ha^{-1}).

2.3. Essential oil extraction

Essential oil was extracted via steam distillation (300 g from lemon grass and 250 g from palmarosa) in a 2 L steam distillation unit for 60 min as described previously (Zheljazkov et al., 2011; Zheljazkov et al., 2010). The different sample sizes were due to unequal amount of biomass from the two crops that can fit the same 2 L bioflask. Immediately after the end of the distillation, the essential oil was separated, measured on an analytical scale and stored at minus 5 °C for further analysis. The essential oil content of lemongrass and palmarosa was calculated as the weight of oil in g g^{-1} dry plant tissue.

2.4. Gas Chromatography–Mass Spectroscopy Quantitative Analysis

Quantitative analyses of oil samples were performed using gas chromatography–mass spectroscopy (GC-MS) at the National Center for Natural Products Research in Oxford, MS, using the GC-MS methods described in (Zheljazkov et al., 2007). Commercial standards (R)-(+)-limonene (CAS 95327-98-3) and (+)-δ-cadinene (CAS 483-76-1) were purchased from Fluka (Switzerland); citral (geranial and (Z)-citral; CAS 5392-40-5), geraniol (CAS 106-24-1), geranyl acetate (CAS 105-87-3), (-)-trans-caryophyllene (CAS 87-44-5), and

caryophyllene oxide (CAS 1139-30-6) were purchased from Sigma-Aldrich (St. Louis, MO, USA). With five concentration points, an external standard least squares regression for quantification was used. Each specific analyte was used to formulate a separate calibration curve using MS total ion chromatogram (TIC) data. Linearity was imposed by using response factors and regression coefficients independently. Response factors were calculated using the equation RF = DR/*C*, where DR was the detector response in peak area (PA) and *C* was the analyte concentration. Since citral was only available as a mixture of *E* (geranial) and *Z* ((Z)-citral) isomers, the TIC area from both isomers was added together to generate the response factor used for the two individual isomers which were quantified separately using that same RF.

The chromatograms of each of the essential oil samples from the field experiments were compared to the chromatograms from standards. Target analytes were confirmed by retention time and mass spectra. Confirmed integrated peaks were used to determine percentage of each chemical constituent in the essential oil itself. The RF of the target chemical constituent was used to determine the percentage of oil for each sample using the equation PA/RF/*C* x 100 = % analyte in the oil on a wt/wt basis.

2.5. Statistical analysis

Repeated measures analysis of the data collected from the experiments conducted at Verona, Mississippi in 2008 and 2009 were analyzed together with

six blocks, composed of the combinations of the two years and the three blocks in the field. The two harvests were used as the time points for the repeated measures analysis completed using the Mixed Procedure of SAS (Institute, 2003). The analysis was completed using two models; one with both crops in which comparison of the crops is made in terms of the responses, and one for each crop where such and between-crop comparison is not made. We, therefore, had three separate analyses: 1) for dry weight yield and essential oil yield responses, a 2×4×4 factorial in six blocks, with the factors of interest being crop (lemongrass [LG], palmarosa [P]), N (0, 40, 80, 160 kg ha^{-1}), and S (0, 30, 60, 90 kg ha^{-1}) and the repeated measures factor, harvest (1, 2) was used; 2) for essential oil (EO) content, and the composition and yield of β-caryophyllene, (Z)-citral and geranial of lemongrass, a 4×4 factorial in six blocks; and 3) for essential oil content, and the composition and yield of geraniol and geranylacetate of palmarosa, a 4×4 factorial in six blocks were used. For both 2) and 3) models, the factors of interest were N (0, 40, 80, 160) and S (0, 30, 60, 90) applications, and Harvest (1, 2) as the repeated measures factor. For each response, the validity of model assumptions was verified by examining the residuals as described (Montgomery, 2008). Some of the responses required cubic root transformation to achieve normality of the error terms, however, the means shown in the tables and figures are back-transformed to the original scale. For significant (p-value < 0.05) and marginally significant (p-value between

0.05 and 0.1) effects, further multiple means comparison was completed for by comparing the least squares means of the corresponding treatment combinations. Letter groupings were generated using a 1% level of significance for two-factor and three-factor interaction effects, and using a 5% level of significance for main effects.

2.6. Pretreatment, ethanol fermentation, and high performance liquid chromatography analysis

Dried *C. martinii* and *C. flexuosus* biomass samples from field trials were ground in a Wiley mill with a 1 mm screen. A portion of whole dried plant biomass was separated into leaf and stem biomass for biomass composition analysis. Bench-scale ethanol fermentations and pretreatment of biomass were conducted as described previously (Fu et al., 2011). Unless stated otherwise, 3.0 g (15% of total fermentation volume) of dried biomass was loaded into each 70 mL Septi-Chek glass vials (Becton Dickson, Franklin Lakes, NJ), and fermentations were done in triplicate. The fermentation microbe was *S. cerevisiae* D5α culture and enzyme was Accellerase 1500 (Genencor, Rochester, NY). The solution was diluted to 20 mL total volume, weighed for calculation of weight loss in later time points, and placed in a 37 °C incubator. At each time point, the bottles were removed from the incubator, and the cap of the bottle was pierced with a needle to allow carbon dioxide to escape from the closed vessel. The remaining weight was then recorded and weight loss was calculated.

After the fermentations were complete, the remaining biomass solutions were centrifuged and 1.0 mL of the supernatant was removed with a 1 mL syringe and filtered through a 13 mm syringe 0.2 μm filter (Millipore, Massachusetts). Ethanol and fermentation liquids were quantified by HPLC (Agilent 1200 Series LC system with 1200 Series refractive index detector) equipped with an Aminex HPX-87P column (Agilent Technologies).

2.7. Determination of saccharification efficiency and lignin composition

Enzymatic saccharification efficiency of all biomass samples was carried out using a high-throughput plate hydrothermal pretreatment and enzymatic saccharification procedure developed at the National Renewable Energy Laboratory (Decker et al., 2009). Samples were run twice with three technical replicates for each run. Lignin composition, content, and S/G ratio was determined using pyrolysis molecular beam mass spectrometry (PyMBMS) method as previously described (Sykes et al., 2009). Samples were run with three technical replicates for each plant for enzymatic saccharification and with two technical replicates for PyMBMS.

3. Results and Discussion

Currently, lignocellulosic biofuel production models have a similar logical flow despite the many different technological systems that are being investigated.

First, plant feedstocks are grown, collected, processed, and transported to biorefinery facilities. Second, plant feedstocks are converted into fuels and potential coproducts either through biological methods, e.g. saccharification and fermentation, or through chemical methods, e.g. thermochemical conversion. Last, products from biomass are converted into final market products and unused biomass fractions are processed as waste and most commonly burned to generate heat (Dutt et al., 2007). To date, plant biofuel feedstock research has focused primarily on reduction of lignocellulosic feedstock recalcitrance and increasing overall biomass yield. Likewise, investigation into mechanisms to produce chemicals of interest, i.e. biofuels and coproducts, has occurred almost exclusively at the microbial and thermochemical conversion steps. However, this logical flow ignores the potential for plant feedstocks to produce myriad biochemicals that can be used as biofuels or coproducts directly from sunlight.

We sought here to examine the capability of relatively novel feedstocks to yield biofuel and established high-value coproducts at a pilot-scale system in tandem. *Cymbopogon flexuosus* and *C. martinii* offer a unique opportunity for agronomic production of biofuel and coproducts. Though *C. flexuosus* and *C. martinii* are native to Southeast Asia, we have selected *C. flexuosus* and *C. martinii* specifically as an agronomically-relevant model system as their agricultural production mirrors switchgrass, they naturally produce high-value coproducts, and the high-value coproducts are composed of terpenoids that have

been proposed for use as advanced biofuels. Additionally, the terpenoids in *C. flexuosus* and *C. martinii* essential oils have been directly linked to *S. cerevisiae* inhibition and toxicity (Helal et al., 2006; Prashar et al., 2003). Taken together, these species offer two attractive model systems to study *in planta* production of advanced biofuels and coproducts.

However, several considerations need to be investigated to understand the feasibility of direct-coproduct biosynthesis in lignocellulosic feedstock crops whether through breeding or biotechnological modification: 1) agronomic inputs and resulting yields that offset costs of inputs, 2) biochemical coproduct inhibition of downstream fermentation processing for biofuel production, and 3) economically feasible ways to remove plant-derived coproducts. Here we investigate the first two considerations for a dual use lignocellulosic crop biomass, i.e. lemongrass and palmarosa, for production of biofuels and low-volume high-value essential oil coproduct.

3.1. C. flexuosus and C. martinii biomass and essential oil yield in field trials

Small plot field trials were conducted to determine growing conditions of *C. flexuosus* and *C. martinii* in the growing regions of the United States. The interaction effect of crop and harvest was significant on dry weight yields and essential oil yield, whereas, the interaction effects of crop and N, and N and harvest were significant on dry weight yield, but not on essential oil yield (Table

17). Overall, increasing N application rates increased dry weight yields of both lemongrass (ranged from 7 673 to 15 196 kg ha^{-1}) and palmarosa (ranged from 11 078 to 19 006 kg ha^{-1}), with palmarosa producing more biomass than lemongrass within each application rate of N (Fig. 21A). The main effect of N was also significant on essential oil yield (Table 17). Increasing N application rate brought a stepwise increase in essential oil yields of both crops (Fig. 21B). The interaction of N application rate and harvest increased dry weight yields of both crops (Fig. 21C). It has been reported that increasing N application rates caused a linear increase in yields of lemongrass biomass up to 150 kg N ha^{-1} (Singh, 2001; Singh et al., 1996). Also, (Singh, 1999) reported that under irrigated conditions, 100 kg N ha^{-1} would provide optimal lemongrass yields, while 75 kg N ha^{-1} to 80 kg N ha^{-1} would be sufficient under non-irrigated conditions. Moreover, (Zheljazkov et al., 2011) tested 4 N rates and found 80 kg N ha^{-1} to be sufficient for optimal biomass yields of lemongrass under conditions in Mississippi. Also, (Ram et al., 1997) concluded 100 kg N ha^{-1} for the establishment year and 150 kg N ha^{-1} for the second growing year optimized palmarosa biomass yield. Contrary to these reports, the current work suggests that both lemongrass and palmarosa biomass yields increase with increasing N fertilization. Wullschleger suggested that optimum biomass production for lowland switchgrass would require approximate N fertilization rates of 100 kg N ha^{-1} (Wullschleger et al.,

2010). Thus, lemongrass and palmarosa will respond similarly to switchgrass under the same management practices.

Within a crop and N application rate, harvest 2 provided higher yields than harvest 1 (Fig. 21E). Dry biomass yields from the second year of lemongrass and palmarosa field trials had a mean yield of 12.83 Mg ha^{-1} and 15.11 Mg ha^{-1}, respectively (Fig. 21C). Upland switchgrass cultivars have been reported to have a mean yield (8.7 ±4.2) Mg ha^{-1} whereas lowland switchgrass cultivars yield (12.9 ±5.9) Mg ha^{-1} across growing ranges, harvest years, land quality, stand size, precipitation, and other agronomic variables (Wullschleger et al., 2010). Biomass yield of lemongrass and palmarosa in Mississippi is therefore comparable to switchgrass production ranging from the mean of upland varieties to exceeding upland switchgrass cultivar means depending on nitrogen fertilization rate, crop, and harvest year. However, the lemongrass and palmarosa plots were irrigated in this trial, whereas most switchgrass plots are rainfed which likely had a positive effect on the lemongrass and palmarosa biomass yields.

Generally, essential oil yields of lemongrass were greater than the oil yields of palmarosa within each harvest: 56.1 kg ha^{-1} for lemongrass essential oil and 50.4 kg ha^{-1} for palmarosa from harvest 1, and 85.7 and 67.0 kg ha^{-1}, respectively, from harvest 2 (Fig. 21D). In general, palmarosa and lemongrass essential oil content (% biomass) and essential oil yields (kg ha^{-1}) in Mississippi were similar to those reported for traditional growing countries such as southern

India (Singh, 2001; Singh & Sharma, 2001). Palmarosa yields were also comparable to yields obtained from 4 harvests and eight N applications (Ram et al., 1997; Rao, 2001), and palmarosa yields from 6 harvest (Kulkarni, 1994). Hence, lemongrass and palmarosa in the Southeastern US can provide similar productivity to traditional producing regions in the world. Lemongrass and palmarosa frost hardiness will dictate their growing range in the US. In a previous study, lemongrass had a 30% winter survival after the first year of transplanting into the field in plant hardiness zone 8b, but only 1% in zone 7b (Zheljazkov et al., 2011). Therefore, the perennial growing range of these crops will be from southern Texas to South Carolina.

3.2. Lemongrass essential oil responses

Additionally, the essential oil characteristics of lemongrass and palmarosa grown on the small field plots were investigated to determine potential yield of high-value secondary coproducts and their response to agronomic parameters such as fertilization rates. The essential oil content of lemongrass was higher in the N0 and N40 treatments (0.66 and 0.65%, respectively), and lower in the N160 treatment (0.61%), indicating that increasing N rates may reduce essential oil content (Fig. 22A). The essential oil content was significantly higher in S0, S30 and S60 rates than in the S90 rate (Fig. 22B). The essential oil content from harvest 2 was higher than that from harvest 1, 0.67 and 0.61%, respectively (Fig. 22C). Previously, lemongrass essential oil content was reported to vary between

0.55 and 1.03% (Sarma & Sarma, 2005) with some selected clones reaching essential oil up to 1.3-1.5% (Kulkarni et al., 1992). However, such high-essential oil content clones did not seem to get established as most reports had essential oil content similar this study. For example, lemongrass essential oil content was found to vary between 0.35 to 0.6% of the dried biomass (Zheljazkov et al., 2011). N, S, and harvest, separately had a significant effect on lemongrass essential oil; but they all interacted on the concentrations of β-caryophyllene and (*Z*)-citral (Table 18). The main effect of S was significant, and the interaction effect of N and harvest was marginally significant on the concentration of (*E*)-citral; the main effect of N, and the interaction effect of S and harvest were significant on β-caryophyllene yield; whereas, N and harvest were individually significant on the yields of (*Z*)-citral and geraniol (Table 18).

The yield of (*E*)-citral was higher in the N80 and N160 (32.5 and 36.2 kg ha^{-1} respectively), lower in N40, and lowest in N0 (20.4 kg ha^{-1}) rate (Fig. 23D). The concentration of (*E*)-citral was higher in S30 and S90 and lower in the S0 and S60 rates, indicating S30 as a possible optimum rate for maximizing the yield and the concentration of some oil constituents (Fig. 22E). The yields of (*E*)-citral was higher from harvest 2 than from harvest 1 (Fig. 22F). The yields of (*Z*)-citral was highest in N160 (28.4 kg ha^{-1}), lower in N80, a step lower in N40 and the lowest in N0 (15.6 kg ha^{-1}) application rate (Fig. 22G). The yields of (*Z*)-citral was higher from harvest 2 than from harvest 1 (Fig. 22H).

The yield of *β*-caryophyllene was higher at N80 (1.18 kg ha^{-1}) and lower in N0 and N40 (0.68 and 0.81 kg ha^{-1} respectively) rates (Fig. 22I). Increased sulfur reduced β-caryophyllene yield in both harvest years (Fig. 22J). The three-way interaction of N, S, and Harvest affected the concentrations of *β*-caryophyllene and (*Z*)-citral in the lemongrass essential oil (Fig. 26). For example, with harvest 1 and N160 rate, the addition of S increased the concentration of β-caryophyllene relative to the S0 rate. However, within harvest 2 and N80 rate, S application at 90 kg ha^{-1} had a lower concentration of β-caryophyllene relative to no application of S. Increased N or S fertilization application had lower mean essential oil content (%) of lemongrass biomass, but increased N and S fertilization application had higher mean yield (kg ha^{-1}) of most essential oil components (Fig. 22D, E, G, I, J). This could have resulted from increased biomass yield (dry kg ha^{-1}) from fertilization with only a slight reduction in essential oil content of the biomass. Essential oil composition varied in reports from India with 78-95% citral in lemongrass oil (Sarma & Sarma, 2005). Some lemongrass clones are rich in (*E*)-citral (66-73%), with approximately 10% (*Z*)-citral, and 8-10% linalool (Kulkarni et al., 1992). N application rates, other agroecological conditions, and even leaf position within one plant were shown to alter essential oil composition of lemongrass (Singh et al., 1989). Additionally, as the response of β-caryophyllene and (*Z*)-citral concentrations in our study did not have a direct

trend in relation to fertilization rates we may assume the presence of other factors modifying the concentrations of these two terpenes in lemongrass oil.

3.3. Palmarosa essential oil responses

Harvest year had a significant effect on essential oil content of palmarosa biomass which was higher in harvest 2 than harvest 1 (Fig. 23A). Harvest year also had a significant effect on geranylacetone content of palmarosa essential oil (Fig. 23B). Interaction between N and S had significant effect on the concentration of geraniol in palmarosa essential oil (Fig. 23C). Geraniol yield was positively affected by harvest (Fig. 23D) and nitrogen and sulfur interactions (Fig. 23E). Geranyl acetate yield was negatively affected by harvest (Fig. 23F) and nitrogen and sulfur interaction (Fig. 23G). Geraniol concentration in palmarosa oil ranged between 70-85% and geranyl acetate varied from 4 to 15% (Rajeswara Rao et al., 2009). Other studies reported geraniol at 82% and geranyl acetate of 10% [29], or up to 93% geraniol, 3-4% linalool, and 2% geranyl acetate (Chowdhury et al., 1998; Rao et al., 2005). Lemongrass and palmarosa oil composition can be altered by various and numerous factors; our study demonstrated some of these alterations. The general biosynthetic pathways of lemongrass and palmarosa oil were recently reviewed by (Ganjewala & Luthra, 2010).

3.4. Biomass composition and recalcitrance

The composition of lemongrass and palmarosa biomass was investigated to determine feasibility of use in lignocellulosic biofuel and high-value coproducts fermentation. Lignin has been identified as a major inhibitor of lignocellulosic biomass fermentation. Specifically, recalcitrance of lignocellulosic biomass to enzyme degradation and fermentation has been linked to the overall S/G ratio of lignin subunits in biomass (Fu et al., 2011). Palmarosa biomass had the largest lignin fraction and the highest S/G ratio (Table 15). Stem tissues of both lemongrass and palmarosa biomass had greater lignin fraction and a higher S/G ratio than leaf tissue. Average glucose yield from lemongrass whole biomass was about double that of palmarosa whole biomass (Table 15). However, palmarosa whole biomass had 22.2% more cellulose content than lemongrass whole biomass. This suggests that palmarosa biomass should produce more ethanol than lemongrass if complete hydrolysis is achieved and no inhibition results from hydrolysis products or native metabolites. Xylan content was similar for both lemongrass and palmarosa biomass (Table 15). Therefore, the differences seen in ethanol production are not likely due to xylan content although xylans have previously been found to inhibit biomass hydrolysis by endoglucanases and cellobiohydrases (Zhang et al., 2012).

Palmarosa recalcitrance to enzymatic degradation was mirrored in principal during SSF optimization. Lemongrass biomass showed a coordinated

response to enzyme dosage with a plateau between 15 (FPU g^{-1} biomass) and 20 (FPU g^{-1} biomass), whereas palmarosa biomass yielded a maximum between 10 (FPU g^{-1} biomass) and 15 (FPU g^{-1} biomass) (Fig. 24). Prior steam distillation extraction of essential oils did not affect lignin fraction, lignin S/G ratio, enzymatic saccharification cellulose release, or enzymatic saccharification xylose release (Table 15). These results suggest that the steam distillation process to remove essential oils does not pretreat biomass.

3.5. Ethanol fermentation and inhibition from biomass metabolites

Benchtop-scale simultaneous saccharification and fermentation (SSF) were carried out to investigate the potential to produce biofuels, i.e. ethanol in this work, from lemongrass and palmarosa biomass and how the role of secondary metabolites found in these two species might affect biofuel production efficiency. Lemongrass extracted (EX) biomass yielded less ethanol than not extracted biomass (NE), but EX palmarosa biomass yielded more ethanol than NE palmarosa biomass (Fig. 25A). These patterns were the same in biomass treated with and without enzyme, suggesting that results were from biomass properties rather than inhibition of enzymatic hydrolysis. Lemongrass and palmarosa essential oils were found to interact with *S. cerevisiae* cell membranes and cause the leakage of ions until cellular death (Helal et al., 2006; Prashar et al., 2003). These two reports found a concentration of 0.1% for either lemongrass or palmarosa essential oil was toxic to *S. cerevisiae*. The major essential oil

constituents of lemongrass (citral) and palmarosa (geraniol) have low solubility in water because they are primarily nonpolar terpenoid hydrocarbons; however, geraniol is 38.8% more soluble in water, 420 mg L^{-1} and 686 mg L^{-1} at 20 ºC respectively (GESTIS database, October 2012). Therefore, fermentation of palmarosa biomass would result in *S. cerevisiae* toxicity in relatively lower biomass concentrations than lemongrass or switchgrass as observed (Fig. 25A). These considerations support two different mechanisms for the difference in ethanol fermentation potential. First, fermentable glucose in lemongrass biomass is lost by processing during steam distillation which leads to higher ethanol yields from NE biomass. Secondly, the essential oil present in palmarosa inhibits *S. cerevisiae* fermentation leading to higher ethanol yields from biomass that has had the essential oils removed.

Overall, both extracted and not-extracted lemongrass biomass yielded similar amounts of ethanol as BESC switchgrass and palmarosa biomass yielded the least amount of ethanol of all the biomass types tested. The difference in ethanol yield between the two lots of BESC switchgrass can be explained further by the remaining biomass fractions in the fermentation liquid. Lot #1 had three times the amount of residual cellulose (glucose) after fermentation compared to lot #2 (Table 19). BESC switchgrass lot #2 had a similar amount of cellulose leftover after fermentation as lemongrass and palmarosa biomass.

Pretreatment of biomass universally increased production of ethanol from biomass (Fig. 25B; Table 19). It was unclear whether pretreatment of biomass would result in increased concentrations of inhibitory secondary metabolites. After pretreatment, extracted lemongrass and palmarosa biomass produce more ethanol than biomass that had not been extracted (Fig. 25B). Palmarosa biomass turned a dull red after dilute acid pretreatment whereas all other biomass samples remained brown (data not shown). Only NE palmarosa biomass had remaining glucose in the fermentation liquid of all pretreated biomass (Table 20). This supports the previous observation that palmarosa biomass will yield toxic concentrations of essential oils at lower biomass concentrations. In short, pretreatment likely breaks open more cells which would provide better enzyme access in palmarosa biomass resulting in increased concentrations of essential oils in fermentation liquids. In future experiments, analysis of fermentation liquids for essential oil metabolites would yield interesting observations and help to predict specific concentrations of coproducts that are inhibitory to biofuel production in larger processes.

3.6 From field to fermenter: estimated market values of coproduction

Finally, it is important to consider the potential of high-value, low-volume coproducts to offset the inherent low-value, high-volume economics of biofuel production. Coupling our biomass and essential oil agronomic data and the SSF ethanol production data the market value of lemongrass and palmarosa biomass

can be estimated. The second harvest year had a mean biomass yield of 128 000 kg ha^{-1} for lemongrass and 151 000 kg ha^{-1} for palmarosa. Pretreated lemongrass yielded 156 mg g^{-1} biomass or a volume of 198 mL g^{-1} biomass considering ethanol's density of 0.789 g mL^{-1} (Fig. 25B). Pretreated palmarosa had an ethanol yield of 134 mg g^{-1} biomass or 170 mL g^{-1} biomass. This equates to a volume of ethanol yield of 2541 L ha^{-1} for lemongrass and 2569 L ha^{-1} for palmarosa biomass, or $1600 and $1620, respectively, for current spot ethanol prices at $0.63 L^{-1} (Chicago Board of Trade October 2012). The ethanol production for these dual use crops falls within the theoretical maximum ethanol production (100% conversion assumed) range, 1749 L ha^{-1} to 3691 L ha^{-1}, for switchgrass (Schmer et al., 2012) and compares to actual observed yield means from switchgrass, 3091 L ha^{-1} (Vogel et al., 2011). A total of $1950 ha^{-1} can be produced from switchgrass fields, assuming 3091 L ha^{-1} for average switchgrass production as in (Vogel et al., 2011).

Previous reports in India note a value of $10.00 (USD) for lemongrass essential oil per kilogram and $15.00 for palmarosa essential oil (Rao et al., 2005; Sarma & Sarma, 2005). It is likely that these essential oils would have a greater value in the United States or in international markets (Singh et al., 1996). The mean value for the second-year harvest for lemongrass essential oil yield was 85.7 kg ha^{-1} and 67 kg ha^{-1} for palmarosa. Therefore, lemongrass would yield $857 ha^{-1} and palmarosa would yield $1005 ha^{-1} in essential oil coproduct

sales. At a refinery, this translates to an additional \$66.80 Mg^{-1} for lemongrass and \$66.51 Mg^{-1} for palmarosa based on the mean biomass production values per hectare. Taken together with the above spot ethanol prices of \$0.63 L^{-1} a total of \$2457 lemongrass ha^{-1} and \$2625 palmarosa ha^{-1} for ethanol and essential oil production can be produced while only \$1950 for switchgrass spot ethanol ha^{-1} alone can be realized currently. While these results help to quantify the value of coproducts to the biofuels industry, these estimates are derived from small plots and bench-top ethanol fermentation scales. In other work plot size has previously been found to not skew biomass yield data (Wullschleger et al., 2010). However, further investigation across multiple years and growing climates will be needed to determine whether lemongrass and palmarosa biomass could be used as feasible dual use lignocellulosic feedstocks.

4. Conclusions

We report agronomic production, essential oil, and ethanol production from two novel dual use lignocellulosic crops. Extrapolation of the results lead to an ethanol yield of 2541 L ha^{-1} of lemongrass and 2569 L ha^{-1} of palmarosa biomass with an additional essential oil yield of 85.7 kg ha^{-1} and 67 kg ha^{-1}. This leads to a combined value of \$2457 ha^{-1} lemongrass and \$2625 ha^{-1} palmarosa for ethanol and essential oil compared to \$1950 for switchgrass spot ethanol ha^{-1} alone. These results support the potential value of coproduct economics in the

emerging biofuel industry, and have identified two feasible dual uses for biofuel and coproduct commercialization.

Acknowledgements

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 thank them for funding. This field research was funded in part by ARS Specific Crop Agreement 58-6402-026 with Mississippi State University. Specific project: “Field establishment of medicinal herbs and potential for commercial production” awarded to Dr. V.D. Jeliazkov (Zheljazkov). We thank Thomas Horgan, S. Marie Rogers, and Amber Reichley for excellent help with experiments and analysis.

APPENDIX

Tables

Table 18. Lemongrass and palmarosa biomass without pretreatment fraction composition and enzymatic saccharification efficiency.

Biomass Type	Fraction	Lignin (%)	S/G ratio	Biomass Cellulose content (g g^{-1})	Biomass cellulose release (g g^{-1})	Average Cellulose Yield† (%)	Biomass xylan content (g g^{-1})	Biomass xylan release (g g^{-1})	Average Xylan Yield† (%)
LNE	Whole	19.87	0.48	0.2827	0.1629	51.86	20.06	0.1411	61.88
	Leaf	18.77	0.46	0.2713	0.1768	58.67	18.48	0.1506	71.70
	Stem	22.23	0.54	0.3486	0.1863	48.10	20.02	0.1404	61.68
LEX	Whole	21.39	0.46	0.3024	0.1706	50.77	20.80	0.1483	62.72
	Leaf	20.30	0.44	0.2463	0.1663	60.79	18.62	0.1472	69.56
	Stem	24.20	0.56	0.3861	0.1758	40.97	38.61	0.1511	64.19
PNE	Whole	23.76	0.65	0.3676	0.1165	28.53	19.67	0.1298	58.06
	Leaf	16.91	0.42	0.2276	0.1872	74.03	16.82	0.1218	63.75
	Stem	27.57	0.77	0.4158	0.0850	18.40	19.85	0.1289	57.13
PEX	Whole	24.06	0.68	0.3885	0.1174	27.19	20.52	0.1339	57.42
	Leaf	18.79	0.48	0.2714	0.1710	56.71	17.49	0.1281	64.49
	Stem	28.79	0.84	0.3804	0.0950	22.49	20.74	0.1380	58.57

L – Lemongrass

P – Palmarosa

NE – not-extracted

EX – extracted biomass

† $\text{Average Yield} = \text{Release (g g}^{-1}\text{ cellulose or xylan)} / \text{Content (g g}^{-1}\text{ biomass)}$

Table 19. The repeated measures main and interaction effects of crop, nitrogen (N), sulfur (S) fertilization, and harvest on biomass dry weight yield (DW) and essential oil composition (EO).

Source	DW yield (kg ha^{-1})	EO yield (kg ha^{-1})
Block	0.001	0.553
Crop	0.001	0.001
N	0.001	0.001
Crop×N	0.065	0.451
S	0.215	0.654
Crop×S	0.742	0.212
N×S	0.263	0.184
Crop×N×S	0.971	0.558
Harvest	0.001	0.001
Crop×Harvest	0.001	0.017
N×Harvest	0.043	0.484
Crop×N×Harvest	0.663	0.102
S×Harvest	0.288	0.778
Crop×S×Harvest	0.529	0.979
N×S×Harvest	0.688	0.758
Crop×N×S×Harvest	0.469	0.743

Table 20. The effect of N, S and harvest on lemongrass EO content, and the composition and yield of β-caryophyllene, (Z)-citral and (E)-citral.

Source	EO content of biomass (%)	Content in essential oil (%)			Yield (kg ha^{-1})		
		β-caryophyllene	(Z)-citral	(E)-citral	β-caryophyllene yield	(Z)-citral yield	(E)-citral yield
Block	0.128	0.069	0.270	0.136	0.071	0.811	0.661
N	0.029	0.009	0.832	0.262	0.001	0.001	0.001
S	0.024	0.688	0.025	0.001	0.153	0.704	0.262
N×S	0.316	0.032	0.649	0.458	0.157	0.514	0.281
Harvest	0.002	0.001	0.964	0.367	0.007	0.046	0.047
N×Harvest	0.587	0.432	0.382	0.086	0.506	0.832	0.670
S×Harvest	0.648	0.003	0.390	0.703	0.013	0.245	0.236
N×S×Harvest	0.944	0.089	0.086	0.236	0.416	0.277	0.127

Table 21. P-values showing the effect of N, S and harvest on palmarosa EO content, and the composition and yield of geraniol and geranylacetate.

Source	EO content of biomass (%)	Content in essential oil (%)		Yield (kg ha^{-1})	
		Geraniol	Geranyl acetate	Geraniol yield	Geranyl acetate yield
Block	0.006	0.467	0.001	0.001	0.001
N	0.598	0.474	**0.001**	0.001	0.001
S	0.958	0.318	0.274	0.394	0.142
N×S	0.927	**0.001**	0.451	**0.003**	**0.034**
Harvest	**0.001**	0.705	**0.001**	**0.001**	**0.028**
N×Harvest	0.444	0.962	0.701	0.718	0.574
S×Harvest	0.390	0.526	0.970	0.753	0.953
N×S×Harvest	0.831	0.974	0.444	0.842	0.451

Table 22. Biomass fractions remaining in fermentation liquid after SSF for enzyme optimization, dried harvested biomass, and pretreated biomass as determined by HPLC.

Enzyme Optimization	Enzyme Concentration (FPU g^{-1} biomass)	Biomass Fraction Remaining in Fermentation Liquid (mg g^{-1} biomass)					
		Cellobiose	Glucose	Xylose	Arabinose	Acetic acid	Ethanol
Lemongrass Extracted	0	3.71	0.97	0.00	1.58	2.16	20.05
	10	1.85	1.75	10.16	3.30	4.59	40.46
	15	3.46	0.00	16.06	4.76	6.34	63.16
	20	3.01	1.83	13.10	5.54	7.33	72.24
Palmarosa Extracted	0	1.44	0.00	0.00	1.23	1.85	20.10
	10	2.04	2.02	11.17	3.64	4.83	44.79
	15	1.13	2.17	20.89	4.77	4.66	47.16
	20	2.12	2.14	13.73	4.26	6.27	49.54
No Pretreatment Biomass Fermentation	**Enzyme Concentration (FPU g^{-1} biomass)**	**Cellobiose**	**Glucose**	**Xylose**	**Arabinose**	**Acetic acid**	**Ethanol**
Lemongrass Extracted	15	3.04	0.00	13.48	6.02	7.39	70.38
	0	3.77	1.13	0.00	1.91	2.13	20.30
Lemongrass Not-extracted	15	3.57	1.50	14.99	5.70	7.00	72.45
	0	5.08	1.27	9.25	1.88	3.25	28.30
Palmarosa Extracted	15	2.07	1.82	14.97	4.80	5.66	47.38
	0	3.52	2.07	0.00	1.57	2.13	19.89
Palmarosa Not-extracted	15	2.43	1.31	15.57	4.63	4.66	37.72
	0	3.59	1.60	0.00	1.26	1.77	11.48
Switchgrass Lot #1	15	1.21	3.22	13.87	3.34	6.82	52.11
No Biomass	15	0.31	0.00	0.46	0.00	0.00	0.00
Switchgrass Lot #2	15	1.89	1.11	11.29	3.76	5.91	67.64

Table 22. Continued.

Pretreated Biomass	Enzyme Concentration (FPU g^{-1} biomass)	Cellobiose	Glucose	Xylose	Arabinose	Acetic acid	Ethanol
Lemongrass Extracted	15	1.61	0.00	12.88	0.83	14.11	156.27
Lemongrass Not-extracted	15	1.79	0.00	18.55	1.55	14.29	124.21
Palmarosa Extracted	15	1.36	0.00	13.31	0.85	9.99	134.15
Palmarosa Not-extracted	15	0.85	1.77	13.22	0.68	6.39	90.11
Switchgrass Lot #1	15	2.39	0.00	21.42	1.03	12.36	108.23
No Biomass	15	0.76	0.00	1.39	0.00	0.00	0.00

Figures

Figure 23. Main effects and interaction plots for dry weight biomass (kg ha^{-1}) and essential oil (kg ha^{-1}) production across both crops (lemongrass and palmarosa).

For each plot means sharing the same letter are not significantly different.

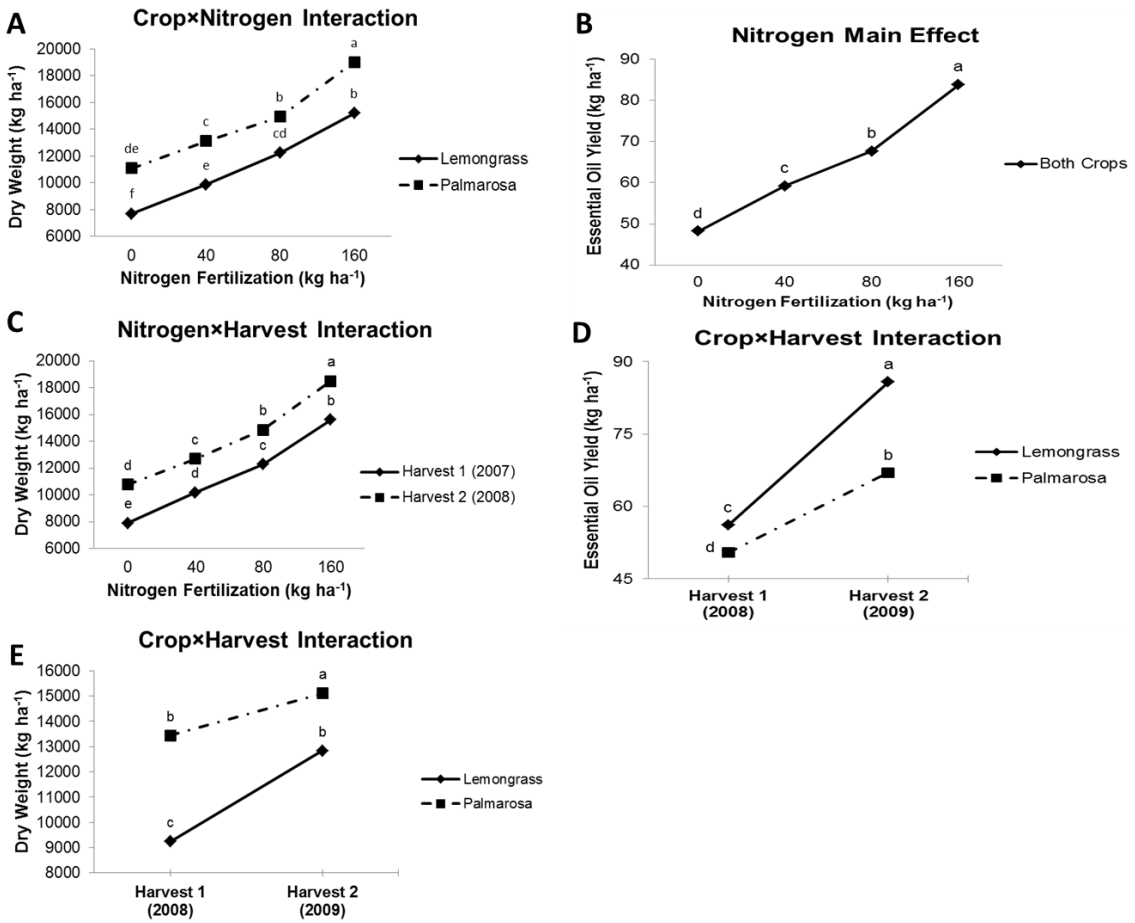
A
Crop×Nitrogen Interaction
Dry Weight (kg ha-1)
Nitrogen Fertilization (kg ha-1)
Lemongrass
Palmarosa
B
Nitrogen Main Effect
Essential Oil Yield (kg ha-1)
Nitrogen Fertilization (kg ha-1)
Both Crops
C
Nitrogen×Harvest Interaction
Dry Weight (kg ha-1)
Nitrogen Fertilization (kg ha-1)
Harvest 1 (2007)
Harvest 2 (2008)
D
Crop×Harvest Interaction
Essential Oil Yield (kg ha-1)
Harvest 1 (2008)
Harvest 2 (2009)
Lemongrass
Palmarosa
E
Crop×Harvest Interaction
Dry Weight (kg ha-1)
Harvest 1 (2008)
Harvest 2 (2009)
Lemongrass
Palmarosa

Figure 23.

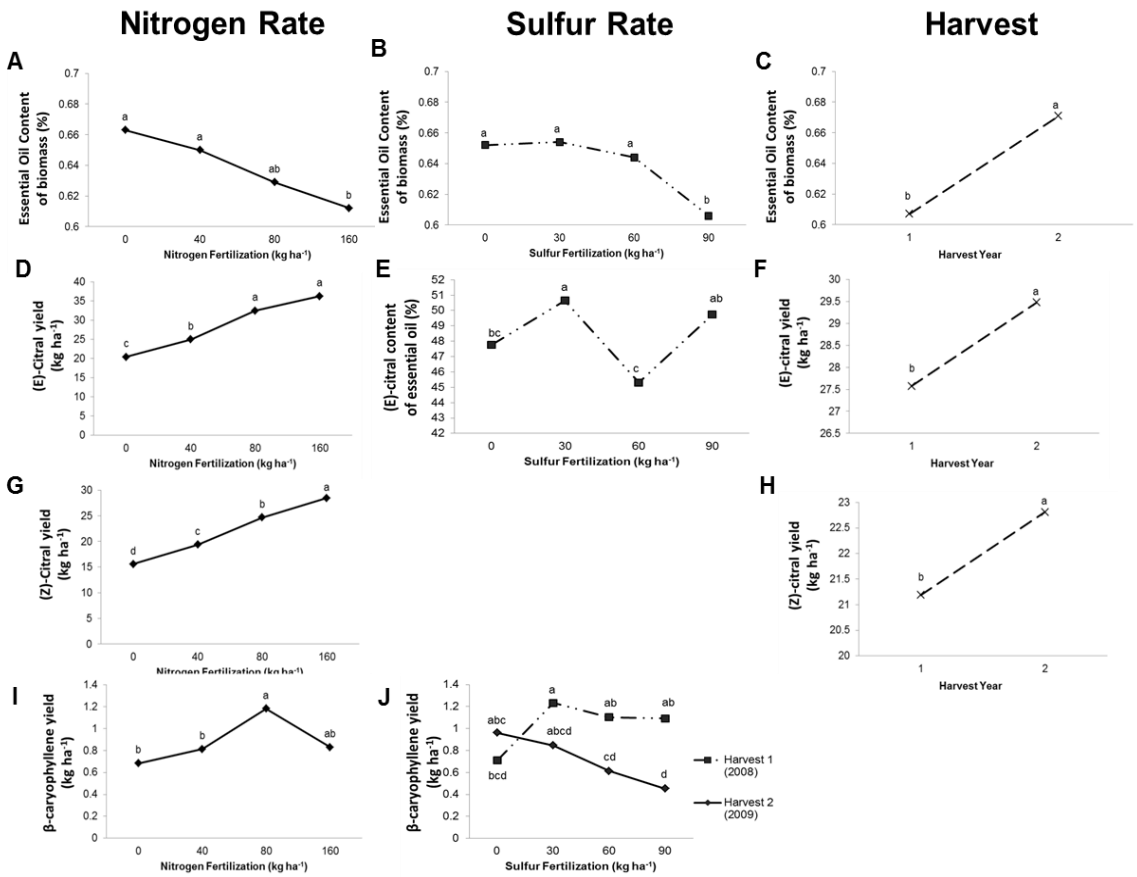

Figure 24. Means of lemongrass essential oil content, (*E*)-citral content and yield, (*Z*)-citral yield, and â-caryophyllene yield.

For each plot means sharing the same letter are not significantly different.

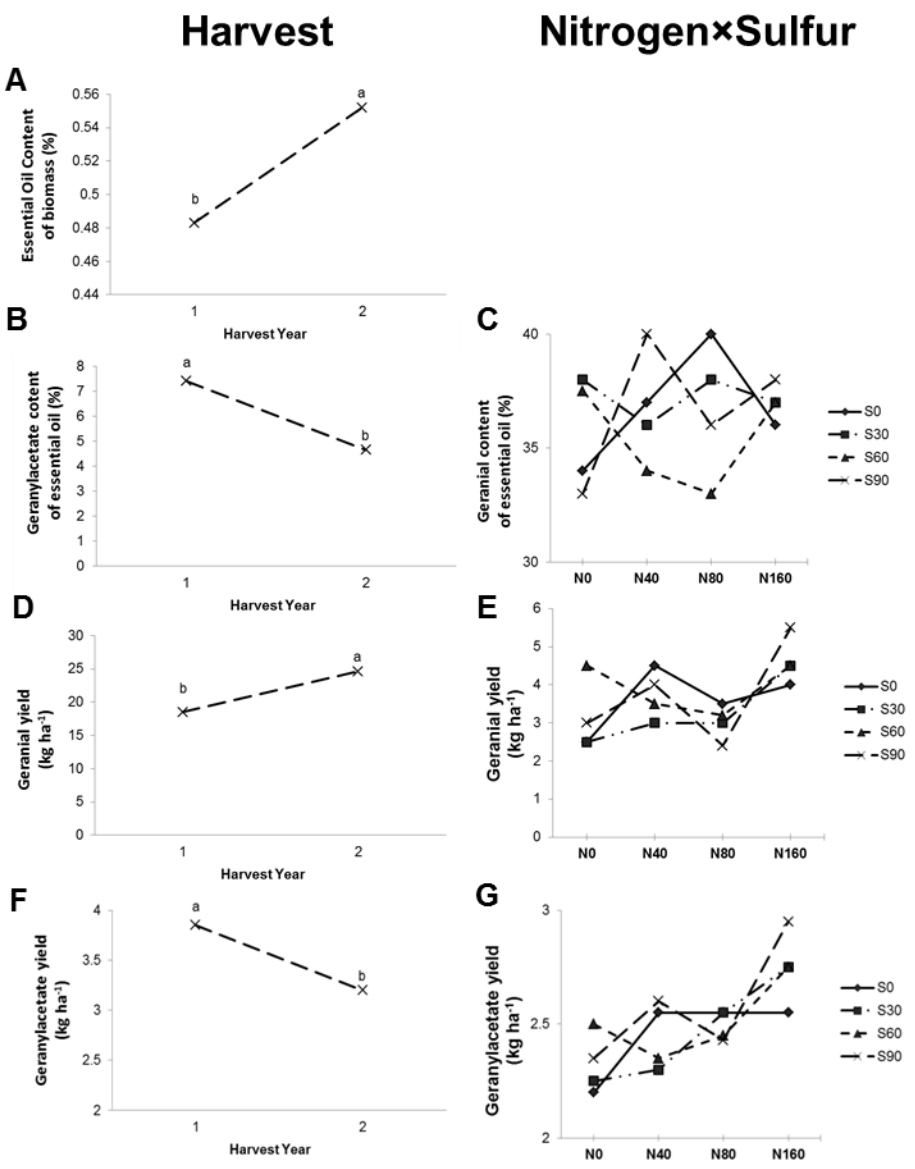

Figure 25. Main effects and interaction plots palmarosa essential oil content, geranylacetate content and yield, and geraniol content and yield.

For each plot means sharing the same letter are not significantly different.

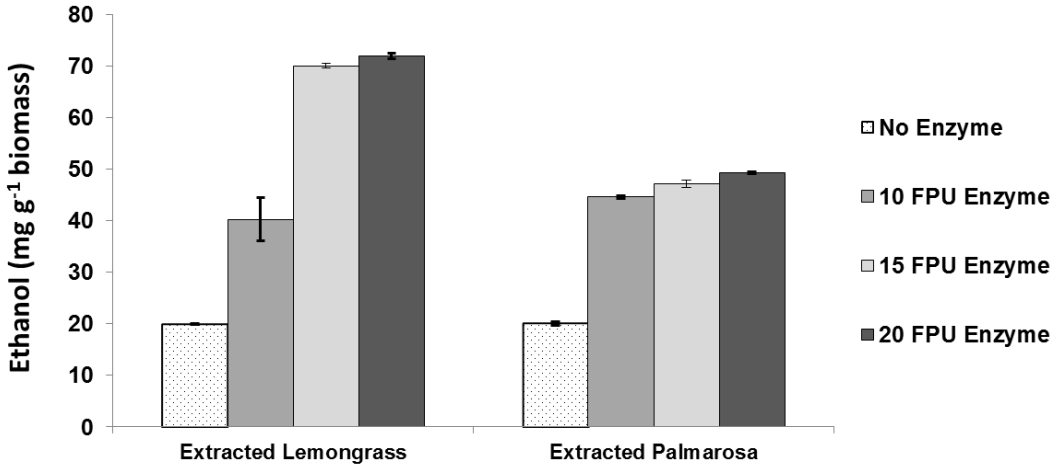

Figure 26. Bench top simultaneous saccharification and fermentation (SSF) ethanol yield from extracted whole biomass based on filter paper units of enzymes.

Fermentation of lemongrass biomass reached maximum yields at enzyme concentrations of 15 FPU g^{-1} biomass. Fermentation of palmarosa biomass reached maximum yields at an enzyme concentration of 10 FPU g^{-1} biomass.

Figure 27. Final ethanol yield from lemongrass biomass (mg g^{-1}) and palmarosa (mg g^{-1}) that was (EX) or was not (NE) previously extracted for essential oils in comparison to two lots of BioEnergy Science Center (BESC) control switchgrass.

A) Final ethanol concentration of biomass that was not pretreated in fermentation liquids. B) Final ethanol yield (mg g^{-1} biomass) of dilute acid pretreated lemongrass and palmarosa biomass that was (EX) or was not (NE) extracted for essential oils in comparison to lot #1 BESC control switchgrass.

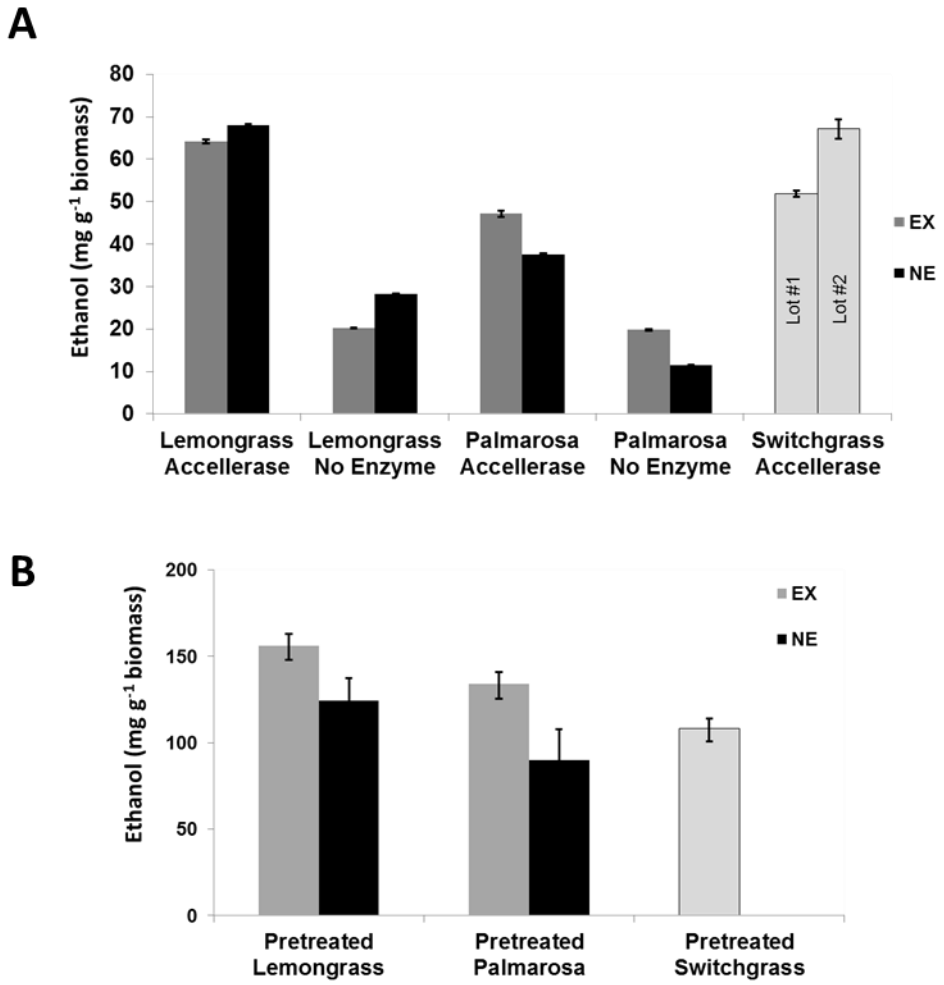
A
Ethanol (mg g-1 biomass)
80
70
60
50
40
30
20
10
0
EX
NE
Lot #1
Lot #2
Lemongrass Accellerase
Lemongrass No Enzyme
Palmarosa Accellerase
Palmarosa No Enzyme
Switchgrass Accellerase
B
Ethanol (mg g-1 biomass)
200
150
100
50
0
EX
NE
Pretreated Lemongrass
Pretreated Palmarosa
Pretreated Switchgrass

Figure 27.

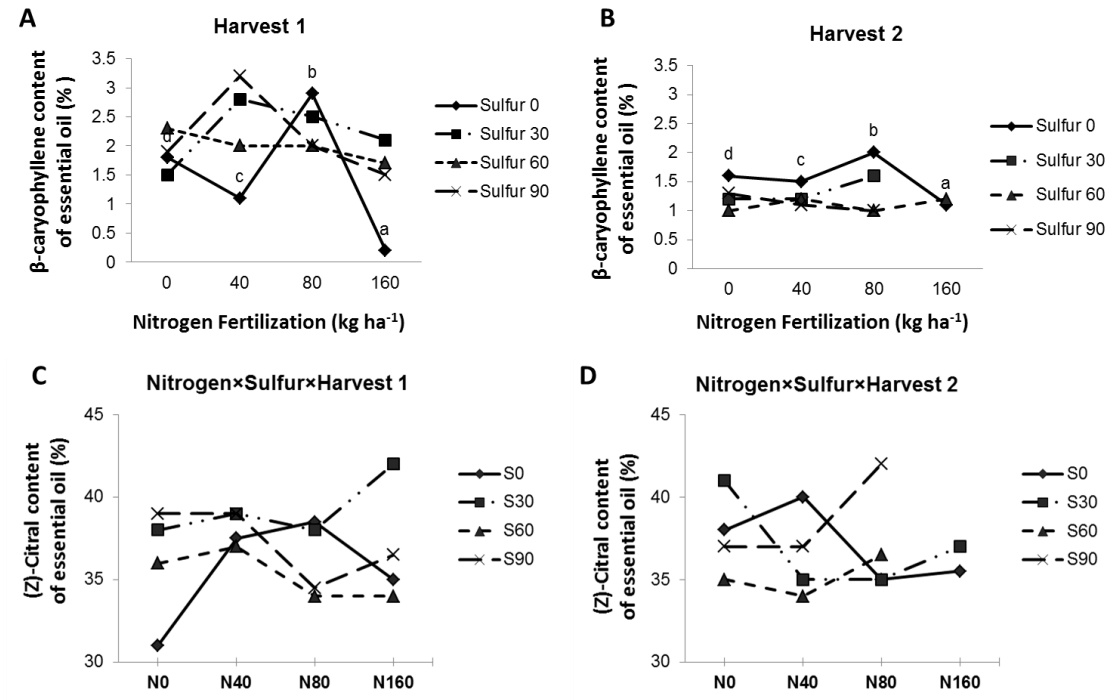

Figure 28. Interaction effect of nitrogen×sulfur×harvest on β-caryophyllene and (*Z*)-citral content in lemongrass essential oil.

CONCLUSION

Drop-in ready biofuels need several characteristics if direct production is to be feasible in rural areas. Firstly, the biofuel should yield supplies of liquid fuels frequently, preferably yearly or several times a year to be sustainable. In this work, *Pittosporum resiniferum*, *Cymbopogon flexuosus,* and *Cymbopogon martinii* almost meet this requirement. *P. resiniferum* fruits twice a year in its native ranges while the *Cymbopogon* spp. can be harvested in much the same way as switchgrass. Current and proposed biofuels meet this demand with feedstocks such as oilseeds, corn, switchgrass, and algae producing at least one harvest in a year with some of these proposing several harvests a year.

Secondly, the biofuels should require minimal infrastructure to extract, process, and deliver finished liquid fuel products to consumers. As presented in Chapter III, *P. resiniferum*, *Copaifera reticulata*, and *Cymbopogon flexuosus* and *C. martinii* all produce extractable oils that are usable directly as biofuels in at least B20 blends. Specifically, *P. resiniferum* oil can be harvested simply by squeezing and filtering the fruit produced on the tree. The trunks of *Copaifera* species can be tapped and used directly, however steam distillation to remove high molecular weight resin acids did improve fuel characteristics. *C. flexuosus* and *C. martinii* require distillation of biomass to recover oils and as such could prove too complex to be utilized in rural settings. Likewise, current and proposed biofuel feedstocks do not meet this characteristic. Current biofuels from either

fermentation or thermochemical conversion require biorefineries and infrastructure investment to develop into finalized transportation feedstocks. Current biodiesel feedstocks also require chemical conversion and thus must be alkyl esterified and then removed from the waste glycerol.

Lastly, the biofuel should also mix with existing supplies of petroleum fuels without restriction (are completely fungible with petroleum fuels) to meet fluctuations in supply, demand, and therefore cost, of petroleum fuels. *P. resiniferum*, *Copaifera reticulata*, *C. flexuosus*, *C. martinii*, and *D. albus* all can be blended to at least B20 according to this work. However, further road testing in engines will be required to determine long term effects on engines, if any.

P. resiniferum oil meets each of these characteristics and could therefore be useful for production of fuels in rural areas. Additionally, investigating the novel short-alkane biosynthesis pathway in *P. resiniferum* could yield biosynthetic genes useful in creating biochemicals identical to petroleum fuels. Oil from *Copaifera reticulata* and from *C. flexuosus* and *C. martinii* meet several of these characteristics. However, both of these species do not have the required production characteristics. Specifically, *Copaifera* oils are not produced annually and require more than 20 years to mature before harvest. *Cymbopogon* oils require distillation and as such are unlikely to be feasible in remote rural areas as maintaining distillation apparatus would be expensive and require specialty knowledge. However, these three species warrant further study for other

applications. *Copaifera spp.* produce large volumes of sesquiterpenes which is unique compared to other known plants. *Cymbopogon flexuosus* and *C. martinii* could be used as dual-use feedstocks to produce high value essential oils alongside lignocellulosic ethanol.

Copaifera officinalis and *C. langsdorffii* oils, like *C. reticulata*, are composed primarily of sesquiterpenoids. Illumina sequencing of the *C. officinalis* transcriptome allowed for isolation and identification of the terpene synthase genes involved in producing all the major sesquiterpene present in *C. officinalis* tissues. Additionally, upstream terpene biosynthesis genes were also identified. Primers specific to *C. officinalis* sesquiterpene synthases were also able to isolate *C. langsdorffii* sesquiterpene synthases suggesting that sequencing one species in a genus can enable investigation of biosynthetic pathway genes across a genus. Further investigation may lead to understanding how the *Copaifera* genus produces copious volumes of sesquiterpene oils.

Cymbopogon flexuosus and *C. martinii* biomass and ethanol fermentation studies were conducted to determine the feasibility of producing terpenoid advanced fuels in lignocellulosic ethanol feedstocks. *C. flexuosus* and *C. martinii* had a lower biomass yield ha^{-1} to previous switchgrass field plots which corresponded to a lower ethanol yield ha^{-1}. However, *C. flexuosus* and *C. martinii* also produce essential oils that can either be used as advanced biofuels or high-value coproducts. Ultimately, the combined value of the essential oils and

ethanol produced from *C. flexuosus* and *C. martinii* was greater than from switchgrass ethanol production alone. Therefore, dual use feedstocks and addition of value added traits to switchgrass should be considered in future biofuel production plans. Inhibition of fermentation was not directly observed in either *C. flexuosus* or *C. martinii* samples though the essential oils from both species was previously reported to be toxic to *Saccharomyces cerevisiae* used during simultaneous saccharification and fermentation bench-scale studies.

LIST OF REFERENCES

The 1KP Project. [www.onekp.com]
Phytozome. [www.phytozome.net]
Plant Metabolic Network Database. [www.plantcyc.org]
SIG-Pred. [www.signalpeptide.de]
Agarwal, A.K. 2007. Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. *Progress in Energy and Combustion Science*, **33**(3), 233-271.
Ahmad, R., Parfitt, D., Fass, J., Ogundiwin, E., Dhingra, A., Gradziel, T., Lin, D., Joshi, N., Martinez-Garcia, P., Crisosto, C. 2011. Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics*, **12**(1), 569.
Ahring, B., Westermann, P. 2007. Coproduction of bioethanol with other biofuels. in: *Biofuels*, (Ed.) L. Olsson, Vol. 108, Springer Berlin Heidelberg, pp. 289-302.
Alencar, J.C. 1982. Estudos silviculturais de uma populacao natural de *Copaifera multijuga* Hayne (Leguminosae) na Amazonia Central IV. Interpretacao de dados fenologicos em relacao a elementos climaticos. *Acta Amazonica*, **18**, 109-209.
Allison, G.G., Robbins, M.P., Carli, J., Clifton-Brown, J.C., Donnison, I.S. 2010. Designing biomass crops with improved calorific content and attributes for burning: a UK Perspective. in: *Plant Biotechnology for Sustainable Production of Energy and Co-products*, (Eds.) P.N. Mascia, J. Scheffran, J.M. Widholm, Vol. 66, Springer Berlin Heidelberg, pp. 25-55.
Allsopp, C., Fattouh, B. 2011. Oil and international energy. *Oxford Review of Economic Policy*, **27**(1), 1-32.
Anderson, A.B., Riffer, R., Wong, A. 1969. Monoterpenes, fatty and resin acids of *Pinus ponderosa* and *Pinus jeffreyi*. *Phytochemistry*, **8**(5), 873-875.
Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., Flynn, J., Matyszczuk, P., Andryszak, K., Laurelli, M., Golovkin, M., Koprowski, H. 2010. Tobacco as a production platform for biofuel: overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnology Journal*, **8**(3), 277-287.
Ansorge, W.J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, **25**(4), 195-203.
Arcoumanis, C., Bae, C., Crookes, R., Kinoshita, E. 2008. The potential of di-methyl ether (DME) as an alternative fuel for compression-ignition engines: A review. *Fuel*, **87**(7), 1014-1030.

ASTM Standard D6751. 2012. Standard specification for biodiesel fuel blend stock (B100) for middle distillate fuels. ASTM International, West Conshohocken, PA. DOI: 10.1520/D6751-12. www.astm.org.

Bacha, J., Freel, J., Gibbs, A., Gibbs, L., Hemighaus, G., Hoekman, K., Horn, J., Ingham, M., Jossens, L., Kohler, D., Lesnini, D., McGeehan, J., Nikanjam, M., Olsen, E., Organ, R., Scott, B., Sztenderowicz, M., Tiedemann, A., Walker, C., Lind, J., Jones, J., Scott, D., Mills, J. 2007. Diesel fuels technical review, Chevron Corporation.

Bailey, J., Lee, H. 2012. North American oil and gas reserves: prospects and policy.

Baker, D.A., Baker, F.S., Gallego, N.C. 2009. Thermal engineering of ligning for low-cost production of carbon fiber. in: *The Fiber Society 2009 Fall Meeting and Technical Conference*. Athens, GA, USA.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., Ashton, N.W., Axtell, M.J., Barker, E., Barker, M.S., Bennetzen, J.L., Bonawitz, N.D., Chapple, C., Cheng, C., Correa, L.G.G., Dacre, M., DeBarry, J., Dreyer, I., Elias, M., Engstrom, E.M., Estelle, M., Feng, L., Finet, C., Floyd, S.K., Frommer, W.B., Fujita, T., Gramzow, L., Gutensohn, M., Harholt, J., Hattori, M., Heyl, A., Hirai, T., Hiwatashi, Y., Ishikawa, M., Iwata, M., Karol, K.G., Koehler, B., Kolukisaoglu, U., Kubo, M., Kurata, T., Lalonde, S., Li, K., Li, Y., Litt, A., Lyons, E., Manning, G., Maruyama, T., Michael, T.P., Mikami, K., Miyazaki, S., Morinaga, S.i., Murata, T., Mueller-Roeber, B., Nelson, D.R., Obara, M., Oguri, Y., Olmstead, R.G., Onodera, N., Petersen, B.L., Pils, B., Prigge, M., Rensing, S.A., Riaño-Pachón, D.M., Roberts, A.W., Sato, Y., Scheller, H.V., Schulz, B., Schulz, C., Shakirov, E.V., Shibagaki, N., Shinohara, N., Shippen, D.E., Sørensen, I., Sotooka, R., Sugimoto, N., Sugita, M., Sumikawa, N., Tanurdzic, M., Theißen, G., Ulvskov, P., Wakazuki, S., Weng, J.K., Willats, W.W.G.T., Wipf, D., Wolf, P.G., Yang, L., Zimmer, A.D., Zhu, Q., Mitros, T., Hellsten, U., Loqué, D., Otillar, R., Salamov, A., Schmutz, J., Shapiro, H., Lindquist, E., et al. 2011. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*, **332**(6032), 960-963.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., Miyano, S. 2002. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*, **18**(2), 298-305.

Benderoth, M., Textor, S., Windsor, A.J., Mitchell-Olds, T., Gershenzon, J., Kroymann, J. 2006. Positive selection driving diversification in plant secondary metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(24), 9118-9123.

Benjumea, P., Agudelo, J.R., Agudelo, A.S.F. 2010. Effect of the degree of unsaturation of biodiesel fuels on engine performance, combustion characteristics, and emissions. *Energy and Fuels*, **25**(1), 77-85.

Bergmüller, E., Porfirova, S., Dörmann, P. 2003. Characterization of an *Arabidopsis* mutant deficient in γ-tocopherol methyltransferase. *Plant Molecular Biology*, **52**(6), 1181-1190.

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., Legrand, M. 2007. Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell*, **19**(1), 148-162.

Bhasin, M., Raghava, G.P.S. 2004. ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. *Nucleic Acids Research*, **32**, W414-W419.

Boe, A. 2007. Variation between two switchgrass cultivars for components of vegetative and seed biomass. *Crop Science*, **47**(2), 636-640.

Bond, J.Q., Alonso, D.M., Wang, D., West, R.M., Dumesic, J.A. 2010. Integrated catalytic conversion of γ-valerolactone to liquid alkenes for transportation fuels. *Science*, **327**(5969), 1110-1114.

Bozell, J.J. 2008. Feedstocks for the future - Biorefinery production of chemicals from renewable carbon. *Clean-Soil Air Water*, **36**(8), 641-647.

Bozell, J.J., Black, S.K., Myers, M., Cahill, D., Miller, W.P., Park, S. 2011. Solvent fractionation of renewable woody feedstocks: organosolv generation of biorefinery process streams for the production of biobased chemicals. *Biomass and Bioenergy*, **35**(10), 4197-4208.

Braga, W.F., Rezende, C.M., Antunes, O.A.C., Pinto, A.C. 1998. Terpenoids from *Copaiba cearensis*. *Phytochemistry*, **49**(1), 263-264.

Bunting, B., Bunce, M., Joyce, B., Crawford, R. 2011. Investigation and optimization of biodiesel chemistry for HCCI combustion. in: *Sustainable Automotive Technologies 2011*, (Eds.) S. Hung, A. Subic, J. Wellnitz, Springer Berlin Heidelberg, pp. 51-58.

Bunting, B.G., Eaton, S.J., Crawford, R.W. 2009. Performance evaluation and optimization of diesel fuel properties and chemistry in an HCCI engine. *SAE Technical Paper*, 01-2645.

Calvin, M. 1980. Hydrocarbons from plants - analytical methods and observations. *Naturwissenschaften*, **67**(11), 525-533.

Calvin, M. 1981. Growing and engineering of hydrocarbons. *In Vitro-Journal of the Tissue Culture Association*, **17**(3), 215-215.

Calvin, M. 1983. New sources for fuel and materials. *Science*, **219**(4580), 24-26.

Camilli, R., Reddy, C.M., Yoerger, D.R., Van Mooy, B.A.S., Jakuba, M.V., Kinsey, J.C., McIntyre, C.P., Sylva, S.P., Maloney, J.V. 2010. Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon. *Science*, **330**(6001), 201-204.

Cao, R., Zhang, Y., Mann, F.M., Huang, C., Mukkamala, D., Hudock, M.P., Mead, M.E., Prisic, S., Wang, K., Lin, F.-Y., Chang, T.-K., Peters, R.J., Oldfield, E. 2010. Diterpene cyclases and the nature of the isoprene fold. *Proteins: Structure, Function, and Bioinformatics*, **78**(11), 2417-2432.

Caretto, S., Quarta, A., Durante, M., Nisi, R., De Paolis, A., Blando, F., Mita, G. 2011. Methyl jasmonate and miconazole differently affect arteminisin production and gene expression in *Artemisia annua* suspension cultures. *Plant Biology*, **13**(1), 51-58.

Carlsson, A.S. 2009. Plant oils as feedstock alternatives to petroleum - A short survey of potential oil crop platforms. *Biochimie*, **91**(6), 665-670.

Carroll, A., Somerville, C. 2009. Cellulosic biofuels. *Annual Review of Plant Biology*, **60**(1), 165-182.

Cascon, V., Gilbert, B. 2000. Characterization of the chemical composition of oleoresins of *Copaifera guianensis* Desf., *Copaifera duckei* Dwyer and *Copaifera multijuga* Hayne. *Phytochemistry*, **55**(7), 773-778.

Cenci, A., Combes, M.-C., Lashermes, P. 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology*, **78**(1), 135-145.

Chen, F., Al-Ahmad, H., Joyce, B., Zhao, N., Kollner, T.G., Degenhardt, J., Stewart, C.N. 2009. Within-plant distribution and emission of sesquiterpenes from *Copaifera officinalis*. *Plant Physiology and Biochemistry*, **47**(11-12), 1017-1023.

Chen, F., Dixon, R.A. 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology*, **25**(7), 759-761.

Chen, F., Tholl, D., Bohlmann, J., Pichersky, E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, **66**(1), 212-229.

Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E., Gershenzon, J. 2003. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell*, **15**(2), 481-94.

Cheng, A.X., Lou, Y.G., Mao, Y.B., Lu, S., Wang, L.J., Chen, X.Y. 2007. Plant terpenoids: biosynthesis and ecological functions. *Journal of Integrative Plant Biology*, **49**(2), 179-186.

Cherubini, F., Jungmeier, G. 2010. LCA of a biorefinery concept producing bioethanol, bioenergy, and chemicals from switchgrass. *The International Journal of Life Cycle Assessment*, **15**(1), 53-66.

Chow, J., Kopp, R.J., Portney, P.R. 2003. Energy resources and global development. *Science*, **302**(5650), 1528-1531.

Chowdhury, J.U., Yusuf, M., Begum, J., Mondello, L., Previti, P., Dugo, G. 1998. Studies on the essential oil bearing plants of Bangladesh. Part IV. composition of the leaf oil of three *Cymbopogon* species: *C. flexuosus*

(Nees ex Steud.) Wats., *C. nardus* (L.) Rendle var. confertiflorus (Steud.) N. L. Bor and *C. martinii* (Roxb.) Wats. var. martinii. *Journal of Essential Oil Research*, **10**(3), 301-306.
Christian, D.P., Lederle, P.E. 1984. Seed properties and water balance in desert granivores. *The Southwestern Naturalist*, **29**(2), 181-188.
da Costa, M.M., Cohen, C., Schaeffer, R. 2007. Social features of energy production and use in Brazil: Goals for a sustainable energy future. *Natural Resources Forum*, **31**(1), 11-20.
Dagaut, P., Cathonnet, M. 2006. The ignition, oxidation, and combustion of kerosene: A review of experimental and kinetic modeling. *Progress in Energy and Combustion Science*, **32**(1), 48-92.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, **27**(8), 1164-1165.
Dassanayake, M., Oh, D.-H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.-J., Bressan, R.A., Zhu, J.-K., Bohnert, H.J., Cheeseman, J.M. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics*, **43**(9), 913-918.
de Almeida, P., Silva, P.D. 2009. The peak of oil production—timings and market recognition. *Energy Policy*, **37**(4), 1267-1276.
Decker, S., Brunecky, R., Tucker, M., Himmel, M., Selig, M. 2009. High-throughput screening techniques for biomass conversion. *BioEnergy Research*, **2**(4), 179-192.
Degenhardt, J., Köllner, T.G., Gershenzon, J. 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, **70**(15), 1621-1637.
Demirbas, A. 2009a. Emission characteristics of gasohol and diesohol. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, **31**(13), 1099 - 1104.
Demirbas, A. 2009b. Progress and recent trends in biodiesel fuels. *Energy Conversion and Management*, **50**(1), 14-34.
Demirbas, M.F. 2009c. Biorefineries for biofuel upgrading: A critical review. *Applied Energy*, **86**, S151-S161.
Di Cesare, S., Binder, U., Maier, T., Skerra, A. 2013. High-yield production of PASylated human growth hormone using secretory *E. coli* technology. *BioProcess International*, **11**, 4.
Du, W., Li, W., Sun, T., Chen, X., Liu, D. 2008. Perspectives for biotechnological production of biodiesel and impacts. *Applied Microbiology and Biotechnology*, **79**(3), 331-337.
Dupain, X., Gamas, E.D., Madon, R., Kelkar, C.P., Makkee, M., Moulijn, J.A. 2003. Aromatic gas oil cracking under realistic FCC conditions in a microriser reactor. *Fuel*, **82**(13), 1559-1569.

Dutt, D., Garg, A.P., Tyagi, C.H., Upadhyay, A.K., Upadhyaya, J.S. 2007. Bio-soda-ethanol-water (Bio-sew) delignification of lignocellulosic residues of *Cymbopogon martinii* with *Phanerochaete chrysosporium*. *Cellulose Chemistry and Technology*, **41**(2-3), 161-174.

Dwyer, J.D. 1951. The Central American, West Indian, and South American species of *Copaifera* (Caesalpiniaceae). *Brittonia*, **7**, 143-172.

Dwyer, J.D. 1954. Further studies on the New World species of *Copaifera*. *Bulletin of the Torrey Botanical Club*, **81**, 179-187.

Egli, D. 2008. Comparison of corn and soybean yields in the United States: historical trends and future prospects. *Agronomy Journal*, **100**(3), 79.

Elgendy, E.M., Khayyat S.A. 2008. Oxidation reactions of some natural volatile aromatic compounds: anethole and eugenol. Russian Journal of Organic Chemistry, **44**(6), 823-829.

Energy Information Administration. 2011. Annual energy review 2011, (Ed.) U.S.Department of Energy. Washington, DC.

Energy Information Administration. 2011. Oil market report. in: *Total OECD 12 Month Moving Average Demand vs Year/Year growth*, Organisation for Economic Co-production and Development. Paris, France.

Escamilla-Treviño, L.L., Shen, H., Uppalapati, S.R., Ray, T., Tang, Y., Hernandez, T., Yin, Y., Xu, Y., Dixon, R.A. 2010. Switchgrass (*Panicum virgatum*) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. *New Phytologist*, **185**(1), 143-155.

Evans, P.H., Becerra, J.X. 2006. Non-terpenoid essential oils from *Bursera chemapodicta*. *Flavour and Fragrance Journal*, **21**(4), 616-618.

FitzPatrick, M., Champagne, P., Cunningham, M.F., Whitney, R.A. 2010. A biorefinery processing perspective: treatment of lignocellulosic materials for the production of value-added products. *Bioresource Technology*, **101**(23), 8915-8922.

Fleisher, A., Fleisher, Z. 2004. Study of *Dictamnus gymnostylis* volatiles and plausible explanation of the "burning bush" phenomenon. *Journal of Essential Oil Research*, **16**, 1-3.

Freire, L.M.S., Filho, J.R.C., Moura, C.V.R., Soledade, L.E.B., Stragevitch, L., Cordeiro, Â.M.T.M., Santos, I.M.G., Souza, A.G. 2012. Evaluation of the oxidation stability and flow properties of quaternary mixtures of vegetable oils for biodiesel production. *Fuel*, **95**(0), 126-130.

Fu, C., Mielenz, J.R., Xiao, X., Ge, Y., Hamilton, C.Y., Rodriguez, M., Chen, F., Foston, M., Ragauskas, A., Bouton, J., Dixon, R.A., Wang, Z.-Y. 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(9), 3803-3808.

Fung, P.K., Krushkal, J., Weathers, P.J. 2010. Computational analysis of the evolution of 1-Deoxy-D-xylulose-5-phosphate reductoisomerase, an

important enzyme in plant terpene biosynthesis. *Chemistry & Biodiversity*, **7**(5), 1098-1110.
Gallego-Giraldo, L., Jikumaru, Y., Kamiya, Y., Tang, Y., Dixon, R.A. 2011. Selective lignin downregulation leads to constitutive defense response expression in alfalfa (Medicago sativa L.). *New Phytologist*, **190**(3), 627-639.
Ganjewala, D., Luthra, R. 2010. Essential oil biosynthesis and regulation in the genus *Cymbopogon*. *Natural Product Communications*, **5**(1), 163.
García, E., Laca, M., Pérez, E., Garrido, A., Peinado, J.n. 2008. New class of acetal derived from glycerin as a biodiesel fuel component. *Energy and Fuels*, **22**(6), 4274-4280.
Garg, R., Patel, R.K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., Bhatia, S., Chattopadhyay, D., Tyagi, A.K., Jain, M. 2011. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiology*, **156**(4), 1661-1678.
Gershenzon, J., Croteau, R. 1991. Terpenoids. in: *Herbivores: Their Interactions with Secondary Plant Metabolites*, (Eds.) G.A. Rosenthal, M.R. Berenbaum, Vol. Vol. 1 The Chemical Participants, Academic Press. New York, pp. 165-219.
Godfray, H.C.J., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Nisbett, N., Pretty, J., Robinson, S., Toulmin, C., Whiteley, R. 2010. The future of the global food system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**(1554), 2769-2777.
Gomes, N.D., Rezende, C.D., Fontes, S.P., Hovell, A.M.C., Landgraf, R.G., Matheus, M.E., Pinto, A.D., Fernandes, P.D. 2008. Antineoplasic activity of *Copaifera multijuga* oil and fractions against ascitic and solid Ehrlich tumor. *Journal of Ethnopharmacology*, **119**(1), 179-184.
Gomes, N.M., Rezende, C.M., Fontes, S.P., Matheus, M.E., Fernandes, P.D. 2007. Antinociceptive activity of Amazonian copaiba oils. *Journal of Ethnopharmacology*, **109**(3), 486-492.
Gonela, A., Sebbenn, A.M., Soriani, H.H., Mestriner, M.A., Martinez, C.A., Alzate-Marin, A.L. 2013. Genetic diversity and mating system of *Copaifera langsdorffii* (Leguminosae/Caesalpinioideae). *Genetics and Molecular Research*, **12**(1), 569-580.
Gramosa, M.V., Silveira, E.R. 2005. Volatile constituents of *Copalifera langsdorfffi* from the Brazilian northeast. *Journal of Essential Oil Research*, **17**(2), 130-132.
Gressel, J. 2008. Transgenics are imperative for biofuel crops. *Plant Science*, **174**(3), 246-263.
Hagvall, L., Bäcktorp, C., Svensson, S., Nyman, G., Börje, A., Karlberg, A. T. 2007. Fragrance compound geraniol forms contact allergens on air

exposure. Identification and quantification of oxidation products and effect on skin sensitization. *Chemical research in toxicology*, **20**(5), 807-814.
Hartmann, T. 2007. From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, **68**(22-24), 2831-2846.
Harvey, B.G., Wright, M.E., Quintana, R.L. 2009. High-density renewable fuels based on the selective dimerization of pinenes. *Energy and Fuels*, **24**(1), 267-273.
Helal, G.A., Sarhan, M.M., Shahla, A.N.K.A., El-Khair, E.K.A. 2006. Effect of *Cymbopogon citratus* L. essential oil on growth and morphogenesis of *Saccharomyces cerevisiae* ML2-strain. *Journal of Basic Microbiology*, **46**(5), 375-386.
Hermann, B., Patel, M. 2007. Today's and tomorrow's bio-based bulk chemicals from white biotechnology. *Applied Biochemistry And Biotechnology*, **136**(3), 361-388.
Hillwig, M.L., Xu, M., Toyomasu, T., Tiernan, M.S., Wei, G., Cui, G., Huang, L., Peters, R.J. 2011. Domain loss has independently occurred multiple times in plant terpene synthase evolution. *The Plant Journal*, **68**(6), 1051-1060.
Hodges, J.D., Elam, W.W., Watson, W.F. 1977. Physical properties of the oleoresin system of the four major southern pines. *Canadian Journal of Forest Research*, **7**(3), 520-525.
Hoekman, S.K. 2009. Biofuels in the U.S. - challenges and opportunities. *Renewable Energy*, **34**(1), 14-22.
Höglund, A., Dönnes, P., Blum, T., Adolph, H.W., Kohlbacher, O. 2006. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics*, **22**(10), 1158-1165.
Hou, D. 1994. Studies in Malesian Caesalpinioideae (Leguminosae). I. The genera *Acrocarpus, Afzelia, Copaifera*, and *Intsia*. *Blumea*, **38**, 313-330.
Howard, J.J., Cazin, J., Wiemer, D.F. 1988. Toxicity of terpenoid deterrents to the leafcutting ant *Atta cephalotes* and its mutualistic fungus. *Journal of Chemical Ecology*, **14**(1), 59-69.
Hyten, D., Song, Q., Fickus, E., Quigley, C., Lim, J.-S., Choi, I.-Y., Hwang, E.-Y., Pastor-Corrales, M., Cregan, P. 2010. High-throughput SNP discovery and assay development in common bean. *BMC Genomics*, **11**(1), 475.
John, A.J., George, V., Pradeep, N.S., Sethuraman, M.G. 2008. Composition and antibacterial activity of the leaf and fruit oils of *Pittosporum neelgherrense* Wight et Arn. *Journal of Essential Oil Research*, **20**(4), 380-382.
Joyce, B.L., Stewart Jr, C.N. 2012. Designing the perfect plant feedstock for biofuel production: Using the whole buffalo to diversify fuels and products. *Biotechnology Advances*, **30**(5), 1011-1022.
Karthikeyan, R., Nallusamy, N., Alagumoorthi, N., Ilangovan, V. 2010. Optimisation of engine operating parameters for turpentine mixed diesel

fueled DI diesel engine using Taguchi Method. *International Journal of Engineering Science*, **2**.
Katoh, K., Misawa, K., Kuma, K.i., Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**(14), 3059-3066.
Kausch, A.P., Hague, J., Oliver, M., Li, Y., Daniell, H., Mascia, P., Watrud, L.S., Stewart, C.N. 2010. Transgenic perennial biofuel feedstocks and strategies for bioconfinement. *Biofuels*, **1**(1), 163-176.
Keeling, C., Bohlmann, J. 2006a. Diterpene resin acids in conifers. *Phytochemistry*, **67**, 2415 - 2423.
Keeling, C., Weisshaar, S., Ralph, S., Jancsik, S., Hamberger, B., Dullat, H., Bohlmann, J. 2011. Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). *BMC Plant Biology*, **11**(1), 43.
Keeling, C.I., Bohlmann, J. 2006b. Diterpene resin acids in conifers. *Phytochemistry*, **67**(22), 2415-2423.
Kerr, R.A. 2012. Technology is turning U.S. oil around but not the world's. *Science*, **335**(6068), 522-523.
Kim, M.Y., Lee, S., Van, K., Kim, T.-H., Jeong, S.-C., Choi, I.-Y., Kim, D.-S., Lee, Y.-S., Park, D., Ma, J., Kim, W.-Y., Kim, B.-C., Park, S., Lee, K.-A., Kim, D.H., Kim, K.H., Shin, J.H., Jang, Y.E., Kim, K.D., Liu, W.X., Chaisan, T., Kang, Y.J., Lee, Y.-H., Kim, K.-H., Moon, J.-K., Schmutz, J., Jackson, S.A., Bhak, J., Lee, S.-H. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(51), 22032-22037.
Kirby, J., Keasling, J.D. 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual Review of Plant Biology*, **60**(1), 335-355.
Kleinert, M., Barth, T. 2008a. Phenols from lignin. *Chemical Engineering & Technology*, **31**(5), 736-745.
Kleinert, M., Barth, T. 2008b. Towards a lignincellulosic biorefinery: direct one-step conversion of lignin to hydrogen-enriched biofuel. *Energy & Fuels*, **22**(2), 1371-1379.
Knothe, G. 2009. Improving biodiesel fuel properties by modifying fatty ester composition. *Energy & Environmental Science*, **2**(7), 759-766.
Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A., Sugi, M. 2010. Tropical cyclones and climate change. *Nature Geoscience*, **3**(3), 157-163.
Kohse-Höinghaus, K., Oßwald, P., Cool, T.A., Kasper, T., Hansen, N., Qi, F., Westbrook, C.K., Westmoreland, P.R. 2010. Biofuel combustion chemistry: from ethanol to biodiesel. *Angewandte Chemie International Edition*, **49**(21), 3572-3597.

Köllner, T.G., O'Maille, P.E., Gatto, N., Boland, W., Gershenzon, J., Degenhardt, J. 2006. Two pockets in the active site of maize sesquiterpene synthase TPS4 carry out sequential parts of the reaction scheme resulting in multiple products. *Archives of Biochemistry and Biophysics*, **448**(1-2), 83-92.

Kudapa, H., Bharti, A.K., Cannon, S.B., Farmer, A.D., Mulaosmanovic, B., Kramer, R., Bohra, A., Weeks, N.T., Crow, J.A., Tuteja, R., Shah, T., Dutta, S., Gupta, D.K., Singh, A., Gaikwad, K., Sharma, T.R., May, G.D., Singh, N.K., Varshney, R.K. 2012. A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. *Molecular Plant*, **5**(5), 1020-1028.

Kulkarni, R.N. 1994. Phenotypic recurrent selection for oil content in East Indian lemongrass. *Euphytica*, **78**(1), 103-107.

Kulkarni, R.N., Mallavarapu, G.R., Ramesh, S. 1992. The oil content and composition of new variants of *Cymbopogon flexuosus*. *Journal of Essential Oil Research*, **4**(5), 511-514.

Lambin, E.F., Meyfroidt, P. 2011. Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(9), 3465-3472.

Lange, B.M., Mahmoud, S.S., Wildung, M.R., Turner, G.W., Davis, E.M., Lange, I., Baker, R.C., Boydston, R.A., Croteau, R.B. 2011. Improving peppermint essential oil yield and composition by metabolic engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(41), 16944-16949.

Lange, J.-P., Price, R., Ayoub, P.M., Louis, J., Petrus, L., Clarke, L., Gosselink, H. 2010. Valeric biofuels:a platform of cellulosic transportation fuels. *Angewandte Chemie International Edition*, **49**(26), 4479-4483.

Langenheim Jean, H. 1996. Biology of amber-producing trees: focus on case studies of *Hymenaea* and *Agathis*. in: *Amber, Resinite, and Fossil Resins*, Vol. 617, American Chemical Society, pp. 1-31.

Langenheim, J.H. 1973. Leguminous resin-producing trees in Africa and South America. in: *Tropical Forest Ecosystems in Africa and South America*, (Eds.) B.J. Meggers, E.S. Ayensu, W.D. Duckworth, Smithsonian Institution Press. Washington, DC, pp. 89-104.

Langenheim, J.H. 2003. *Plant resins: chemistry, evolution, ecology, and ethnobotany.* Timber Press, Portland.

Lapuerta, M., Armas, O., Rodríguez-Fernández, J. 2008. Effect of biodiesel fuels on diesel engine emissions. *Progress in Energy and Combustion Science*, **34**(2), 198-223.

Laza, T., Bereczky, Á. 2011. Basic fuel properties of rapeseed oil-higher alcohols blends. *Fuel*, **90**(2), 803-810.

Lee, S., Chappell, J. 2008. Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. *Plant Physiology*, **147**(3), 1017-1033.
Lee, S.K., Chou, H., Ham, T.S., Lee, T.S., Keasling, J.D. 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, **19**(6), 556-563.
Léonard, J. 1949. Notulae systematicae IV (Caesalpiniaceae-Amherstieae africanae americanaeque). *Bulletin du Jardin botanique de l'État a Bruxelles*, **19**, 384-408.
Léonard, J. 1950. Notulae systematicae IX nouvelles observations sur le genre Guibourtia (Caesalpiniaceae). *Bulletin du Jardin botanique de l'État a Bruxelles*, **20**, 270-284.
Lewinsohn, E., Gijzen, M., Savage, T.J., Croteau, R. 1991. Defense mechanisms of conifers - relationship of monoterpene cyclase activity to anatomical specialization and oleoresin monoterpene content. *Plant Physiology*, **96**(1), 38-43.
Lewinsohn, E., Savage, T.J., Gijzen, M., Croteau, R. 1993. Simultaneous analysis of monoterpenes and diterpenoids of confier oleoresin. *Phytochemical Analysis*, **4**(5), 220-225.
Lichtenthaler, H.K. 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **50**, 47-65.
Lima, N.N., Pio, N.S. 2007. Evaluation of the glue amount in the static bending of the species *Copaifera duckei* Dwyer and *Eperua oleifera* Ducke. *Acta Amazonica*, **37**, 347-352.
Lima, S.R.M., Veiga, V.F., Christo, H.B., Pinto, A.C., Fernandes, P.D. 2003. *In vivo* and *in vitro* studies on the anticancer activity of *Copaifera multijuga* Hayne and its fractions. *Phytotherapy Research*, **17**(9), 1048-1053.
Liu, T., Khosla, C. 2010. Genetic engineering of *Escherichia coli* for biofuel production. *Annual Review of Genetics*, **44**(1), 53-69.
Lv, P., Cheng, Y., Yang, L., Yuan, Z., Li, H., Luo, W. 2013. Improving the low temperature flow properties of palm oil biodiesel: addition of cold flow improver. *Fuel Processing Technology*, **110**(0), 61-64.
Lyko, H., Deerberg, G., Weidner, E. 2009. Coupled production in biorefineries-combined use of biomass as a source of energy, fuels and materials. *Journal of Biotechnology*, **142**(1), 78-86.
Macedo, C.A., Langenheim, J.H. 1989a. A further investigation of leaf sesquiterpene variation in relation to herbivory in 2 Brazilian populations of *Copaifera langsdorffii*. *Biochemical Systematics and Ecology*, **17**(3), 207-216.
Macedo, C.A., Langenheim, J.H. 1989b. Intraplant and interplant leaf sesquiterpene variability in *Copaifera langsdorffii* - relation to

microlepidopteran herbivory. *Biochemical Systematics and Ecology*, **17**(7-8), 551-557.
Macedo, C.A., Langenheim, J.H. 1989c. Microlepidopteran herbivory in relation to leaf sesquiterpenes in *Copaifera langsdorffii* adult trees and their seedling progeny in a Brazilian woodland. *Biochemical Systematics and Ecology*, **17**(3), 217-224.
Malik, S., Cusidó, R.M., Mirjalili, M.H., Moyano, E., Palazón, J., Bonfill, M. 2011. Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: a review. *Process Biochemistry*, **46**(1), 23-34.
Mann, D., Labbé, N., Sykes, R., Gracom, K., Kline, L., Swamidoss, I., Burris, J., Davis, M., Stewart, C. 2009. Rapid assessment of lignin content and structure in switchgrass (*Panicum virgatum* L.) grown under different environmental conditions. *BioEnergy Research*, **2**(4), 246-256.
Marshall, A.G., Rodgers, R.P. 2008. Petroleomics: chemistry of the underworld. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(47), 18090-18095.
Martin, D., Tholl, D., Gershenzon, J., Bohlmann, J. 2002. Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiology*, **129**(3), 1003-1018.
Martin, M.A. 2010. First generation biofuels compete. *New Biotechnology*, **27**(5), 596-608.
Mata, T.M., Martins, A.A., Caetano, N.S. 2010. Microalgae for biodiesel production and other applications: A review. *Renewable & Sustainable Energy Reviews*, **14**(1), 217-232.
Matishev, V.A. 1994. Primary and secondary petroleum refining processes in the pages of the journal chemistry and technology of fuels and oils during the first decade of publication. *Chemistry and Technology of Fuels and Oils*, **30**(9-10), 377-384.
Maugh, T.H. 1979. Unlike money, diesel fuel does grow on trees. *Science*, **206**(4417), 436.
McCaskill, D., Croteau, R. 1997. Prospects for the bioengineering of isoprenoid biosynthesis. *Advances in Biochemical Engineering/Biotechnology*, **55**, 107-146.
McMichael, A.J., Powles, J.W., Butler, C.D., Uauy, R. 2007. Food, livestock production, energy, climate change, and health. *The Lancet*, **370**(9594), 1253-1263.
Medeiros, J.R., Campos, L.B., Mendonça, S.C., Davin, L.B., Lewis, N.G. 2003. Composition and antimicrobial activity of the essential oils from invasive species of the Azores, *Hedychium gardnerianum* and *Pittosporum undulatum*. *Phytochemistry*, **64**(2), 561-565.

Medeiros, R.D., Vieira, G. 2008. Sustainability of extraction and production of copaiba (Copaifera multijuga Hayne) oleoresin in Manaus, AM, Brazil. *Forest Ecology and Management*, **256**(3), 282-288.
Melero, J.A., Clavero, M.M., Calleja, G., Garcia, A., Miravalles, R., Galindo, T. 2010. Production of biofuels via the catalytic cracking of mixtures of crude vegetable oils and nonedible animal fats with vacuum gas oil. *Energy and Fuels*, **24**, 707-717.
Mirov, N. 1952. Composition of gum turpentines of pines. XIV. A report on three mexican pines: *Pinus ayacahuite*, *Pinus cembroides* and *Pinus pinceana*. *Journal of the American Pharmaceutical Association*, **41**(12), 673-676.
Misawa, N. 2011. Pathway engineering for functional isoprenoids. *Current Opinion in Biotechnology*, **22**(5), 627-633.
Moerschbaecher, M., Day Jr., J.W. 2011. Ultra-deepwater Gulf of Mexico oil and gas: energy return on financial investment and a preliminary assessment of energy return on energy investment. *Sustainability*, **3**(10), 2009-2026.
Montgomery, D.C. 2008. *Design and analysis of experiments*. John Wiley & Sons Inc.
Monti, H., Tiliacos, N., Faure, R. 1999. Copaiba oil: isolation and characterization of a new diterpenoid with the dinorlabdane skeleton. *Phytochemistry*, **51**(8), 1013-1015.
Monti, H., Tiliacos, N., Faure, R. 1996. Two diterpenoids from copaiba oil. *Phytochemistry*, **42**(6), 1653-1656.
Mooney, B.P. 2009. The second green revolution? Production of plant-based biodegradable plastics. *Biochemical Journal*, **418**, 219-232.
Moser, B.R. 2010. Camelina (*Camelina sativa* L.) oil as a biofuels feedstock: golden opportunity or false hope? *Lipid Technology*, **22**(12), 270-273.
Murashige, T., Skoog, F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, **15**(3), 473-497.
Nemethy, E.K., Calvin, M. 1982. Terpenes from Pittosporaceae. *Phytochemistry*, **21**(12), 2981-2982.
Neto, J.D.L., Gramosa, N.V., Silveira, E.R. 2008. Chemical constituents of the fruits of *Copaifera langsdorffii* desf. *Quimica Nova*, **31**(5), 1078-1080.
Newton, P., Watkinson, A.R., Peres, C.A. 2011. Determinants of yield in a non-timber forest product: *Copaifera* oleoresin in Amazonian extractive reserves. *Forest Ecology and Management*, **261**(2), 255-264.
Ng, J., Ng, H.K., and S. Gan. 2010. Advances in biodiesel fuel for application in compression ignition engines. *Clean Technologies and Environmental Policy*, **12**(5): 459-493.
Nigam, P.S., Singh, A. 2011. Production of liquid biofuels from renewable resources. *Progress in Energy and Combustion Science*, **37**(1), 52-68.

Nishiyama, Y., Yun, C.-S., Matsuda, F., Sasaki, T., Saito, K., Tozawa, Y. 2010. Expression of bacterial tyrosine ammonia-lyase creates a novel p-coumaric acid pathway in the biosynthesis of phenylpropanoids in *Arabidopsis*. *Planta*, **232**(1), 209-218.

O'Maille, P.E., Chappell, J., Noel, J.P. 2004. A single-vial analytical and quantitative gas chromatography–mass spectrometry assay for terpene synthases. *Analytical Biochemistry*, **335**(2), 210-217.

Oakes, J., Brackenridge, D., Colletti, R., Daley, M., Hawkins, D.J., Xiong, H., Mai, J., Screen, S.E., Val, D., Lardizabal, K., Gruys, K., Deikman, J. 2011. Expression of fungal diacylglycerol acyltransferase2 genes to increase kernel oil in maize. *Plant Physiology*, **155**(3), 1146-1157.

Octave, S., Thomas, D. 2009. Biorefinery: Toward an industrial metabolism. *Biochimie*, **91**(6), 659-664.

Odell, E., Raguso, R.A., Jones, K.N. 1999. Bumblebee foraging responses to variation in floral scent and color in snapdragons (*Antirrhinum* : Scrophulariaceae). *American Midland Naturalist*, **142**(2), 257-265.

Ohsaki, A., Yan, L.T., Ito, S., Edatsugi, H., Iwata, D., Komoda, Y. 1994. The isolation and *in vivo* potent antitumor-activity of clerodane diterpenoid from the oleoresin of the Brazilian medicinal plant, *Copaifera langsdorffii*. *Bioorganic & Medicinal Chemistry Letters*, **4**(24), 2889-2892.

Ong, Y.K., Bhatia, S. 2010. The current status and perspectives of biofuel production via catalytic cracking of edible and non-edible oils. *Energy*, **35**(1), 111-119.

Osborn, S.G., Vengosh, A., Warner, N.R., Jackson, R.B. 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(20), 8172-8176.

Paiva, L.A.F., Gurgel, L.A., De Sousa, E.T., Silveira, E.R., Silva, R.M., Santos, F.A., Rao, V.S.N. 2004. Protective effect of *Copaifera langsdorffii* oleo-resin against acetic acid-induced colitis in rats. *Journal of Ethnopharmacology*, **93**(1), 51-56.

Paiva, L.A.F., Gurgel, L.A., Silva, R.M., Tome, A.R., Gramosa, N.V., Silveira, E.R., Santos, F.A., Rao, V.S.N. 2002. Anti-inflammatory effect of kaurenoic acid, a diterpene from *Copaifera langsdorffii* on acetic acid-induced colitis in rats. *Vascular Pharmacology*, **39**(6), 303-307.

Paiva, L.A.F., Rao, V.S.N., Gramosa, N.V., Silveira, E.R. 1998. Gastroprotective effect of *Copaifera langsdorffii* oleo-resin on experimental gastric ulcer models in rats. *Journal of Ethnopharmacology*, **62**(1), 73-78.

Paltsev, S., Jacoby, H.D., Reilly, J.M., Ejaz, Q.J., Morris, J., O'Sullivan, F., Rausch, S., Winchester, N., Kragha, O. 2011. The future of U.S. natural gas production, use, and trade. *Energy Policy*, **39**(9), 5309-5321.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., Rokhsar, D.S. 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**(7229), 551-556.

Pauly, M., Keegstra, K. 2010. Plant cell wall polymers as precursors for biofuels. *Current Opinion in Plant Biology*, **13**(3), 304-311.

Pedroni, F., Sanchez, M., Santos, F.A.M. 2002. Fenologia da copaiba (*Copaifera langsdorffii* Desf. – Leguminosae, Caesalpinioideae) emu ma floresta semidecídua no sudeste do Brasil. *Revista Brasil Botanica*, **25**, 183-194.

Peralta-Yahya, P.P., Keasling, J.D. 2010. Advanced biofuel production in microbes. *Biotechnology Journal*, **5**(2), 147-162.

Perlack, R.D., Wright, L.L., Turhollow, A., Graham, R., Stokes, B., Erbach, D. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply, (Eds.) U.S.Department of Energy, U.S.Department of Agriculture, UT-Battele. Oak Ridge, TN.

Pfromm, P.H., Amanor-Boadu, V., Nelson, R., Vadlani, P., Madl, R. 2010. Bio-butanol vs. bio-ethanol: a technical and economic assessment for corn and switchgrass fermented by yeast or *Clostridium acetobutylicum*. *Biomass and Bioenergy*, **34**(4), 515-524.

Pingali, P., Raney, T., Wiebe, K. 2008. Biofuels and food security: missing the point. *Applied Economic Perspectives and Policy*, **30**(3), 506-516.

Pinzi, S., Leiva, D., Arzamendi, G., Gandia, L.M., Dorado, M.P. 2011. Multiple response optimization of vegetable oils fatty acid composition to improve biodiesel physical properties. *Bioresource Technology*, **102**(15), 7280-7288.

Pitz, W.J., Cernansky, N.P., Dryer, F.L., Egolfopoulos, F.N., Farrell, J.T., Friend, D.G., Pitsch, H. 2007. Development of an experimental database and kinetic models for surrogate gasoline fuels. in: *2007 SAE World Congress*, Vol. 2007-01-0175, Lawrence Livermore National Laboratory. Detroit, MI, USA.

Pitz, W.J., Mueller, C.J. 2011. Recent progress in the development of diesel surrogate fuels. *Progress in Energy and Combustion Science*, **37**(3), 330-350.

Plowden, C. 2004. The ethnobotany of copaiba (*Copaifera*) oleoresin in the Amazon. *Economic Botany*, **58**, 729-739.

Plowden, C. 2003. Production ecology of copaiba (*Copaifera* spp.) oleoresin in the Eastern Brazilian Amazon. *Economic Botany*, **57**(4), 491-501.
Poola, R.B., Nagalingam, B., Gopalakrishnan, K. 1994. Performance studies with biomass-derived high-octane fuel additives in a two-stroke spark-ignition engine. *Biomass and Bioenergy*, **6**(5), 369-379.
Prashar, A., Hili, P., Veness, R.G., Evans, C.S. 2003. Antimicrobial action of palmarosa oil (*Cymbopogon martinii*) on *Saccharomyces cerevisiae*. *Phytochemistry*, **63**(5), 569-575.
Prikhodko, V.Y., Pihl, J.A., Lewis, S.S.A., Parks, J.E., Ii. 2012. Effect of hydrocarbon emissions from PCCI-type combustion on the performance of selective catalytic reduction catalysts. *Journal of Engineering for Gas Turbines and Power*, **134**(8), 082804-5.
Prisic, S., Xu, M., Wilderman, P.R., Peters, R.J. 2004. Rice contains two disparate ent-copalyl diphosphate synthases with distinct metabolic functions. *Plant Physiology*, **136**(4), 4228-4236.
Purushothaman, K., Nagarajan, G. 2009. Performance, emission and combustion characteristics of a compression ignition engine operating on neat orange oil. *Renewable Energy*, **34**(1), 242-245.
Rahm, D. 2011. Regulating hydraulic fracturing in shale gas plays: The case of Texas. *Energy Policy*, **39**(5), 2974-2981.
Rahmat, N., Abdullah, A.Z., Mohamed, A.R. 2010. Recent progress on innovative and potential technologies for glycerol transformation into fuel additives: A critical review. *Renewable & Sustainable Energy Reviews*, **14**(3), 987-1000.
Rajeswara Rao, B.R., Rajput, D.K., Patel, R.P. 2009. Essential oil profiles of different parts of palmarosa (*Cymbopogon martinii* (Roxb.) Wats. var. motia Burk.). *Journal of Essential Oil Research*, **21**(6), 519-521.
Ram, M., Patra, D.D., Kumar, S. 1997. Nitrogen fertilization of palmarosa in continuous cropping. *Journal of Herbs, Spices & Medicinal Plants*, **5**(1), 53-61.
Rao, B.R.R. 2001. Biomass and essential oil yields of rainfed palmarosa (*Cymbopogon martinii* (Roxb.) Wats. var. motia Burk.) supplied with different levels of organic manure and fertilizer nitrogen in semi-arid tropical climate. *Industrial Crops and Products*, **14**(3), 171-178.
Rao, B.R.R., Kaul, P.N., Syamasundar, K.V., Ramesh, S. 2005. Chemical profiles of primary and secondary essential oils of palmarosa (*Cymbopogon martinii* (Roxb.) Wats var. motia Burk.). *Industrial Crops and Products*, **21**(1), 121-127.
Renner, T., Waters, E.R. 2007. Comparative genomic analysis of the Hsp70s from five diverse photosynthetic eukaryotes. *Cell stress and Chaperones*, **12**(2), 172.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.-H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., Boore, J.L. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*, **319**(5859), 64-69.

Rigamonte-Azevedo, O.C., Wadt, P.G.S., Oliveira, L.H.W. 2006. Potencial de produção de oleo-resina de copaiba (*Copaifera* spp) de populaces naturais do sudoeste da Amazônia. *Revista Árvore*, **20**, 583-591.

Rigamonte-Azevedo, O.C., Wadt, P.G.S., Wadt, L.H.O. 2004. Copaíba: ecologia e Produção de Óleoresina. *EMBRAPA*. ISSN 0104-9046. In Portuguese.

Rizwan, A.M., Dennis, L.Y.C., Liu, C. 2008. A review on the generation, determination and mitigation of urban heat island. *Journal of Environmental Sciences*, **20**(1), 120-128.

Ro, D.K., Bohlmann, J. 2006. Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*): Functional characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1). *Phytochemistry*, **67**(15), 1572-1578.

Rodrigues, K.C.S., Fett-Neto, A.G. 2009. Oleoresin yield of *Pinus elliottii* in a subtropical climate: seasonal variation and effect of auxin and salicylic acid-based stimulant paste. *Industrial Crops and Products*, **30**(2), 316-320.

Roytrakul, S., Verpoorte, R. 2007. Role of vacuolar transporter proteins in plant secondary metabolism: *Catharanthus roseus* cell culture. *Phytochemistry Reviews*, **6**(2), 383-396.

Sannigrahi, P., Ragauskas, A.J., Tuskan, G.A. 2010. Poplar as a feedstock for biofuels: a review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, **4**(2), 209-226.

Sarma, A., Sarma, T. 2005. Studies on the morphological characters and yields of oil and citral of certain lemongrass [*Cymbopogon flexuosus* (Steud) Wats] accessions grown under agro-climatic conditions of Northeast India. *Journal of Essential Oil Bearing Plants*, **8**, 250-257.

Savage, T.J., Hristova, M.K., Croteau, R. 1996. Evidence for an elongation/reduction/C1-elimination pathway in the biosynthesis of n-heptane in xylem of Jeffrey pine. *Plant Physiology*, **111**(4), 1263-1269.

Schirmer, A., Rude, M.A., Li, X., Popova, E., del Cardayre, S.B. 2010. Microbial biosynthesis of alkanes. *Science*, **329**(5991), 559-562.

Schmer, M.R., Vogel, K.P., Mitchell, R.B., Dien, B.S., Jung, H.G., Casler, M.D. 2012. Temporal and spatial variation in switchgrass biomass composition and theoretical ethanol yield. *Agronomy Journal*, **104**(1), 54-64.

Schmidhuber, J., Tubiello, F.N. 2007. Global food security under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(50), 19703-19708.

Shen, B., Allen, W.B., Zheng, P., Li, C., Glassman, K., Ranch, J., Nubel, D., Tarczynski, M.C. 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiology*, **153**(3), 980-987.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., Burns, P., Davis, T.M., Slovin, J.P., Bassil, N., Hellens, R.P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O.R., Jensen, R.V., Allan, A.C., Michael, T.P., Setubal, J.C., Celton, J.-M., Rees, D.J.G., Williams, K.P., Holt, S.H., Rojas, J.J.R., Chatterjee, M., Liu, B., Silva, H., Meisel, L., Adato, A., Filichkin, S.A., Troggio, M., Viola, R., Ashman, T.-L., Wang, H., Dharmawardhana, P., Elser, J., Raja, R., Priest, H.D., Bryant, D.W., Fox, S.E., Givan, S.A., Wilhelm, L.J., Naithani, S., Christoffels, A., Salama, D.Y., Carter, J., Girona, E.L., Zdepski, A., Wang, W., Kerstetter, R.A., Schwab, W., Korban, S.S., Davik, J., Monfort, A., Denoyes-Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J.L., Salzberg, S.L., Dickerman, A.W., Velasco, R., Borodovsky, M., Veilleux, R.E., Folta, K.M. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics*, **43**(2), 109-116.

Simmons, B.A., Loqué, D., Ralph, J. 2010. Advances in modifying lignin for enhanced biofuel production. *Current Opinion in Plant Biology*, **13**(3), 312-319.

Singh, M. 1999. Effect of irrigation and nitrogen on herbage, oil yield and water use of lemongrass (*Cymbopogon flexuosus*) on alfisols. *The Journal of Agricultural Science*, **132**(02), 201-206.

Singh, M. 2001. Long-term studies on yield, quality and soil fertility of lemongrass (*Cymbopogon flexuosus*) in relation to nitrogen application. *Journal of Horticultural Science & Biotechnology*, **76**(2), 180-182.

Singh, M., Sharma, S. 2001. Influence of irrigation and nitrogen on herbage and oil yield of palmarosa (*Cymbopogon martinii*) under semi-arid tropical conditions. *European Journal of Agronomy*, **14**(2), 157-159.

Singh, M., Shivaraj, B., Sridhara, S. 1996. Effect of plant spacing and nitrogen levels on growth, herb and oil yields of lemongrass (*Cymbopogon flexuosus* (Steud.) Wats. var. I cauvery). *Journal of Agronomy and Crop Science*, **177**(2), 101-105.

Singh, N., Luthra, R., Sangwan, R.S. 1989. Effect of leaf position and age on the essential oil quantity and quality in lemongrass (*Cymbopogon flexuosus*). *Planta Medica*, **55**(3), 254-256.

Sovacool, B.K. 2009. Rejecting renewables: The socio-technical impediments to renewable electricity in the United States. *Energy Policy*, **37**(11), 4500-4513.

Speight, J.G. 1999. *The Chemistry and Technology of Petroleum*. *3rd ed*. Marcel Dekker, New York.

Speight, J.G. 2002. *Handbook of Petroleum Product Analysis*. Wiley-Interscience, New York.

Starck, L., Lecointe, B., Forti, L., Jeuland, N. 2010. Impact of fuel characteristics on HCCI combustion: performances and emissions. *Fuel*, **89**(10), 3069-3077.

Starks, C.M., Back, K., Chappell, J., Noel, J.P. 1997. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science*, **277**(5333), 1815-1820.

Stashenko, E., Wiame, H., Dassy, S., Martinez, J.R., Shibamoto, T. 1995. Catalytic transformation of copaiba (*Copaifera officinalis*) oil over zeolite-ZSM-5. *Hrc-Journal of High Resolution Chromatography*, **18**(1), 54-58.

Steen, E.J., Kang, Y.S., Bokinsky, G., Hu, Z.H., Schirmer, A., McClure, A., del Cardayre, S.B., Keasling, J.D. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, **463**(7280), 559-662.

Steuernagel, B., Taudien, S., Gundlach, H., Seidel, M., Ariyadasa, R., Schulte, D., Petzold, A., Felder, M., Graner, A., Scholz, U., Mayer, K., Platzer, M., Stein, N. 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**(1), 547.

Strzepek, K., Boehlert, B. 2010. Competition for water for the food system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**(1554), 2927-2940.

Stupp, T., de Freitas, R.A., Sierakowski, M.R., Deschamps, F.C., Wisniewski, A., Biavatti, M.W. 2008. Characterization and potential uses of *Copaifera langsdorffii* seeds and seed oil. *Bioresource Technology*, **99**(7), 2659-2663.

Suriyamongkol, P., Weselake, R., Narine, S., Moloney, M., Shah, S. 2007. Biotechnological approaches for the production of polyhydroxyalkanoates

in microorganisms and plants - A review. *Biotechnology Advances*, **25**(2), 148-175.

Sykes, R., Yung, M., Novaes, E., Kirst, M., Peter, G., Davis, M. 2009. High throughput screening of plant cell wall composition using pyrolysis molecular beam mass spectroscopy. *Biofuels: Methods and Protocols. Humana Press, New York*, 169-183.

Tamilvendhan, D., Ilangovan, V. 2011. A performance, emission and combustion investigation on hot air assisted *Eucalyptus* oil direct injected compression ignition engine. *Modern Applied Science*, **5**(4), p53.

Tang, Q., Ma, X., Mo, C., Wilson, I.W., Song, C., Zhao, H., Yang, Y., Fu, W., Qiu, D. 2011. An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics*, **12**(1), 343.

Tholl, D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology*, **9**(3), 297-304.

Thompson, R.C., Swan, S.H., Moore, C.J., vom Saal, F.S. 2009. Our plastic age. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**(1526), 1973-1976.

Tracy, N.I., Chen, D., Crunkleton, D.W., Price, G.L. 2009. Hydrogenated monoterpenes as diesel fuel additives. *Fuel*, **88**(11), 2238-2240.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**(5793), 1596-1604.

Tyson, K.S. 2004. Biodiesel handling and use guidelines, (Ed.) U.S.Department of Energy, Office of Science and Technical Information. Oak Ridge.

Veiga Junior, V.F., Pinto, A.C. 2002. O gênero *Copaifera* L. *Quím Nova*, **25**, 273-286.

Veiga Junior, V.F., Rosas, E.C., Carvalho, M.V., Henriques, M.G.M.O., Pinto, A.C. 2007. Chemical composition and anti-inflammatory activity of copaiba oils from *Copaifera cearensis* Huber ex Ducke, *Copaifera reticulata* Ducke and *Copaifera multijuga* Hayne—A comparative study. *Journal of Ethnopharmacology*, **112**(2), 248-254.

Veiga, V.F., Pinto, A.C. 2002. The *Copaifera* L. genus. *Quimica Nova*, **25**(2), 273-286.

Veiga, V.F., Zunino, L., Patitucci, M.L., Pinto, A.C., Calixto, J.B. 2006. The inhibition of paw oedema formation caused by the oil of *Copaifera multijuga* Hayne and its fractions. *Journal of Pharmacy and Pharmacology*, **58**(10), 1405-1410.

Visser, R., Bachem, C., de Boer, J., Bryan, G., Chakrabati, S., Feingold, S., Gromadka, R., van Ham, R., Huang, S., Jacobs, J., Kuznetsov, B., de Melo, P., Milbourne, D., Orjeda, G., Sagredo, B., Tang, X. 2009. Sequencing the potato genome: outline and first results to come from the elucidation of the sequence of the world's third most important food crop. *American Journal of Potato Research*, **86**(6), 417-429.

Vogel, K., Dien, B., Jung, H., Casler, M., Masterson, S., Mitchell, R. 2011. Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses. *BioEnergy Research*, **4**(2), 96-110.

Wallington, T.J., Kaiser, E.W., Farrell, J.T. 2006. Automotive fuels and internal combustion engines: a chemical perspective. *ChemInform*, **37**(25), 335-347.

Weathers, P., Arsenault, P., Covello, P., McMickle, A., Teoh, K., Reed, D. 2011. Artemisinin production in *Artemisia annua*: studies *in planta* and results of a novel delivery method for treating malaria and other neglected diseases. *Phytochemistry Reviews*, **10**(2), 173-183.

Weathers, P., Towler, M., Xu, J. 2010. Bench to batch: advances in plant cell culture for producing useful products. *Applied Microbiology and Biotechnology*, **85**(5), 1339-1351.

Weiss, E.A. 1997. *Essential oil crops*. Cab International.

Weselake, R.J., Shah, S., Tang, M., Quant, P.A., Snyder, C.L., Furukawa-Stoffer, T.L., Zhu, W., Taylor, D.C., Zou, J., Kumar, A., Hall, L., Laroche, A., Rakow, G., Raney, P., Moloney, M.M., Harwood, J.L. 2008. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *Journal of Experimental Botany*, **59**(13), 3543-3549.

Weston, R.J. 2004. Composition of essential oils from the leaves of seven New Zealand species of *Pittosporum* (Pittosporaceae). *Journal of Essential Oil Research*, **16**(5), 453-458.

Wijffels, R.H., Barbosa, M.J. 2010. An outlook on microalgal biofuels. *Science*, **329**(5993), 796-799.

Williams, D.C., McGarvey, D.J., Katahira, E.J., Croteau, R. 1998. Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry*, **37**(35), 12213-12220.

Wiselogel, A.E., Agblevor, F.A., Johnson, D.K., Deutch, S., Fennell, J.A., Sanderson, M.A. 1996. Compositional changes during storage of large round switchgrass bales. *Bioresource Technology*, **56**(1), 103-109.

Wohlgemuth, R. 2009. The locks and keys to industrial biotechnology. *New Biotechnology*, **25**(4), 204-213.

Wullschleger, S.D., Davis, E.B., Borsuk, M.E., Gunderson, C.A., Lynd, L.R. 2010. Biomass production in switchgrass across the United States: database description and determinants of yield. *Agrononmy Journal*, **102**(4), 1158-1168.

Xue, J., Grift, T.E., Hansen, A.C. 2011. Effect of biodiesel on engine performances and emissions. *Renewable and Sustainable Energy Reviews*, **15**(2), 1098-1116.

Yaman, S. 2004. Pyrolysis of biomass to produce fuels and chemical feedstocks. *Energy Conversion and Management*, **45**(5), 651-671.

Yan, J., Hu, Z., Pu, Y., Brummer, C.E., Ragauskas, A.J. 2010. Chemical compositions of four switchgrass populations. *Biomass and Bioenergy*, **34**(1), 48-53.

Yang, Y., Yuan, J.S., Ross, J., Noel, J.P., Pichersky, E., Chen, F. 2006. An *Arabidopsis thaliana* methyltransferase capable of methylating farnesoic acid. *Archives of Biochemistry and Biophysics*, **448**(1-2), 123-132.

Yang, Z., Ohlrogge, J.B. 2009. Turnover of fatty acids during natural senescence of *Arabidopsis*, *Brachypodium*, and switchgrass in *Arabidopsis* β-oxidation mutants. *Plant Physiology*, **150**(4), 1981-1989.

Yuan, J., Kollner, T., Wiggins, G., Grant, J., Degenhardt, J., Chen, F. 2008a. Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *The Plant Journal*.

Yuan, J.S., Tiller, K.H., Al-Ahmad, H., Stewart, N.R., Stewart Jr, C.N. 2008b. Plants to power: bioenergy to fuel the future. *Trends in Plant Science*, **13**(8), 421-429.

Yuan, J.S., Himanen, S.J., Holopainen, J.K., Chen, F., Stewart, C.N. 2009. Smelling global climate change: mitigation of function for plant volatile organic compounds. *Trends in Ecology & Evolution*, **24**(6), 323-331.

Yumrutas, R., Alma, M.H., Özcan, H., Kaska, Ö. 2008. Investigation of purified sulfate turpentine on engine performance and exhaust emission. *Fuel*, **87**(2), 252-259.

Zerbe, P., Chiang, A., Bohlmann, J. 2012. Mutational analysis of white spruce (*Picea glauca*) *ent*-kaurene synthase (*PgKS*) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. *Phytochemistry*, **74**, 30-39.

Zhang, H., Memelink, J. 2009. Regulation of secondary metabolism by jasmonate hormones. in: *Plant-derived Natural Products*, (Eds.) A.E. Osbourn, V. Lanzotti, Springer New York, pp. 181-194.

Zhang, J., Tang, M., Viikari, L. 2012. Xylans inhibit enzymatic hydrolysis of lignocellulosic materials by cellulases. *Bioresource Technology*, **121**, 8-12.

Zhang, P., Dreher, K., Karthikeyan, A., Chi, A., Pujar, A., Caspi, R., Karp, P., Kirkup, V., Latendresse, M., Lee, C., Mueller, L.A., Muller, R., Rhee, S.Y. 2010. Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiology*, **153**(4), 1479-1491.

Zharkikh, A., Troggio, M., Pruss, D., Cestaro, A., Eldrdge, G., Pindo, M., Mitchell, J.T., Vezzulli, S., Bhatnagar, S., Fontana, P., Viola, R., Gutin, A., Salamini, F., Skolnick, M., Velasco, R. 2008. Sequencing and assembly of highly heterozygous genome of *Vitis vinifera* L. cv Pinot Noir: problems and solutions. *Journal of Biotechnology*, **136**(1–2), 38-43.

Zheljazkov, V.D., Cantrell, C.L., Astatkie, T., Cannon, J.B. 2011. Lemongrass productivity, oil content, and composition as a function of nitrogen, sulfur, and harvest time *Agrononmy Journal*, **103**(3), 805-812.

Zheljazkov, V.D., Cantrell, C.L., Astatkie, T., Ebelhar, M.W. 2010. Productivity, oil content, and composition of two spearmint species in Mississippi. *Agronomy Journal*, **102**(1), 129-133.

Zheljazkov, V.D., Cantrell, C.L., Tekwani, B., Khan, S.I. 2007. Content, composition, and bioactivity of the essential oils of three basil genotypes as a function of harvesting. *Journal of Agricultural and Food Chemistry*, **56**(2), 380-385.

Zhou, W., Nguyen, H.T., Nes, W.D. 2008. Plant sterol methyltransferases: phytosterolomic analysis, enzymology, and bioengineering strategies. in: *Advances in Plant Biochemistry and Molecular Biology*, (Eds.) H.N. Hans J. Bohnert, G.L. Norman, Vol. Volume 1, Pergamon, pp. 241-281.

Zoghbi, M.D.B., Lameira, O.A., Oliveira, E.C.R. 2007. Seasonal variation of oleoresin and volatiles from Copaifera martii Hayne growing wild in the state of Para, Brazil. *Journal of Essential Oil Research*, **19**(6), 504-506.

VITA

Blake Lee Joyce was born in Atlanta, GA on March 20, 1984. He was raised in Marietta, GA where he attended elementary school, middle school, and graduated from Wheeler High School in 2002. From there he went to the University of Georgia in Athens, GA and received a B.S. in Biology and another B.S. in Ecology in May, 2006. While working for Dr. Wayne Parrott at the University of Georgia, he met Dr. Neal Stewart in Minneapolis, MN during the Society for *In Vitro* Biology meeting May, 2006 and started work as a graduate research assistant in January, 2007. He received his Master of Science degree in Plant Sciences with a minor in Statistics from the University of Tennessee, Knoxville in December 2008. Finally, he received his Ph.D. in August 2013 with the help of his committee, friends, and wife.