NSF PRFB – Proposal – Project Summary

Title: Genes controlling wax biosynthesis in *Sorghum bicolor*: potential for improving crop performance and value
Project Director:
Primary Mentor:

Overview

Sorghum bicolor is a multi-use crop with exceptional water- and nitrogen-use efficiency whose expanded use will likely decrease water and fertilizer consumption and create a more sustainable agricultural system. Currently, sorghum is generally not economically competitive with maize when irrigation is available, so sorghum harvest value needs to be increased. Waxes cover aerial plant surfaces where they protect against dry climates and increase water-use efficiency. On sorghum, waxes accumulate to levels higher than nearly all other plant species and sorghum kernel waxes are an emerging industrial bioproduct that can be extracted during grain processing. Sorghum waxes are therefore an important research target because understanding them will enable us to (i) potentially transfer some of sorghum's water efficiency to other crops such as maize, and (ii) rapidly navigate routes to high-wax sorghum lines with enhanced grain value.

Intellectual Merit

The goal of the proposed project is to combine wax biochemistry with genomics and bioinformatics to identify genes controlling sorghum wax deposition via two complimentary specific aims: (i) identify high-confidence sorghum wax gene candidates by correlating epidermis-specific transcriptomes and comprehensive wax profiles, then evaluate their functionality by performing complementation tests in Arabidopsis wax mutants, and (ii) use bulked segregant RNA-seq mapping to pinpoint causal mutations in existing sorghum mutants affected in both kernel and leaf sheath waxes (as putative transcription factor mutants). I am an expert in wax biochemistry. Mentored by a crop biochemist, a sorghum genomics expert, and a bioinformatician, I will carry out a rigorous training and research plan that will equip me for a career in developing and applying knowledge of plant waxes to sustaining and improving agriculture.

Broader Impacts

The broader impacts of this project include the recruitment of two undergraduate chemistry students and their training in phytochemical analysis. In addition to carrying out research, these students will help develop and administer (i) a hands-on active learning module that will be presented at the University of Nebraska Science Museum's Sunday with a Scientist and (ii) a hands-on laboratory workshop for the Women in Science Conference in Lincoln, NE. This module and workshop will educate children, adults and young women about the economic, nutritional, and environmental benefits gained through the study of plant chemistry, plant diversity, and crop improvement and inspire interest in STEM fields, education, and careers. In addition, to initiate and catalyze future sorghum wax research and the further development of applications for sorghum wax knowledge, the whole tissue and epidermis-specific gene expression data collected here will be hosted on an online server that will be developed as part of this project: the Sorghum Epidermis Gene Expression Almanac, where users can BLAST query sequences against the sorghum transcriptome and view the expression of similar genes in whole tissue and the epidermis of sorghum leaves, leaf sheathes, and developing kernels.

1. Introduction - Sorghum is a close, diploid relative of maize that is grown for grain, sugar, and forage. Aside from its naturally gluten-free grain, the exceptional water- and nitrogen-use efficiency of sorghum offer great potential for its expanded use. However, sorghum grain yields are often lower than those of irrigated maize (Farre and Faci, 2006). Accordingly, maize is grown on substantial acreage previously dedicated and better suited to sorghum (Laingen, 2015), which has increased producer input costs, aquifer depletion, and fertilizer run-off. To make sorghum more economically competitive, the value of a sorghum harvest needs to be enhanced. Meeting this need quickly will not only require increasing grain yield, but also parallel research efforts focused on creating value from diverse aspects of sorghum.

Figure 1: Sorghum with leaf sheath waxes (white; left) and a leaf sheath wax mutant (green; right). Photo: Dr. Matthew Jenks.

Sorghum kernel waxes, lipids that can be extracted during grain processing, have properties similar to waxes from the carnauba palm (Weller et al., 2000), which are used widely in industry. The U.S. has no domestic source for carnauba waxes, even though domestic and global demand for these natural waxes is increasing (current global demand: 200 million pounds/year at ~$4.30/pound; Grand View Research, 2016). Waxes cover virtually all above-ground external surfaces of land plants including sorghum, and are well known for protecting plants from drying out (e.g. Jenks et al., 1994; Lü et al., 2012). Consistent with this, drought-tolerant crops such as pearl millet, which can grow on the edge of the Sahara Desert, have substantially expanded wax-related gene families (Varshney et al., 2017). Wax biosynthesis in sorghum leaf sheaths is so active that, midway through development, waxes accumulate visibly as a white coating (Fig. 1, left). Thus, basic knowledge of sorghum wax biosynthesis has application in (i) improving kernel waxes, an emerging sorghum bioproduct with great potential to increase grain value and (ii) furthering our understanding the role sorghum's leaf sheath waxes play in its exceptional water-use efficiency.

Most of our basic knowledge of plant wax biosynthesis comes from *Arabidopsis thaliana*. Since epidermal cells biosynthesize waxes and wax composition can vary between surfaces of the same plant, information on epidermis-specific gene expression (Suh et al., 2005) and correlations between gene expression and wax profiles (Busta et al., 2016b) provide high-confidence candidates for testing and functional wax gene discovery (e.g. Bird et al., 2007; Greer et al., 2007; Li et al., 2008, Hegebarth et al., 2017). However, not all wax genes are upregulated in the epidermis, so complimentary approaches are necessary to discover certain participants in wax biosynthesis, including regulatory machinery (Hooker et al., 2007). Since wax mutants have visibly altered exteriors (for sorghum: Fig. 1, right), visual screens and genetic mapping can pinpoint underlying wax genes (e.g. Millar et al., 1999; Xia et al., 1996). By combining comparative gene expression and mutant mapping approaches, wax biosynthesis in Arabidopsis has been well characterized: approximately 15-20 genes are responsible for coating its aerial surfaces in fatty acids, alcohols, alkanes, and ketones. Mutant lines for nearly all Arabidopsis wax biosynthesis enzymes and several associated transcription factors are available.

Epidermis-specific gene expression and mutant mapping have also led to wax gene discovery in crops (e.g. rice, Wang et al., 2017; maize, Li et al., 2013; and wheat, Lu et al., 2015). Based on these studies, species with fatty acid, alcohol, alkane, or ketone waxes (like Arabidopsis) also have wax lipid biosynthesis genes similar to those in Arabidopsis that can be identified by homology (e.g. Wang et al., 2015, Buda et al., 2013;), though gene families can complicate this process, especially in species with large or complex genomes, such as sorghum (Paterson et al., 2009). Even so, functional gene family members can be identified via complementation testing in Arabidopsis wax mutants, as exemplified by genes from cotton and wheat (Qin et al., 2007; Wang et al., 2015).

Unlike Arabidopsis, wheat, rice, and maize, our knowledge of sorghum wax biosynthesis is extremely limited. We do know that sorghum waxes are mainly alcohols (on kernels; Leguizamón et al., 2009) and fatty acids (on leaf sheaths; Jenks et al., 2000), indicating that some sorghum wax genes can probably be found by homology. However, when sorghum wax mutants were generated and subjected to allelism tests, 31 non-allelic lines were isolated (Peters et al., 2009), suggesting that the number of wax genes in sorghum is probably much higher than in Arabidopsis, and that mapping will be necessary to find some of these. So far, only one sorghum wax gene, a transporter, has been identified (Mizuno et al., 2013). Thus, there is still a ***problem*** as about 30 sorghum wax genes remain unidentified, including transcription factors and biosynthetic genes that would allow us to increase wax yields or optimize wax composition for industry. This is preventing us from taking full advantage of sorghum waxes as industrial products and from fully understanding sorghum's water-use efficiency. Therefore, there is a ***critical need*** to advance our basic knowledge of sorghum wax biosynthesis. In the absence of such information, it will not be possible to navigate long-term, parallel avenues toward (i) a complete understanding of plant water-use efficiency and (ii) increasing sorghum kernel wax amounts and thus the value of a sorghum harvest.

2. Specific Training and Research Goals - My *long-term career goal* is to become a leader in integrating chemistry and plant genomics to develop and apply fundamental knowledge of plant surface lipids to sustaining and improving agriculture. The *objective in this proposal* is to identify genes controlling the accumulation and composition of sorghum wax lipids. The *rationale* for this research is that knowledge of these genes is a prerequisite for investigations of how sorghum waxes influence water-use efficiency and for rapidly breeding valuable high-wax sorghum lines. This work is therefore of potentially high *significance* because it is expected to (i) enable studies on how to increase the water efficiency of other crops like maize, (ii) create a domestic natural wax source, and (iii) increase the economic viability of producing sorghum, which via its water and nitrogen use efficiency will decrease the risk of crop failures and increase food security.

I am particularly well-positioned to carry out this research as I recently completed my Ph.D. on wax biosynthesis. However, near the end of the program, I took a short course in bioinformatics and attended a conference that led to a realization: in order to make impactful technical advances and achieve my long-term goal, I need to (i) strongly link my (bio)chemistry skillset with basic genomics and bioinformatics and (ii) further develop my teaching abilities and my skills as an independent investigator.

In the past 11 months, I have gained experience with bioinformatic identification and heterologous expression of biosynthetic genes under the guidance of Dr. Edgar Cahoon (University of Nebraska – Lincoln; UNL). With this award, I aim to improve my technical skills further by (i) carrying out both conventional RNA-seq expression analysis and bulked segregant RNA-seq (BSR-seq) mapping projects of my own, (ii) becoming proficient in best-practice RNA-seq data analysis (Conesa et al., 2016), (iii) applying results to making impactful technical advances that can readily be translated into applications, and (iv) making the sequence data acquired available in a user-friendly way to other researchers.

I have also been working to develop my abilities as an independent investigator. I mentored a rotating graduate student whose small project was presented at two symposia and a conference at which it won the best poster award. I have also delivered two guest lectures and participated in grant writing and budget development workshops. As part of the proposed work I will hire and mentor two part-time undergraduate researchers as they carry out, document, and present their own small projects in support of the project proposed here. Such a project could be, for example, the profiling of kernel wax composition and yield from diverse lines in the Sorghum Association Panel that is grown at UNL each summer, the data from which could be used in future genome-wide association studies. Undergraduate assistants will also be asked to contribute one or more small guest articles to the plant chemistry blog I have been running for almost two years (>2000 views since Jan. 2016). I will also attend at least two laboratory or personnel management workshops offered by UNL. I will seek out additional opportunities to develop my teaching

abilities, such as delivering further guest lectures in, for example, the primary mentor's plant biochemistry course, and participate in two UNL teaching workshops or seminars, including those focused on developing and implementing new pedagogy methods in higher education. Finally, I will also write a review on the state of sorghum wax research and identify essential critical advances.

To help me attain these training and research goals I have assembled a *team of mentors* whose expertise spans the breadth of the work proposed. The primary mentor, Dr. Edgar Cahoon (U. Nebraska – Lincoln), is an authority on crop biochemistry and lipid biosynthesis gene discovery. My non-primary mentors, Drs. David Holding and Chi Zhang (both U. Nebraska – Lincoln), are experts in sorghum and maize functional genomics and bioinformatics, respectively. With their mentorship, I will not only receive guidance as I prepare to obtain a permanent position (i.e. teaching/research statement preparation, future grant proposal development, feedback on teaching and mentoring of undergraduates, expansion of my professional network, etc.) but also develop a unique skillset. Genomics and bioinformatics are seldom combined with expertise in plant chemistry to address challenges in agronomy, but this approach holds promise for sustaining and improving agriculture (e.g., Lee et al., 2014). In the proposed work, I will identify sorghum wax genes and build the bedrock of my future research career via two complimentary goals:

Research Goal 1: Obtain (i) whole tissue and epidermis-specific transcriptomes and (ii) comprehensive wax chemical profiles from kernels, leaf sheaths, and leaves of the sorghum reference line BTx623 before and after appearance of the visible wax phenotype. Correlate these data sets and use homology-based searching to identify sorghum wax gene candidates. Express select, high-confidence candidate genes in Arabidopsis wax mutants as a prerequisite to future complementation studies in sorghum.

Research Goal 2: Perform reciprocal crosses between five sorghum mutants lacking waxes on both sheaths and kernels (as putative transcription factor mutants) and the sorghum reference BTx623 line, then use BSR-seq in the resulting F2 mapping population to (i) obtain gene expression data from the mutant lines and (ii) map the causal mutations in the sorghum wax mutants.

Upon completion of this project, my *expectation* is that at least several genes involved in sorghum wax biosynthesis will be identified, and a high-quality list of additional sorghum wax gene candidates will be obtained. These outputs are prerequisites for understanding both how sorghum waxes contribute to high water-use efficiency and how to increase sorghum kernel wax amounts. Accordingly, the *positive impacts* of this award will be the enabling of investigations into how to further enhance and potentially transfer some of sorghum's water efficiency to other crops such as maize and the means to produce sorghum lines with increased kernel wax loads and thus substantially increased grain value (Fig. 2). Ultimately, this research will lay groundwork for improving our agricultural practices by minimizing their consumption of natural resources and making them resilient in the face of environmental challenges.

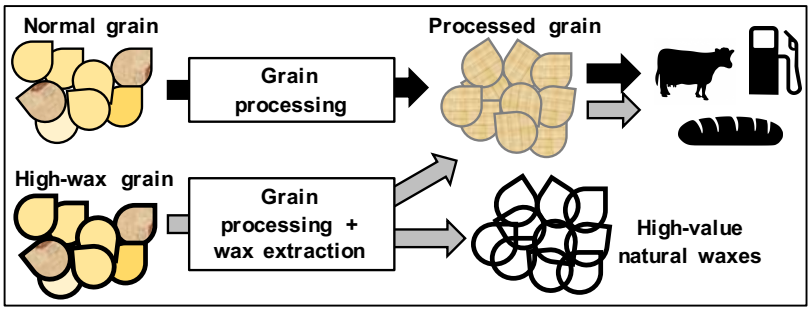

Figure 2: Sorghum grain is currently used for forage, fuel, and food (black path). High-wax sorghum lines could fulfill these same uses, but also increase grain value by making sorghum a convenient, domestic source of high-value natural waxes (grey path), which are currently all imported.

3. Experimental Approach to Research Goals

3.1 Research Goal 1: Sorghum wax genes via transcriptomics Arabidopsis complementation: Ten greenhouse-grown BTx623 sorghum plants will be used to prepare replicate ($n \geq 3$) sets of wax and RNA samples from leaves, leaf sheaths, and kernels from both before and after the developmental stage at which sheath waxes appear. At each time point replicate whole-tissue and epidermis-specific RNA will be prepared. Wax samples will be analyzed with gas chromatography-mass spectrometry. I have extensive experience with wax analyses (Busta et al., 2016a, 2016b; Busta and Jetter, 2017). The RNA samples will be sent to a commercial sequencing center. Gene expression levels will be obtained by quantifying reads against the *Sorghum bicolor* reference transcriptome using the alignment-free RNA quantification tool Salmon (Patro et al., 2017). I am an experienced user of this program. I will then identify candidate wax biosynthesis genes by searching for sorghum genes with homology to known wax biosynthesis from Arabidopsis and/or maize that either (i) are upregulated in the epidermis, and/or (ii) have expression levels that correlate with variation in specific wax compound abundance on different organs or at different developmental stages, as determined by a linear regression model.

As a prerequisite to future complementation tests in sorghum, high-confidence candidate sorghum wax biosynthesis genes will be functionally tested by expressing them in Arabidopsis wax mutants. With input from other sorghum breeders/researchers at UNL, candidate transcription factors (as regulators of wax yield) or enzyme-coding genes (as controllers of wax composition) will be prioritized. Sorghum RNA will be reverse transcribed into a cDNA template from which each top candidate will be cloned into an expression vector containing the DsRed fluorescent marker and transferred to Arabidopsis using *Agrobacterium tumefaciens*. Transformed seeds, identified by their red fluorescence, will be raised in a growth chamber, genotyped by PCR, and their waxes will be analyzed to evaluate complementation. I recently used a similar approach to characterize lipid biosynthesis genes from other crop species.

A potential pitfall of this aim is that functional sorghum wax genes may not compliment corresponding Arabidopsis mutants, thus generating false negatives in the proposed complementation tests. Based on the ability of other crops' genes to complement Arabidopsis wax mutants, we think this is unlikely. However, in the event that none of the top candidates test positive, structural wax gene candidates will be expressed in yeast engineered to produce precursors for wax biosynthetic pathways. Even though via the yeast approach I will not be able to test putative transcription factors, the comparative transcriptomics/wax profiling results will still be indispensable in future sorghum wax mutant complementation, future sorghum knockout lines created to test gene candidates, and in the mapping approach described below.

3.2 Research Goal 2: Sorghum wax genes via bulked segregant RNA sequencing: For species with complex genomes like sorghum, mutations are currently efficiently mapped using bulked segregant RNA sequencing (BSR-seq; Li et al., 2013, Liu et al., 2012). Dr. Gebisa Ejeta has made available to this project the 31 non-allelic sorghum wax EMS mutants developed in his lab (Peters et al., 2009). I will first grow all the mutant lines, eliminate from consideration any that are not affected both in leaf and kernel wax, then of the remaining I will select eight for further study. These eight mutants and the reference BTx623 line will be grown in the greenhouse. In one planting, two plants of each mutant line and ten BTx623 plants will be started. Three plantings will be made, each a week apart, to ensure sufficient synchronicity in flowering times between the lines for successful crosses. Reciprocal crosses will be performed between the mutants and the reference plants, some panicles on each line will be allowed to self-pollinate, and comprehensive wax analyses of each mutant will be performed. Seeds from crosses (F1 seeds) and self-pollinating events will be collected. Five F1 plants from each cross will be grown in the greenhouse, their flowers covered with bags to ensure self-pollination, and resulting F2 seeds will be collected. I have not crossed sorghum before, but have done so with Arabidopsis, and such crosses are routine in the Holding lab; we do not anticipate difficulties generating the F2 seeds.

Next, 200 F2 plants from each cross will be grown in the field. When waxes are visible (~6 weeks of age), I will make one collection of leaf punches from all plants arising from one cross that lack wax and another collection from plants from that same cross that are waxy, resulting in pools of leaf punches from at least 50 mutant and 50 non-mutant plants arising from one cross. Such collections will be made from the offspring of each cross, then RNA will be extracted from each of the eight mutant and eight non-mutant pools and sent to a commercial sequencing center. Reads will be analyzed using the Holding/Zhang lab BSR-seq mapping pipeline: Samtools, Bowtie to map reads to reference transcriptome, VarScan for variant calling, and R for mapping interval visualization (Jia et al., 2016). I have not used this particular pipeline, but have experience with R (Busta et al., 2016b) and Bowtie. We expect to obtain a relatively small mapping interval as high-quality, high-polymorphic single nucleotide polymorphisms can be detected in sorghum even with low-coverage sequencing (Bekele et al., 2013). The BSR-seq analysis, though using only one mutant and non-mutant RNA-seq sample per mutant, will still yield extensive, relevant gene expression data. The causal locus itself will likely be down regulated, so the expression level of each coding sequence in the interval will be determined, expression levels of high-confidence wax gene candidates in mutant and non-mutant RNA pools verified with qRT-PCR, and the effects (e.g. stop codon insertion, non-synonymous substitutions, etc.) of mutations near candidate genes will be assessed. Together, all this information will be used to identify the causal mutation.

A *potential pitfall* of this second specific aim is that the mutants were created in a background that is not the BTx623 reference line, and, though I will cross the mutant with this line, it is possible that issues with read mapping and noisy mapping intervals may arise. Nevertheless, with the expertise and experience of my co-mentors, we expect such a challenge to be manageable. It is also possible that, since I have not used BSR-seq previously, the data analysis could take longer than anticipated. However, I have two non-primary mentors that are experts in this technique, and a sufficient portion of the Project Timetable has been allocated for the analysis. We are confident that this specific aim can be achieved and the technical skills I develop in doing so will enable me to use this technique independently in the future.

4. Broader Impacts - As part of this fellowship I will work to make positive impacts on (i) other local, young researchers in my field, (ii) individuals in the community of Lincoln, Nebraska, (iii) female high school students in greater Nebraska, and (iv) other researchers working on sorghum waxes worldwide.

First, I will lead a journal club focused on crop improvement for UNL students and postdocs that will meet twice a month during summers. I will recruit members, coordinate the events, and initiate each session. Next, together with the undergraduates I hire, we will promote and communicate our work to the public through two venues. In one, using some of the sorghum we grow as props, we will develop an educational module to present at the UNL Science Museum's Sunday with a Scientist event, which brings together one team of scientists and, on average, 100 children and parents for 4 hours. In the module, we will use hands-on mini experiments as an active learning technique to educate and stimulate discussion among children and parents about plant chemistry, diversity, and crop improvement, and foster children's interest in STEM fields. I recently helped to coordinate a similar event to teach museum-goers about plant root chemistry. In addition, we will work to spark interest in plant science among the high school girls that come to UNL for the annual Women in Science Conference by developing a 2-hour hands-on laboratory workshop featuring plant chemistry, plant diversity, and crop improvement. I participated in a similar event earlier this year to teach students about plant pigment chemistry. After each of these sessions, participants will complete a survey we will use to assess and improve the module or workshop so it can be modified and used in years to come for further public engagement.

Finally, in addition to sharing the raw sequence data collected via NCBI's Sequence Read Archive, I will create an R Shiny app for online, rapid visualization of sorghum epidermis-specific gene expression (the Sorghum Epidermal Gene Almanac. I am an advanced user of R and have experience creating Shiny apps for data visualization. Users will be able to enter query sequences which will then be BLASTed against

the sorghum transcriptome. BLAST hits will be displayed as a phylogenetic tree that is annotated with the expression level of each gene in whole tissues (leaf, leaf sheath, and developing kernel), alongside the epidermis-specific expression level of each gene in these tissues. This tool will catalyze further research and facilitate the rapid discovery of additional genes controlling sorghum wax composition and yield.

5. Coordination of Research and Project Timeline – To track and evaluate progress towards the training and research objectives detailed in the previous sections, I will consult my mentors and the UNL Office of Postdoctoral Studies in the development of an Individual Development Plan (IDP) during the first month of the fellowship. It will list the (i) workshops I will attend to develop my communication, leadership, and writing skills; (ii) outreach activities in which I will coordinate during the fellowship; (iii) dates for guest lectures I will deliver; and (iv) conferences I will attend. The IDP will also list milestones for the training and research goals of the fellowship. The IDP will serve as a master document for bimonthly meetings of all mentors to assess the efficacy of training and research activities and make adjustments the plan as needed. Assessments may also be conducted via observed guest lectures, formal testing of sequence data analysis skills, and via presentations/feedback at stakeholder meetings. Once per year and the fellowship's end, UNL's Office of Postdoctoral Studies will provide an outside evaluation of the efficacy of the IDP in keeping the project on track and achieving project goals.

Month ->	1,2 3,4 5,6 7,8 9,10 11,12	13,14 15,16 17,18	19,20 21,22 23,24 25,26	27,28 29,30 31,32 33,34 35,36
Training Objectives	Communication and teaching workshops (2 & 2), Guest lectures (2), Journal Club leader (1), Extension activities (2), PD meeting (1), Review writing (1), Undergraduate mentee (1)		Communication, teaching workshops (2 & 2), Guest lectures (2), Journal Club (1), Extension events (2), Undergraduate mentee (1+), Conference (1+), Grant writing (2), Interview prep	
Research Aim 1: transcriptomics/ wax profiling	Grow sorghum, prepare samples; Analyze wax and RNA data, clone candidate genes	Grow/ transform Arabidopsis	Grow and analyze Arabidopsis transformants	Data consolidation /archival, figure / manuscript preparation and submission, Sorghum Epidermal Gene Almanac creation
Research Aim 2: mutant mapping	Grow sorghum, make crosses; Grow F1, self pollinate	Grow F2, bulk segregants, prep RNA	Analysis of BSR-seq data, qRT- PCR, coding sequence analysis	Data consolidation /archival, figure / manuscript preparation and submission

Figure 3: Project Timetable.

6. Expected Outcomes - We expect that the two approaches outlined above (Research Goals 1 and 2) will lead to the identification of several sorghum wax genes and a list of high-quality wax gene candidates. This project is also expected to:

- **Advance agriculture** by providing publically-available tools to investigate how to transfer some of sorghum's water efficiency to other crops.
- **Increase sorghum harvest value** by providing markers for breeding programs that target high-wax or wax composition-optimized sorghum lines.
- **Improve food security** by increasing the revenue obtained from producing sorghum, a more drought-tolerant crop.
- **Catalyze future crop improvement** by making public a Sorghum Epidermis Gene Expression Almanac to inspire and facilitate future research projects.
- **Recruit new plant scientists** from chemistry backgrounds, train them in phytochemical analysis, involve them in teaching and public outreach, and help them develop professional networks.
- **Engage children and women** in hands-on activities emphasizing the economic, environmental, and health benefits gained through the study of plant chemistry, diversity, and crop improvement.
- **Train me as a unique PI** with diverse skills and a network of expert collaborators so I can obtain a permanent position from which to lead projects and train students to combine chemical analysis with molecular biology, genomics, and bioinformatics to study economically important plant product