

DNA and Protein Analysis throughout the Industrial Refining Process of Sugar Cane

C. A. Cullis^{1*}, A. L. Contento² and Mark Schell^{1,3}

¹ Department of Biology, Case Western Reserve University, Cleveland, Ohio, USA

² Department of Biological Sciences, State University of New York at Oswego, Oswego, New York, USA

³ Professor Emeritus, Department of Microbiology, University of Georgia, Athens GA, USA

Abstract

The amount, nature, and fate of DNA and protein in the major purification fractions generated during industrial scale refining of sugar cane into raw sugar by the diffuser and tandem roller mills was determined. The presence and size of sugar cane DNA were estimated using PCR and sugar cane specific primers that amplified fragments of various sizes from different segments of the repetitive intergenic region (IGS) of the 25S rDNA. Both the maximum fragment size capable of amplification and the amount of DNA decreased as refining progressed, indicating sequential degradation during the milling. However, PCR still detected minute quantities of sugar cane DNA in raw sugar (<10e-3 ppm). Using a bicinchoninic acid assay on trichloroacetic acid precipitated sodium dodecylsulfate-extracts, protein was found in all mill fractions and decreased from 4500 to 10 ppm as sugar cane was refined to raw sugar. Analysis of these extracts by one- and two-dimensional gel electrophoresis suggested a gradual degradation of proteins during refining. Shotgun proteomic analyses identified complex populations of sugar cane proteins, or peptides thereof, in all mill fractions, but the population complexity decreased during processing. Retail-purchased refined cane sugar showed no detectable protein or DNA (<2 ppm protein; 0.001 ppm DNA).

Keywords: *Sugarcane, Refining Fractions, DNA, Protein, Detection, Proteomics*

Introduction

Crystalline sugar is manufactured from sugar cane by a multistep extraction and purification procedure using either a tandem roller mill or diffuser mill (Figure 1). Cut cane pieces (CP) are first shredded, imbibed with water and then crushed between sets of rollers to release the primary juice (PJ) (tandem mill); alternatively shredded cane is extensively lixiviated (rinsed and percolated with recycling ~80°C water) to obtain the primary juice (diffuser mill). Cane residue

(bagasse, BG) is dried and used as boiler fuel, animal feed, or in paper manufacture. Primary juice is filtered and then clarified by heating in the presence of lime to precipitate impurities, especially protein. After filtration to remove the precipitate called filter cake (FC), the clarified juice (CJ) is concentrated by evaporation resulting in precipitation of raw sugar crystals (RW). The evaporation/crystallization process is repeated until no more sucrose crystallizes.

* Corresponding author: cac5@case.edu

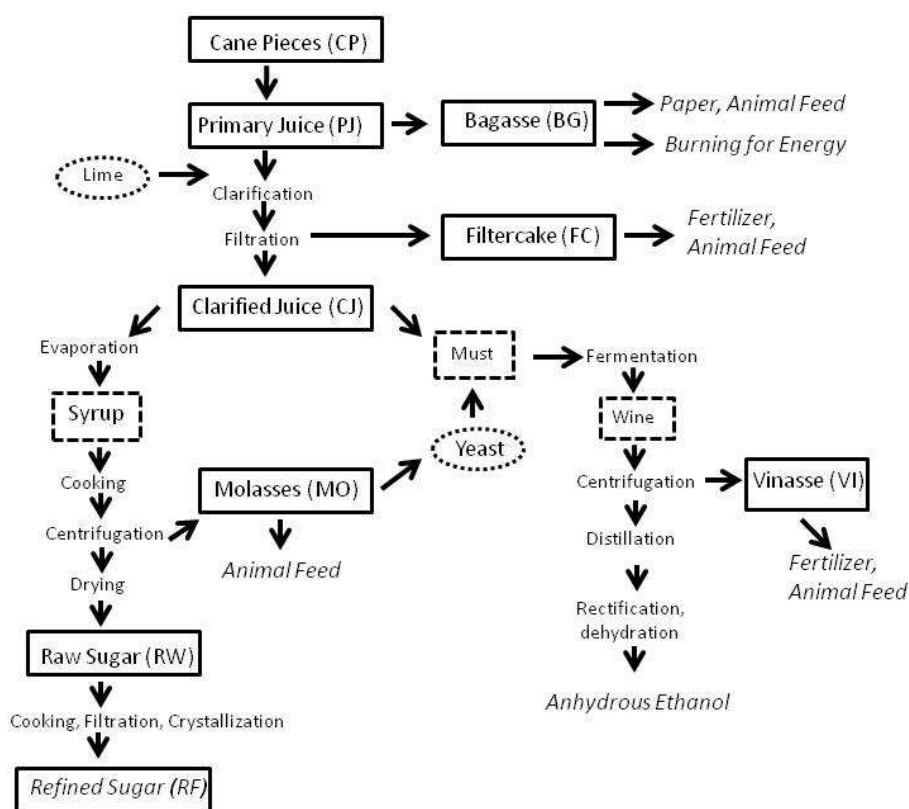


Figure 1: Processing of sugar cane into various end products. Major mill fractions analyzed in this work are in solid boxes. End products are italicized; intermediate products are in dashed boxes. Exogenous additions to fractions are in ellipses.

The residual liquid called molasses (MO) is mixed with yeast and fermented to produce ethanol. After recovery of the ethanol the residual fermentation solids are removed by centrifugation to yield vinasse (VI) which can be used as fertilizer or animal feed.

The objective of this study was to determine the fate and concentrations of sugar cane DNA and proteins throughout the industrial refining process of sampling eight different fractions (Figure 1). The fate of nucleic acid and proteins during sugar manufacture from sugar beet and sugar cane has been published (Klein et al., 1998; Oguchi et al., 2009; Taylor et al., 1999; Joyce et al., 2013; Potter et al., 1990). These reports focused on the detection of specific fragments of the foreign DNA present in the transgenic plants used as starting materials. Here we report the development of robust and highly sensitive methods that enable detection and sizing of the total DNA and protein of conventional sugar cane throughout the refining

process. Methods were tested on eight mill fractions from two commercial sugar cane mills in Brazil, one using the newer diffuser extraction process and one using the more traditional tandem roller mill process. The methods were shown to be rapid and very sensitive, routinely capable of detecting protein and DNA in very concentrated, dirty, and complex samples above the 5 and 0.001 ppm levels, respectively. Results showed DNA and protein are present in all mill fractions; however the amount and size significantly decrease as the cane was refined to raw sugar.

Materials and Methods

Sample Acquisition and Preliminary Processing

Samples were collected at two different industrial sugar cane mills in Brazil: a diffuser type ("mill V") and a tandem roller type ("mill G"). Eight sample types were collected in triplicate from each mill: CP, BG, FC, PJ, CJ, MO, VI and RW (Figure

1). CP samples were collected in plastic bags and immediately placed on dry ice. All other samples were collected in polypropylene bottles and placed immediately on dry ice. Samples were then stored at -80°C until shipment on dry ice to the US. Upon arrival CP, BG, and FC samples were lyophilized for four to five days, resulting in a moisture content of <5%. FC samples were placed at -80°C, while CP and BG samples were ground to a homogeneous powder in a GM200 Grindomix (Retsch) before storage at -80°C. CJ, PJ, MO, and VI samples were not processed before analysis. Sugar cane leaf samples were harvested from greenhouse-grown plants (varieties CP-89-2143 and SP-70-1143), placed immediately on dry ice, and then ground to a homogeneous powder as above before lyophilization for three days and stored at -80°C.

DNA Isolation

DNA was isolated from samples using Plant DNA miniprep or maxiprep kits (Qiagen) as per the manufacturer's instructions except where noted. When the miniprep kit was used, 100 mg samples were ground in a mortar and pestle with 400 µL of extraction buffer, except for the BG and FC samples, where 1200 µL of buffers were used and volumes of all other reagents were increased proportionally. When the maxiprep kit was used 1g of material was extracted. The final elution volume was 200 µL for the miniprep extractions and 1 ml for the maxiprep extractions. A 600 µL aliquot from each maxiprep extraction was then reduced to 20 µL by vacuum reduction at room temperature. DNA isolations were independently performed from each mill fraction at least three times.

Total DNA Quantitation

DNA concentrations were determined using both spectrophotometric and fluometric methods. Purified DNA samples were measured at 280 nm, 260 nm, and 230 nm (NanoDrop 8000; Thermo-Fisher) and in an Applied Biosystems Step-One qPCR system using SYBR green. DNA samples from each extract were also analyzed on 1% agarose gels with TBE buffer and visualized under ultraviolet light (UV) following ethidium bromide staining.

PCR Primers

The intergenic spacer region of the sugar cane

rDNA repeat unit was amplified using a pair of primers designed from the sequence of the conserved 18S and 25S coding regions (25SCONS and 18SCONS; Table 1) and purified sugar cane DNA as the target. The amplicon was purified by agarose gel electrophoresis and sequenced (Supplementary Figure 1S). The sequence was analyzed using BLASTN at NCBI and the region that was specific to sugar cane and other related monocots (e.g. *Sorghum bicolor*, *Zea mays*) identified. Primers were designed to specifically amplify fragments of different sizes from the sugar cane intergenic spacer region (IGS; Figure 2).

The sequences of the primers are given in Table 1. The primer pairs and the expected fragment sizes are: 4L/4R – 273 bp, 4L/5R – 552 bp, 7L/7R – 800 bp, 2L/3R – 1760 bp.

Table 1

Primer Sequences

Primer Name	Primer Sequence (5' – 3')
25SCONS	AAAACACCAGCTCCCAAC
18SCONS	AAGGCAACACGTAACCGAAC
Cane2L	AGTGAAATCACCGGGCATC
Cane3L	GGTTACGTGTTGCCCTTCGTC
Cane4L	AAAAGCACAAGCGAGGTGTC
Cane4R	TTTGGACCTGTGACTGCTTG
Cane5R	GTTGGGATTGTTTCGGTGAC
Cane7L	ACCAAAAGCACAAGCGAGGT
Cane7R	GCCTAGCTGGGAGGGCTA

PCR Method

PCR reactions were performed in 25 µL with variable amounts of template, up to a maximum of 50 ng per reaction with primer concentrations of 0.2 µM. When DNA concentrations were lower than 5ng/ µL, 10 µL of the DNA preparation was used. The amplification program used an initial denaturation step at 98°C for 2 min, followed by 35 cycles of denaturation at 94°C for 15 sec, annealing at 60°C for 30 sec and extension at 72°C for 2 min. Two different polymerases, GoTaq (Promega) and Speedstar (Clontech) (1.25 units each) were added to ensure that fragments of different sizes could be amplified. Following amplification, PCR products were visualized on 1.5% agarose gels run in 0.5X TBE buffer, stained with ethidium bromide, and photographed under UV.

Sugar cane sequence of amplified bands from conserved region of 25S ribosomal RNA into the intergenic spacer region.

```
GTGGCGCAAGCTACCGTGTGCCGATTATGACTGAACGCCTCTAAGTCAGAATCCAAGCTAGCAACCGGCGCCTGTGCC
CGCCGCTCGCCCCGACCCACATTAGGGCGTTCGCGCCCCAAGGGCCCGTGCCTACTGGCTCAGCCCGCCCGGCCGACGCG
CCGCGGCGGGCCCCCTCGAAGCTCCCTTCCCAACGGGCGGCGGGCTGAATCCTTTGCAGACGACTTAAACGCGACGGG
GCATTGTAAGTGGCAGAGTGGCCTTGTGCCACGATCCACTGAGATCCAGCCCTGCGTCGCACGGATTCTCCCTCCCCC
CTCCACGCCCCCGGGCCGAGTCAGGCTCCACCCCGCGTGGCCGTGGCCGACCCTGCCGAACGGCACCGCGCTCGTCTC
CGGCCAAAGGCAAGGCAGCCANAAATGAAATTGGCTAAGTGAATCACCGGGCATCCGCTCCGTGCGCCCTGGGCAAAA
ACACCAGCTCCCCAACGCGTCCGGGGCCAAAAACACCAGTCT
GGCCACATGGTCCGTGAGCANCCCAAGTCCCCGAAAGCCGGCTTTCTACAGTGACAGATATGCGCCCCGGGGCAGTTC
TCCCANGCGGCANGGCTGAAAACNNCGTTCTCCANGCGCCNGAGCTGAGACACANCGNTNNNCGGACTGGCCAANTC
CCCCNGGCCGACGGCACCCGGTGGGGGAAACANCGTTCTCCCTGGCGCCCGGGCCNAAAACANNAGTTCTCNCANG
NNNTCTNCGGAAGACAGNCAAGTCCCCGGGCGGACTGCACCCACGGGGCGNAAACAGCGGGNTCNCCTGGCGCCANG
GCCNAAACNNNNNTTNGCCANGCGCCGGCGCTGAGACNCANCGNTTCGGCAGGNGGTNCNCNGAAGACAGNCNAGTCC
CCGGNCGANNNGCCNCGGGGNGGAAACAGCTGTTCCNCTGGCGNCAGGGCTGAAACAGCAGT
```

Sugar cane sequence of amplified bands from conserved region of 18S ribosomal RNA into the intergenic spacer region.

```
GATTGGGTAATTTGCGCGCCTGCTGCCTTCCTTGGATGTGGTAGCCGTTTCTCAGGCTCCCTCTCCGGAATCGAACCTAAT
TCTCCGTACCCGTCACCACCATGGTAGGCCCTATCCTACCATCGAAAGTTGATAGGGCAGAAATTTGAATGATGCGTCGC
CGGCACGAGGGCCGTGCGATCCGTCAAGTTATCATGAATCATCGGATCAGCGAGCAGAGCCCGCGTCAGCCTTTATCTAA
TAAATGCGCCCCCTCCCGGAAGTCGGGGTTTGTTCACGTATTAGCTCTAGAATTACTACGGTTATCCGAGTAGCACGTACCA
TCAAACAACTATAACTGATTTAATGAGCCATTGCGAGTTTCACAGTTCGAATTAGTTGACATGCGATGGCTTAATC
TTTGAGACAAGCATATGACTACTGGCAGGATCAACCAGGTAGCACGTCTCGCCGACGAGCCGGCGCCGGCCTCCGCACG
AGGCATCGGCCGGGCTGGCTGTCGTTGTTTCAAACGGACTCGGTTTCTAGGCACGGGAGGCCAACGCCCTCCGGCCTTC
GGCATCAGCCGCATCCGCAACCAAAAGCACAAGCGAGGTGTCCTTGGCGGTGCCGCCCTAGGAGGACGACATCACGAGGC
AACGCCGAAGTGCTCTAGAGCCGACGAGCCACACCCAGAGGGTGCCTCGACGAAGGCAACACGTAACCGAACGCAACT
TCCGTGGGACAGGTAGCAGCAGCAAGCACTCGTCAACGCAGCAGGCAAAAGGTGCCCGCACGAATGAAGGGACGCTG
GCGCTAGGAGTCAGCCGACGCCGGCGGGGGTCTCCAAGCAGTCACAGGTCCAAAACAACATGCGCCAGCGTAGCCA
CGACGGTTAAGCCATCCAAAGCATCCCTCCGCGTAGGCACGGTGAGTCTGCTTGCAGGACGGCAACCGAGGGTCCACA
GAGCGCGAGAGAAACCGAACTTATCCAACAGCGGGCCAAAACCGTGACGCCGAAACCGTTCCAGCAACAAGATTCTCCC
ATGGGAGCAGTGCCCGCAGCAGCCCGCAGAAAGCAGGCCACACAGGCTGCCGGCCGTCCAAGAAGGCCGCGAGTCAC
CGAAACAATCCCAACAACACGTTGGCCACCCAGCCAGCCAGCTACCCACAGCATCCCGCCAAAGCAGGCCGCGCAGGC
TTCCGCCCGGCCAAGAAGGCGGCGCATCATAAAACAGTTCCAACAACAAGCTAGCCACCCTAGGCAGCCAGCTGCCAAA
GCAGCCCGCCGAAGCAGGCCACCGCAGGCTGCCGACCGTCCAAGGAGGCCGCAAGTCACCGAATCCGTTCCAACAACAC
GCTAGCCCTCCAGCTAGGCAGTCGCCCCGAGCAGCCACCCAAGCAGGCCGCGCTAGGCAGCCGCCACCCNGGAGGC
CGGGAGTCGTTCCGACACCACGCTGGCCACCGAGCCAGCCCGCGGCCCATAGCAGCCG
```

Sequence of the fragment amplified from raw sugar using primers 4L and 4R

```
AAAAGCACAAGCGAGGTGTCTTGGCGGTGCCACCCTAGGAGGACGACATCACGAGGCAACGCCGCAAGTGCTCTAGAGC
CGACGAGCCACACCCAGAGGGTGCCTCGACGAAGGCAACACGTAACCGAACGCAACTTCCCGTGGGACAGGTAGCAGCA
CGCAAGCACTCGTCAACGCAGCAGGCAAAAGGTGCCCGCAGCAATGAAGGGACGCAGGCGCTAGGATTCAGCCGCACGC
CGGCGGGGGTCTCCAAGCAGTCACAGGTCCAAA
```

Supplementary Figure 1S: Single pass sequences of PCR amplified fragments of the sugar cane large ribosomal RNA intergenic spacer regions used for designing primers:

Protein Extraction Method

Four mL of 3% (w/v) solutions of CP, BG, FC and VI were prepared in 15-mL polypropylene tubes (Sarstedt) and adjusted to 1% sodium dodecyl-sulfate (SDS) +10 mM dithiothreitol (DTT) + 10 mM Tris-HCl pH 7.5 (SDS-extraction buffer, SEB). Similarly 15-20% solutions of PJ, CJ, MO, RW, or retail-purchased refined sugar (RF) were prepared and identically adjusted. All samples

were heated at 65°C for 60 min with occasional mixing by inversion. Tubes were centrifuged for 15 min at 25°C and 6500 x *g* in a swinging bucket rotor. CJ-containing samples were not centrifuged as some SDS precipitated at 25°C but not 65°C. Three ml of 1% sodium deoxycholate (DOC) was mixed into 2mL of each supernatant followed by 1.25 ml of 50% trichloroacetic acid (TCA). After mixing well and sitting 15 min on ice, the tubes

were centrifuged at 6500 x g for 15 min at 7°C and drained for 5 min. Then, 1.5 ml of acetone was added, the tubes were vortexed for 15 sec, and incubated 25°C for 15 min with an occasional vortexing. If necessary the tubes were sonicated for 1 min in a Branson 1510. Samples were placed on ice for 10 min, centrifuged, supernatants removed, and the tubes drained. Then 1.5 ml of 85% acetone was added and the tubes were vortexed, centrifuged, drained and dried at 37°C for 15 min. The precipitate was dissolved in 0.2-0.5 ml of 0.5 % SDS + 10 mM Tris-HCl 7.5 at 65°C for 20 min with occasional vortexing and/or sonication. For preparative SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins extracted from CJ, MO, RW and RF the procedure was scaled up seven-fold. Protein content was determined using the microplate Micro BCA protein assay (Thermo Fisher) as recommended by the manufacturer. Protein at concentrations as low as 4-8 µg/ml of mill fraction could be detected with this protocol. The Overall recovery of protein from fractions with low protein concentrations (CJ, MO, RW, RF) was >50% as determined by analyzing mill fractions spiked with known amounts of protein added (7 µg /ml each of bovine serum albumin and lysozyme).

In some preliminary experiments, proteins in mill samples were extracted with LBT (7 M urea + 2 M thiourea + 12 mM Tris-HCl 7.5 + 4% CHAPS + 12 mM DTT) (Amalraj et al., 2010), precipitated with DOC/TCA, washed and redissolved in the same fashion as when using SEB. A P-PER kit (Pierce) was also used for additional preliminary extractions, according to manufacturer's instruction.

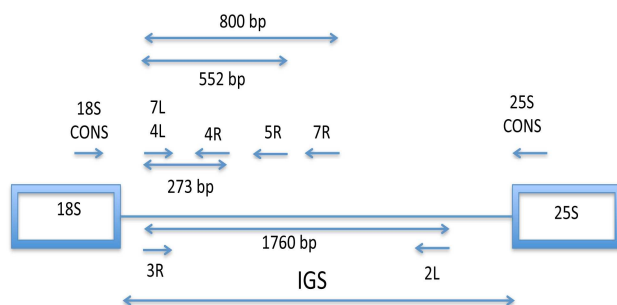


Figure 2: Position of primers within the rDNA repeat unit. 25S CONS and 18S CONS are conserved primers within the 25S and 18S coding regions, respectively. The remaining primers are within the Intergenic Spacer Region (IGS).

Gel Electrophoresis

Analytical and preparative SDS-PAGE was performed using 4-20% gradient gels (Thermo Scientific 0025244). For analytical-scale samples, 10-60 µg of protein in 15 µL, were adjusted to 2% SDS + 2% beta-mercaptoethanol + 15% glycerol + 0.05% bromophenol blue and boiled for 5 min before loading. The gels were run at 50V, treated with BioSafe Coomassie Stain as per manufacturer's instructions and photographed with a digital camera. Peptide mass fingerprinting to identify single protein bands excised from analytical gels was performed essentially as described by Schell et al. (2011).

Preparative SDS-PAGE was run analogously, except 150 µg of protein was loaded on the gel; for RW, 35 µg was loaded. Electrophoresis was carried out until the bromophenol blue dye had migrated one-third of the way down the gel. The stained portion of each lane was excised and cut in half producing a high-molecular mass containing a slice (50-250 kDa) and low-molecular mass (6-50 kDa) containing slice.

Samples for two dimensional gel electrophoresis (2-DGE) were prepared by SEB extraction and DOC/TCA precipitation as described above. Samples were processed using a ReadyPrep 2D Cleanup Kit (BioRad) followed by dissolving in Ready-Prep Rehydration Buffer (BioRad). Samples containing 100 µg of protein were applied to 11 cm pH 4-7 Ready-Strip IPG Strips (BioRad) for 12 hr. IEF was performed using the BioRad Protean IEF cell at 20°C with the following program: 250V with a linear slope for 20 min; 8000V with linear slope for 150 min; and 8000V with a rapid ramp for 20000 V-hr). IEF the strips were processed as recommended by the manufacturer. The strips were applied to 1.0 mm Criterion precast 12.5% gel (BioRad) and run at 200 V for 1 h. Spectra Multicolor Broad Range molecular mass standards (Thermo Scientific) were run adjacent to each strip. Gels were treated with BioSafe Coomassie stain as above and then scanned using the Amersham BioSciences image scanner and the associated 2D Platinum software used to view, analyze, and compare the images. Spots were identified using intensity >4000, area >2, volume >6000, and saliency >200.

Mass Spectrometry

Shotgun proteomic analyses (liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)) were performed on a Thermo-Fisher LTQ Orbitrap Elite Mass Spectrometer coupled with a Proxeon Easy NanoLC system. Tryptic peptides were derived from the high and low molecular mass preparative SDS-PAGE gel slices essentially as described by Schell et al. (2011), except reagent volumes were increased 3 fold. Extracted peptides ($\sim 0.25 \mu\text{g}$) were first passed through a PepMap 5 mm x 300 μm precolumn (Dionex). The eluate was loaded on a self-packed 12 cm x 100 μm column/emitter containing 200 Å 5 μM Bruker MagicAQ C18; peptides were eluted with a 500 μL gradient of solution A (0.1% formic acid in acetonitrile) and solution B (0.1% formic acid). The gradient started with 5% B, increased to 40% B in 40 min, increased to 60% B in 15 min, and increased to 95% B in 10 min. The Top 10 data-dependent acquisition method was used to acquire data. Proteome Discoverer 1.3 (Thermo-Fisher) software was used by Mascot database search program (Matrix Science) for protein identifications. Databases searched were: 1) a set of 121,000 expressed sequence tags (ESTs) from sugar cane (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s_officinarum) translated in all six possible reading frames); 2) a *Sorghum bicolor* proteome predicted from its complete genome sequence (<http://www.ncbi.nlm.nih.gov/genome/108>); and 3) the NCBI nr protein database of all nonredundant predicted protein sequences. We defined positive protein identifications as those which had Mascot scores of >70 based on data from at least 2 unique peptides; at this stringency identifications are at the $>99\%$ confidence level.

Results

Partial Sequence of the Sugar Cane Intergenic Spacer Region

Primer pair 25SCONS/18SCONS was designed from conserved regions of the plant 25S and 18S ribosomal RNA genes, and used to amplify purified sugar cane DNA isolated from sugar cane leaves. The amplified fragments were gel-purified, sequenced, and then primers specific to the sugar cane intergenic spacer region (IGS) near either the 18S or 25S coding regions were designed. The

primers were tested on sugar cane leaf DNA, and in all cases the expected fragment size was observed. The new primers 2L and 3R were used with sugar cane DNA and the resulting fragment isolated and sequenced. This sequence was used to design additional primers that amplified a range of fragments of different sizes from the IGS. The primer sequences are in Table 1. The positions of the various primers and the sizes of the fragments amplified are shown in Figure 2.

DNA Quantitation

The quantity of DNA isolated from the various mill fractions was determined by three different methods. The amounts measured by UV spectroscopy decreased from 1.3 ppm in CP to 0.3 ppm in PJ, while BG and FC contained amounts similar to those found in CP (Table 2). The amount of DNA in the subsequent processing fractions CJ, MO and RW was below the limit of detection even when using the most sensitive (SYBR green) method. Using agarose gel electrophoresis, DNA was detected only in the CP, BG and FC fractions (Figure 3). The DNA from the BG and FC fractions was degraded into much smaller fragments than the DNA from the CP, an observation that was supported by the amplification products observed in PCR reactions discussed below.

Table 2

Amount of DNA in Processing Fractions of Mills G and V

Sample	DNA in Eluate ng/ μL	Total μg DNA Recovered per g Starting Material (ppm)
PJ-G	$0.003 \pm .002$	0.3 ± 0.2
PJ-V	$0.003 \pm .003$	0.3 ± 0.3
FC-G	11 ± 4	1.1 ± 0.4
FC-V	14 ± 3	1.4 ± 0.3
BG-G	10 ± 3	1.0 ± 0.3
BG-V	15 ± 4	1.5 ± 0.4
CP-G	11 ± 5	1.1 ± 0.5
CP-V	13 ± 5	1.3 ± 0.5

For samples CJ, MO, VI, and RW from both mills, as well as retail-purchased refined sugar, DNA was below the limit of detection with SYBR green: $< 0.002 \text{ pg}/\mu\text{L}$; $< 0.4 \text{ ng/g}$; 0.0004 ppm .

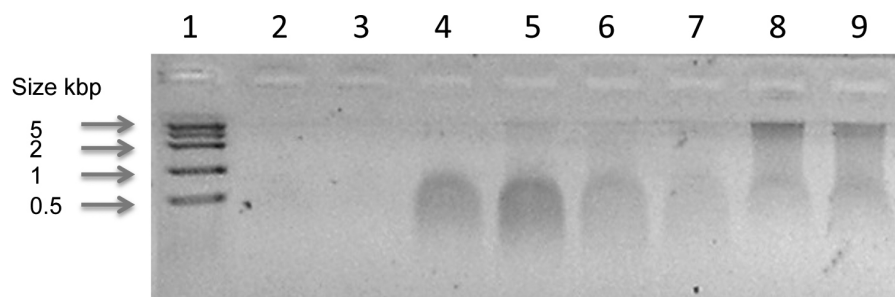


Figure 3: Agarose gel electrophoresis of DNA extracted from mill processing fractions. 15 μ L of DNA extract was loaded. Lanes: 1, molecular weight marker II; 2, PJ-G; 3, PJ-V; 4, FC-G; 5, FC-V; 6, BG-G; 7, BG-V; 8, CP-G; 9, CP-V.

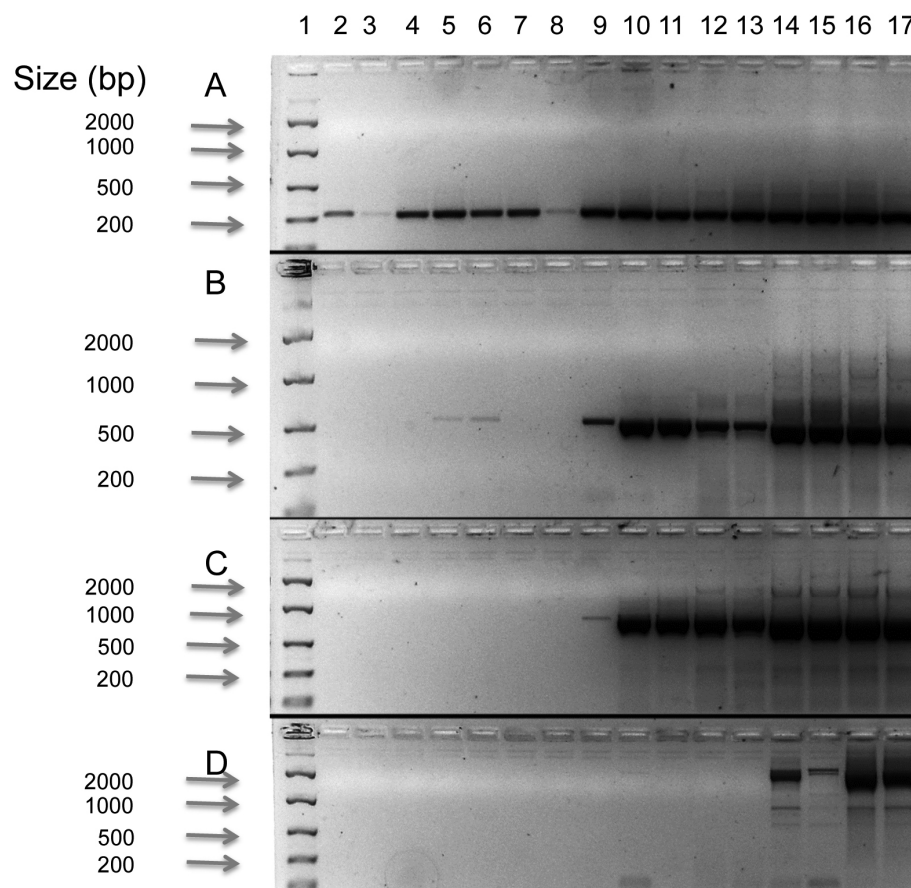


Figure 4: PCR amplifications of DNA extracted from mill processing fractions. Lanes on the 1.5% agarose gel show a representative PCR reaction from each fraction from mill G and mill V. Lanes: 1, molecular weight marker I; 2, RW-G; 3, RW-V; 4, MO-G; 5, MO-V; 6, VI-G; 7, VI-V; 8, CJ-G; 9, CJ-V; 10, PJ-G; 11, PJ-V; 12, FC-G; 13, FC-V; 14, BG-G; 15, BG-V; 16, CP-V; 17, CP-V. Panel A: Primers 4L/4R; Panel B: Primers 4L/5R; Panel C: Primers 7L/7R; Panel D: Primers 2L/3R.

Detection of Sugar Cane DNA Sequences in Extracts by PCR

For the determination of the presence of sugar cane DNA in the mill fractions, PCR amplifications

with a series of primer pairs designed from the repetitive IGS region of the large ribosomal RNA genes were performed (Figure 4). Using primer combinations that amplified different sized

fragments, the maximum size of DNA remaining in each fraction was estimated. The primer pair 4L/4R amplified the shortest fragment (273 bp) while the pair 2L/3R amplified the largest fragment (1760 bp) (Figure 2). PCR products of the expected size were amplified from the DNA preparations from all of the mill fractions when using the primer pair 4L/4R, although in very different final amounts (Figure 4; Panel A).

These data are from endpoint PCR with different amounts of input DNA. Therefore the intensities of the bands are not directly reflective of the actual differences in DNA in the reaction since those reactions with more input DNA reach saturation at a lower number of cycles. More realistic comparisons can be made by altering the amplification cycle number. Amplification from the FC, BG and CP was observed even when the number of PCR cycles was reduced from 35 to 20. Therefore the amplification differences apparent in Figure 4 represent a more than 10000-fold difference in target DNA. The amplicons from RW, VI, MO and CJ using 4L and 4R were gel-purified, reamplified, and sequenced. The sequence matched that of the previously determined sugar cane IGS region, confirming the presence of sugar cane DNA in these fractions (Figure 1 and Supplementary Figure 1S). In all of the presented data amplifications from only one of the samples is shown. However, DNA samples from all of the fractions were tested using all of the primers, including all the 6 replicate extractions for RW, VI, MO and CJ which gave results similar to those presented.

To test for possible contamination during the extraction process, an equivalent volume of double distilled water was substituted for the RW mill fraction and simultaneously extracted using the same reagents. These water extractions showed no amplification with any of the primers. Therefore, all the mill fractions contain sugar cane DNA with a size of >273 bp. As the PCR-target size increases from 273 bp to 1760 bp, the amplification from the samples from the later stages of the processing chain is reduced (Figure 4; Panel B-D). None of the DNA preparations from RW supported amplification of fragments as large as 550 bp. VI or MO did not support amplification of 800-bp long fragments and only one of the CJ derived samples gave detectable amplification of an 800-bp product (Figure 4, panel C). Finally, only DNA

preparations from the BG and CP directed amplification of the 1760-bp fragment. These data indicate progressive degradation of sugar cane DNA into smaller pieces during the refining process. These data are also consistent with previous studies, which failed to identify fragments >700 bp in the downstream sugar refining fractions (Joyce et al., 2013).

To estimate the amounts of sugar cane DNA in mill fractions with low DNA, we performed endpoint PCR comparing the amount of amplified product obtained using the 4L/4R primer pair and serial dilutions of the DNA from CP and the concentrated DNA samples from other fractions. DNA extraction samples from RW, VI, MO and CJ were estimated to contain less than 1 pg DNA per g of starting fraction. This is approximately equal to the DNA from a single cell per 4 g of fraction, using 4 pg as the value for the 1C nuclear DNA content of sugar cane (Kew Plant DNA C-values, data.kew.org/cvalues/).

Measurement of Protein in Sugar Cane Mill Fractions

Our next goal was to develop robust and sensitive methods to extract and characterize the protein content of sequentially derived processing fractions from the two mill types. A preliminary evaluation was made of the performance of two plant protein extraction methods: 1) 60 min at 65°C in LBT; 2) the commercial P-PER Plant Protein Extraction Kit (Pierce); and one general protein extraction method: 60 min at 65°C in SEB. Initial extracts required subsequent concentration/purification by precipitation with DOC + TCA before protein measurement since most of the downstream mill fractions (CJ, MO, RW, and RF) have a low protein content relative to the amount of DTT and dissolved solids which were likely to interfere with the protein assay.

In general, SEB extraction yielded 20 to 100% more protein from most mill fractions compared to LBT; with the exception of VI, the P-PER kit yielded 10 to 50% less protein than SEB extractions. The limit of protein detection after LBT extraction was 3 fold higher than for SEB extractions, in part due to the dilution of the sample caused by the addition of large amounts of urea/thiourea in LBT to the aqueous samples. Some LBT extracts gave poor results during SDS-PAGE analysis (not shown). Given these limita-

tions and fact that LBT and P-PER are optimized for extraction of protein from plant tissues (only two of eight mill fractions are “plant tissues”), the SEB extraction method was chosen for all subsequent analyses. Using our SEB extraction and TCA precipitation method of water or 20% refined sugar samples containing 1 to 100 µg per ml of a protein standard indicated the lower limit of detection as 5 µg/ml protein.

Two independently collected samples of mill fractions were extracted with SEB, precipitated with TCA/DOC, dissolved in 0.5 x SEB without DTT and assayed for protein content with a BCA-based assay (Table 3). For the most part variation between values obtained for each fraction from either mill type was not large (<30%). The protein content of CP was found to range between 3700 and 4600 ppm, consistent with the range of 2000-5000 ppm reported by Amalraj et al. (2010) for the 5 extraction methods they tested on experimental farm grown sugar cane stalks and the 20000 ppm average reported for dried cane stalks (feedipedia.org/node/14465). The first mill products, BG and PJ, each contain about half of the protein from the shredded cane stalk starting material (assuming 50% water imbibition). During clarification of the PJ fraction >95% of the protein is removed by the majority (>75%) ending up in the FC. During further processing of the CJ, the majority of protein in this fraction (~35 µg/ml protein; 200

ppm) ends up in the MO with only a small part (4%) remaining in the RW. We estimate that there is <1 g of protein in the RW derived from one metric ton of cane; this amount represents 0.005 - 0.020% of the original protein in the CP.

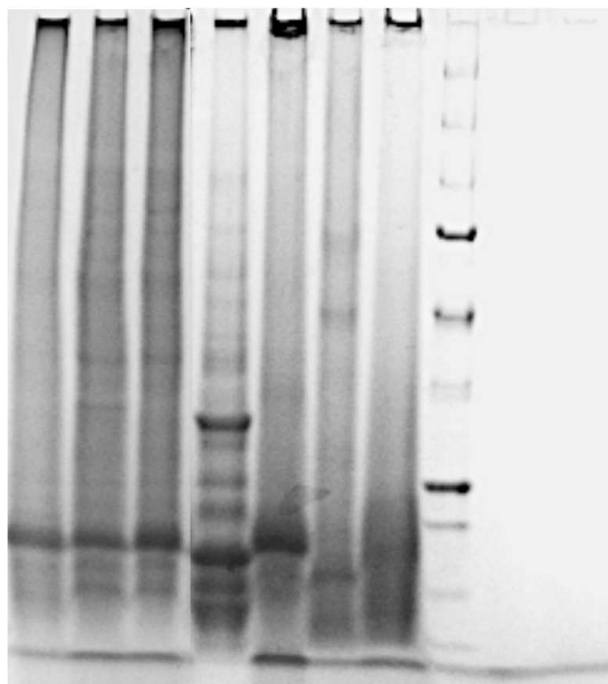
Since the protein levels detected in RW were so low, a protease digestion experiment was performed to ensure that the BCA reactive material is indeed protein. Digestion of 20 µg of RW-derived protein with 2 µg each of pronase and proteinase K reduced the intensity of the BCA reaction by 75%; an analogous treatment of 20 µg of bovine serum albumin also caused a 75% reduction. Thus, we are confident in our detection value of 9 to 15 ppm protein in raw sugar. It is consistent with the previous estimate of 20-60 ppm reported by Goodshall and Roberts (1976).

We also analyzed 3 batches of locally-purchased refined sugar. The average protein content was 1.4 ppm. However, this is close to the limit of detection as SEB buffer controls gave an equivalent value of 0.5 ppm. Furthermore SDS-PAGE analysis of 10 µg of ‘protein’ extracted from refined sugar by SEB showed no protein as indicated by Coomassie Blue staining (see below). Thus, we could not reliably detect protein in refined sugar; if present, the level of protein is probably well below the 1 ppm detection limit of our assay.

Table 3

Protein Content in PPM of Mill Fractions from Sugarcane Processing

Mill Fraction	Mill V (diffuser)		Mill G (tandem roller)		% Dry matter Mill V/MillG
	Sampling 1	Sampling 2	Sampling 1	Sampling 2	
CP	3840 ± 330	4150 ± 1060	4620 ± 420	3700 ± 960	100/100
BG	3300 ± 230	2680 ± 450	2810 ± 215	3290 ± 360	100/100
PJ	3620 ± 435	3510 ± 316	3140 ± 365	4240 ± 435	21/23
CJ	244 ± 24.2	234 ± 27.2	183 ± 19.6	148 ± 15.4	18/22
FC	10400 ± 1360	13500 ± 1520	11600 ± 1870	12300 ± 3230	100/100
MO	286 ± 88	274 ± 24.7	295 ± 52	254 ± 32	87/92
VI	3260 ± 195	3500 ± 225	3770 ± 237	3950 ± 375	3.1/3.2
RW	14.3 ± 4.1	8.6 ± 2.5	14.7 ± 4.6	13.3 ± 3.9	100/100



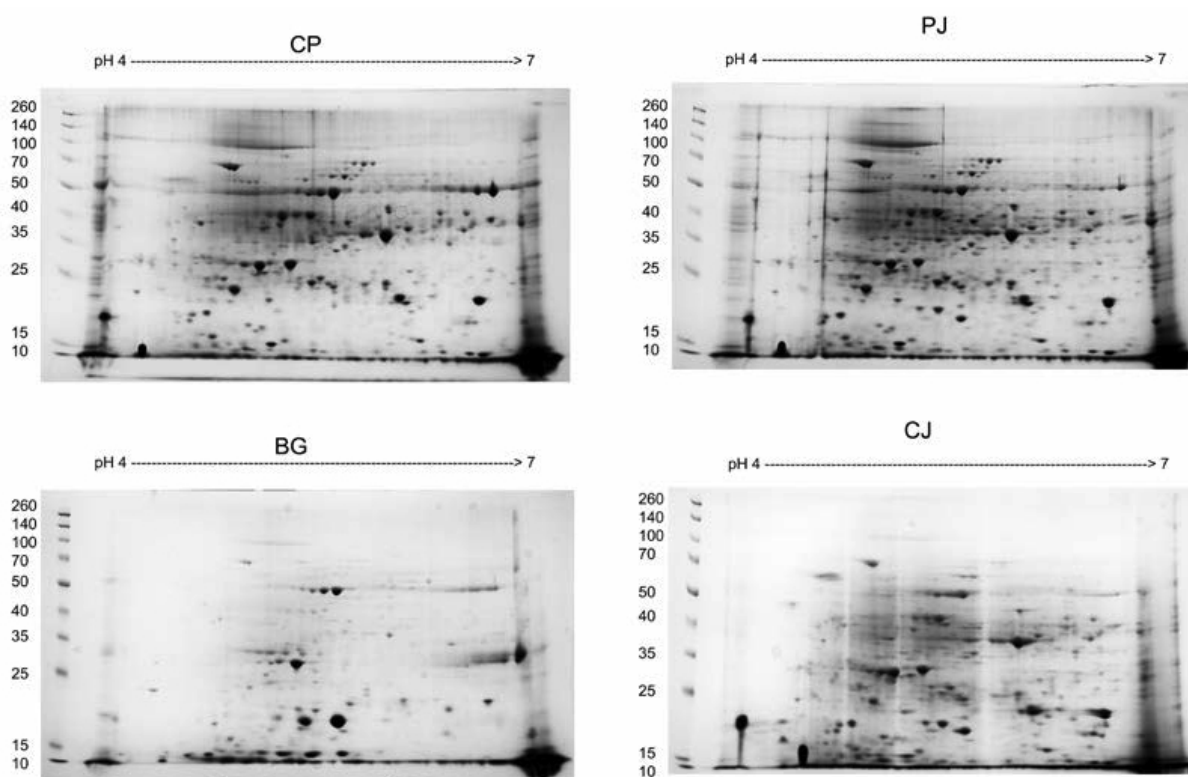


Figure 6: Representative 2-DGE for 100 µg of protein from early stage processing samples of mill V. Molecular masses of standard proteins are on the left; IEF range is at top.

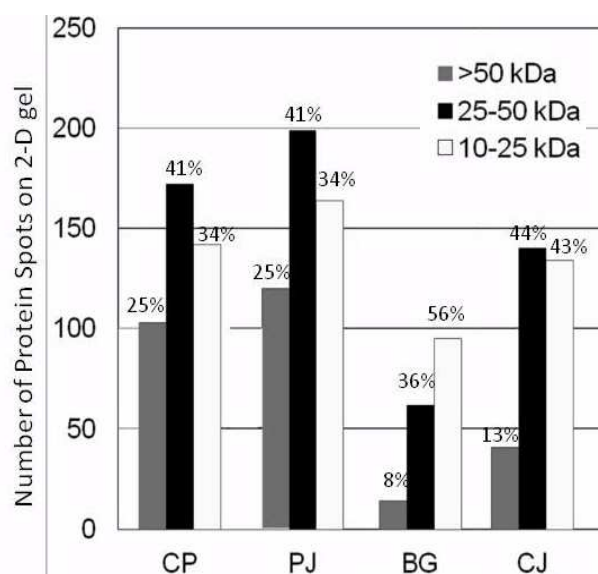


Figure 7: Analysis of protein spot number and size range for 2-DGE gels in figure 6. Numbers on top are percentages of total spots in each sample in designated molecular mass range. Total spot numbers are: CP, 417; PJ, 483; BG, 171; CJ, 315.

BG samples showed a different 2-DGE pattern and had less than half the number of protein spots observed with CP and PJ samples, suggesting BG may contain a subset of sugar cane proteins that are not efficiently extracted from the milling process. Relative to its immediate precursor fraction PJ, the total number of protein spots observed with CJ samples was 35% less with the most conspicuous loss in the higher molecular mass range. The number of protein spots >50 kDa was reduced from 120 in PJ to 45 in CJ. Less than 13% of the CJ protein spots were >50 kDa, compared to 25% for PJ.

Image analysis of the gels in Figure 6 shows that the majority of individual specific proteins present in CP and PJ are missing or reduced in CJ. This is not surprising since milling of PJ into CJ involves heating with lime which precipitates >90% of the protein into the FC (Table 3). All gels were loaded with equivalent amounts of protein to enable visualization of spots. If loading had been based on equal volumes of the fractions, some gels would have far fewer visible spots. These results demon-

strate that the milling process clearly reduces the complexity of the protein population.

To determine the identity of the proteins and/or peptides in each mill fraction, we used shotgun proteomics (i.e. LC-MS/MS analysis of tryptic peptides derived *en masse* from bulk, largely unfractionated proteins). Preparative SDS-PAGE was performed on the TCA-precipitated protein from each fraction to further eliminate non-protein materials and SDS which would interfere with trypsin digestion or mass spectrometry. Tryptic peptides were derived from whole lane gel-slices (not individual bands) and subjected to LC-MS/MS. The outputs were searched against the predicted proteome databases of sugar cane and *Sorghum bicolor*, as well as NCBI-nr to identify parent proteins of peptides identified by MS/MS. We used the *S. bicolor* proteome database because it is the closest evolutionary relative of sugar cane with a complete genome sequence. Furthermore, a comparison of 282,000 sugar cane ESTs of the coding DNA sequences (CDSs) of ESTs from the sorghum genome showed that 70% of sugar cane ESTs matched a *Sorghum* CDS with average amino acid sequence identity of 94% and vice versa (Nishiyama et al., 2011)

Peptides from 835 unique predicted sugar cane proteins were identified with 99% confidence in protein extracts from CP from mill G; 624 were identified from the cane used by mill V (Supplementary Table 1S). Peptides from 506 sugar cane homologs of *Sorghum* proteins from mill G were identified; 388 from mill V. More than 300 of these individual proteins were detected in cane from both mills. Of the 28 proteins identified in the cane stalk by Amalraj et al. (2010) using PMF, 23 were among the 835 identified here. PJ from mill G contained proteins or peptides from 475 predicted sugar cane proteins, while the CJ derived from it contained 244.

44 and 33 sugar cane proteins were identified in the raw sugar from mills G and V, respectively; 13 of these individual proteins or peptides thereof were found in the raw sugar from both mills (Supplementary Table 1S). Similar results were obtained searching the *Sorghum* database.

Searches of LC-MS/MS output against the NCBI-nr protein database implied the presence of a few nonplant proteins in some fractions, mostly keratin and a few bacterial, fungal or insect proteins. One

exception was the detection of a large number of bacterial proteins/peptides in PJ and BG of mill V. Although 75% of the proteins identified in PJ by searching NCBI-nr were from sugar cane-related plants, the remaining 25% were from members of the bacterial family *Leuconostocaceae*, mostly *Weissella* sp. Proteins from *Weissella* were not detected in the CP of mill V, but were found in its BG (~ 10% of the total identifications). *Weissella* are associated with sugar cane plants and often cause spoilage or souring after harvest (Bjorkroth et al., 2002). No *Weissella* proteins were detected in any fractions from mill G or downstream fractions of mill V.

The mill fraction with the most nonplant proteins was VI. Over 40% of the proteins identified in VI from either mill by searching NCBI-nr were from yeast (*Saccharomyces cerevisiae*); another 50% of the proteins identified in the VI from mill V were from *Lactobacillus* sp. In contrast, only 10% of the VI proteins identified in this fraction from mill G were from Lactobacilli. Nonetheless, many sugar cane proteins/peptides remained as significant components of this fermentation residue from both mills. VI from mill G contained ~three times as many residual sugar cane protein/peptides as VI from mill V. Presence of yeast and lactobacilli proteins in VI is not unexpected as VI is a residue from a yeast-based alcoholic fermentation of MO and contamination of the fermentation by *Lactobacilli* has been reported previously (Lucena et al., 2010; Gallo, 1990.)

Analysis of MO derived from the CJ of mill V showed the presence of 311 sugar cane proteins/peptides, while that from mill V showed 197. Approximately 108 of these were found to be coincident in MO from both mills. Although most of the protein from RW appeared degraded as judged by SDS-PAGE, we still identified sugar cane proteins or peptides: 44 from RW of mill G and 60 from mill V, and as previously noted, 13 unique sugar cane proteins/peptides were coincident in raw sugar from both mills.

Comparing all LC-MS/MS data, we found six predicted sugar cane proteins/peptides present in all sequentially produced sugar cane processing fractions (CP, PJ, CJ, MO, and RW) from both mills, but only 3 that were found in all fractions from all mills (Table 4).

Supplementary Table 1S

Results of LC-MS/MS Analyses of Mill Fraction Proteins

Fraction	Database searched	Unique protein identifications from Mill G	Unique protein identifications from Mill V	Proteins identified in fractions from both mills
Cane pieces (CP)	Cane	835	624	382
Cane pieces	Sorghum	506	388	305
Cane pieces	NCBI	695	552	nd
Primary juice (PJ)	Cane	475	202	184
Primary juice	Sorghum	338	299	171
Pressed juice	NCBI	348	400	nd
Bagasse (BG)	Cane	768	348	205
Bagasse	Sorghum	497	253	213
Bagasse	NCBI	608	305	nd
Clarified juice (CJ)	Cane	244	247	46
Clarified juice	Sorghum	107	147	74
Clarified juice	NCBI	145	183	nd
Filtercake (FC)	Cane	627	266	138
Filtercake	Sorghum	355	177	148
Filtercake	NCBI	601	117	nd
Molasses (MO)	Cane	311	197	108
Molasses	Sorghum	170	109	85
Molasses	NCBI	163	111	nd
Vinasse (VI)	Cane	108	33	20
Vinasse	Sorghum	53	21	16
Vinasse	NCBI	200	202	nd
Raw sugar (RW)	Cane	44	60	13
Raw sugar	Sorghum	28	38	13
Raw sugar	NCBI	12	56	nd

nd; not done

Table 4

Sugar Cane Proteins (ESTs) with Peptides in Most or All Mill Fractions

Accession Number	Description/Annotation from Compbio
TC123310	Similar to UniRef100_A5H454 Cluster: Plasma membrane-bound peroxidase 3-2 from <i>Zea mays</i>
*CA146186	Similar to UniRef100_Q2XX98 Cluster: Pathogenesis-related protein 5 from <i>Zea mays</i>
*TC113487	Homologue to UniRef100_A2X1D2 Cluster: Phosphoglycerate kinase from <i>Oryza sativa</i>
*TC128535	Homologue to UniRef100_Q6TGM7 Cluster: Dirigent from <i>Saccharum</i> hybrid
TC135999	Similar to UniRef100_Q38769 Cluster: Permatin from <i>Avena sativa</i>
CA273807	Homologue to UniRef100_A6N0M9 Cluster: Nucleoside diphosphate kinase from <i>Oryza sativa</i>

Data is from Compbio sugar cane EST database (http://compbio.dfc.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=s_officinarum);

*- protein found in all fractions from both mills.

These are apparently refractory to the harsh treatments that occur during processing of cane to raw sugar and/or are present in large amounts in the initial starting material. One of these TC128535 is likely the 18 kDa prominent band present in most fractions shown in Figure 5. Identifications and accession numbers of individual sugar cane proteins found in fractions from both mills and those which are shared among sequential mill processing fractions are posted online at:

https://dl.dropboxusercontent.com/u/16262713/milIG_shared_proteins.xlsx

https://dl.dropboxusercontent.com/u/16262713/fc_vi_bg_millGandV.xlsx

https://dl.dropboxusercontent.com/u/16262713/milIV_shared_proteins.xlsx

Conclusions

Methods were developed and applied to characterize the DNA and protein content of all major fractions produced during the milling of sugar cane into raw sugar. In contrast to previous studies, the methodology employed for DNA, particularly the use of a repetitive sequence target, and a series of primers amplifying different sized products within this target, allowed estimation of both amount and size of the DNA fragments present. Sugar cane DNA was found in all fractions from the refining process; however, in the later fractions (CJ, MO, VI, and RW) the amount was very low (<0.4 ng/g; 0.0004 ppm) and detectable only by PCR. The DNA was also of short length (<500 bp) compared to the DNA in early mill fractions. Thus, sugar cane DNA appears to be gradually degraded and lost as refining progresses; the amount of DNA in RW is >2000 fold less than in the CP starting material. PCR failed to detect any sugar cane DNA in refined sugar. These conclusions are consistent with previous studies on the detection of DNA through the refining process (Joyce et al., 2013; Klein et al., 1998; Oguchi et al., 2009; Taylor et al., 1999).

Our method developed to measure protein in mill fractions reliably detected as little as 5 µg/ml or 1 ppm of protein. We used precipitation with TCA/DOC to remove larger quantities of sugars and other contaminants in the complex and sometimes

dirty sugar cane mill fractions that can cause inaccurate protein determinations. Our protein measurements were not subject to such interference since mill samples spiked with known amounts of a standard protein gave accurate values. Furthermore, protease digestion experiments confirmed that >90% of the material in the RW and MO fractions that reacted with the protein detection reagent was indeed protein.

Similar to what was observed for DNA, the concentration of protein in the sequential mill fractions decreased by three orders of magnitude as processing progressed. We estimate that the majority (90%) of the CP protein ends up equally distributed between the BG and FC from the clarification step. We also estimate that <0.02% of the initial protein in CP ends up in the 'final' RW fraction. At the limit of detection of our assay (1 ppm), we could not detect any protein in refined sugar, indicating the additional refining removes most, if not all, of the relatively miniscule amounts of sugar cane protein remaining in RW.

Results of 1-D and 2-D gel electrophoresis of the mill fractions were consistent with the progressive degradation and fragmentation of the cane protein during the refining process. Proteins in RW, FC, and MO appeared to be the most extensively degraded. To circumvent this limitation, we used shotgun proteomics to determine the origin and identity proteins and peptides present in each mill fraction. Many proteins that were reported as abundant in sugar cane stalks by Amalraj et al. (2010) were detected, as were other abundant proteins involved in plant metabolism. The number of unique proteins/peptides detected in each fraction decreased as refining progressed. This is not unexpected as most steps of the refining process are very harsh for proteins (e.g. boiling and drying); such conditions would be expected to select from a small subset of proteins or peptides in later fractions that are thermally stable and resistant to drying. The vast majority of the hundreds of proteins found in most mill fractions were from sugar cane. The three exceptions were the BG and PJ of Mill V which contained proteins of *Weissella* sp., a reported bacterial inhabitant of sugar cane and/or VI from both mills. VI contains the residual solids from a yeast-based fermentation of molasses and CJ, and thus not surprisingly contained mostly yeast proteins as well as proteins

of lactobacilli that can contaminate large-scale alcoholic fermentations.

The methods described here may be useful in both quality control and in monitoring the fate of transgene products, both at the DNA and protein levels, in genetically-modified sugar cane or other plants and plant products.

Acknowledgements

This work was supported by Idea Connection. We acknowledge the contribution of M. Cullis in identifying the intergenic spacer region of the sugar cane ribosomal RNA genes, and J. N. Vignola in the 2-D gel analyses.

The authors have no conflicts of interest to report.

References

- Amalraj J., Selvarai R. S., Veluswamy, G.K., Ramanujan, R.P., Muthurajan, R., Palaniyandi, M., Agrawal, G.K., Rakwal, R., and Viswanathan, R. (2010), "Sugarcane proteomics: Establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map", *Electrophoresis* Vol. 31, pp. 1959-1974
- Bjorkroth, K.J., Schillinger, U., Geisen, R., Weiss, N., Hoste, B., Holzapfel, W.H., Korkealaand, H.J., and Vandamme, P. (2002), "Taxonomic study of *Weissella confusa* and description of *Weissella cibaria* sp. nov. detected in food and clinical samples", *IJSEM*. Vol. 52, pp. 141-148
- Gallo, C.R (1990), "Determinação da microbiota bacteriana de mosto e de dornas de fermentação alcoólica. PhD Thesis", University of Campinas, Campinas
- Goodshall, M.A. and Roberts, M.A. (1976), "Determination of refined sugar proteins by polyacrylamide gel electrophoresis" *Proc. Tech. Session Cane Sugar Refin.Res.*, New Orleans, LA, pp. 57-72
- Joyce, P.A., Dinh, S-Q., Burns, E.M. and O'Shea M.G. (2013), "Sugar from genetically modified sugar cane: Tracking transgenes, transgene products and compositional analysis", *Proc. Int. Soc. Sugar Cane Technol.*, Vol. 28, pp. 1 – 9
- Klein, J., Altenbuchner, J. and Mattes, R. (1998), "Nucleic acid and protein elimination during the sugar manufacturing process of conventional and transgenic sugar beets", *J. Biotechnology*, Vol. 60, pp. 145 – 153
- Lucena, B.T.L., dos Santos, B.M., Moreira, J.L.S., Moreira, A.P.B., Nunes, A.C., Azevedo, V., Miyoshi, A., Thompson, F.L., Morais, M.A. (2010), "Diversity of lactic acid bacteria of the bioethanolprocess", *BMC Micro.*, Vol. 10, pp. 29
- Nishiyama, M., Vicente F., Sato P., Ferreira,S., Feltus, F. and Souza, G. (2011), "The *Saccharinae* transcriptome", in Paterson A.H., (Ed) *Plant Genetics and Genomics: Crops and Models*, Vol. 11: *Genetics and Genomics of the Saccharinae*. Springer, N.Y., pp. 367-389
- Oguchi, T., Chikagawa, Y., Kodama, T., Suzuki, E., Kasahara, M., Akiyama, H., Teshima, R., Futo, S., Hino, A., Furui, S. and Kitta, K. (2009), "Investigation of residual DNAs in sugar from sugar beet (*Beta vulgaris* L.)", *J. Food Hyg. Soc. Japan*, Vol. 50, pp. 42 – 46
- Potter, R. L., Bacheller J. D., et al. (1990), "Isolation of proteins from commercial beet sugar preparations", *J. Agr. Food Chem.* Vol. 38, pp. 1498-1502
- Schell M.A., Zhao P., Wells L. (2011), "Outer membrane proteome of *Burkholderia pseudomallei* and *Burkholderia mallei* from diverse growth conditions", *J Proteome Res.* Vol. 10(5), pp. 2417-24
- Taylor, G.O., Joyce, P.A., Sedl, J.M. and Smith, G.R. (1999), "Laboratory crystallised sugar from genetically engineered sugar cane does not contain transgene DNA", *Proc Aust. Soc. Sugar Cane Technol.*, Vol. 21, pp. 502