LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TH-226437

Genotyping and Bioforensics of Ricinus communis

A. C. Hinckley

November 30, 2006

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Genotyping and Bioforensics of
Ricinus communis

By

AUBREE CHRISTINE HINCKLEY
B.S. (California State Polytechnic University, San Luis Obispo) 2001

THESIS

Submitted in partial satisfaction of the requirements of the degree of

MASTER OF SCIENCE

in

Forensic Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

_____

_____

Committee in Charge

2006

Acknowledgements

I am grateful for the technical assistance of Jessica Wollard, Marina Chiarappa-Zucca, and Shanavaz Nasarabadi at Lawrence Livermore National Laboratory.

I am also appreciative of the guidance from Paul Jackson at Lawrence Livermore National Laboratory, and from Sharon Messenger, the Lawrence Livermore National Laboratory project principal investigator.

Finally, thank you to my University of California at Davis Thesis Committee Members: Sharon Messenger, from Lawrence Livermore National Laboratory, George Sensabaugh, from the UC Davis Master of Science in Forensic Science program, and Paul Gepts, Thesis chair, from UC Davis Plant Sciences.

Table of Contents

Introduction

The castor bean plant (*Ricinus communis*) is a member of the family Euphorbiaceae. In spite of its common name, the castor plant is not a true bean (i.e., leguminous plants belonging to the family, Fabaceae). *Ricinus communis* is native to tropical Africa, but because the plant was recognized for its production of oil with many desirable properties, it has been introduced and cultivated in warm temperate regions throughout the world (Armstrong 1999 and Brown 2005). Castor bean plants have also been valued by gardeners as an ornamental plant and, historically, as a natural rodenticide. Today, escaped plants grow like weeds throughout much of the southwestern United States, and castor seeds are even widely available to the public for order through the Internet.

Although castor bean is the sole species in the genus *Ricinus*, castor plants exhibit great variation in appearance and in growth habit. Some plants grow as perennials, while others have an annual life cycle (Brown 2005). Castor plants exhibit variation in leaf size, shape, and coloring, in stem coloring, in seed pod morphology (e.g., smooth or spiked), and in seed coloration and size. The specific name for castor bean plants, *Ricinus communis*, translates to "common tick", reflecting the striking mottled appearance of the seed which looks like the body of a tick.

The seeds of *R. communis* consist of 40-60% by weight of one of the world's most useful natural plant oils (Shamns 2005). Castor oil is widely used for its lubricating properties and for medicinal purposes. In industry, castor oil is used for the manufacturing of soaps, lubricants, hydraulic and brake fluids, paints, dyes, coatings, inks, cold resistant plastics, waxes and polishes, nylon, pharmaceuticals and perfumes

(Franz and Jaax 1997, Morris 1996-2003 and Duke 1998). Castor oil contains ricinoleic acid, which has been effective at inhibiting the growth of many viruses, bacteria, yeasts, and molds (Williams 1995), which lends credence to the medicinal use of castor oil as a topical treatment for a variety of skin infections and ailments. Castor oil is also used as a laxative, as an emetic, and to induce labor in close to full-term pregnant women (Duke 1998).

The castor bean plant also produces a toxin called ricin that is of increasing interest for use as an agent of biowarfare or biocrimes. Ricin can be found in the bean mash that is left over after castor oil has been extracted from castor seeds (CDC 2004). Ricin is a potent cytotoxic type II ribosome-inactivating protein consisting of an A chain and a B chain linked by a disulfide bridge. The A chain is a RNA N-glycosidase that de-adenylates the 28S rRNA of the 60 S subunit of ribosomes, thus irreversibly inhibiting protein synthesis (Brown 2005, Franz and Jaax 1997). The B chain is a lectin that facilitates recognition, binding, and entry into cells to allow the A chain to cross cell membranes and become activated once inside the cell (Brown 2005, Franz and Jaax 1997). As little as 200 µg of ricin is a lethal dose for humans, depending upon whether the exposure is through inhalation, injection, ingestion, or absorption through the skin (Brown 2005, CDC 2004, Franz and Jaax 1997). Ricin seeds consist of as much as 3% by weight of toxic substances, and 2-3 seeds possess enough toxin to be lethal to the average adult (Griffiths 1987). Unfortunately, there is no known antidote for exposure to ricin. Treatment for ricin intoxication varies depending upon the route of exposure, and although treatment cannot reverse the effects of exposure to the toxin, symptomatic care is administered (CDC 2004, Franz and Jaax1997). Ricin is a stable protein, and recipes

and reagents for crude toxin preparations are easily accessible and widely available on the Internet. These characteristics make ricin a popular choice for use as a biological agent for criminal and bioterrorist activities.

A brief history of the use of ricin in biocrimes

A number of criminal cases involving ricin have been documented. In 1978 Georgi Markov, a Bulgarian dissident, was assassinated by the Bulgarian government and the KGB by using an umbrella to inject a ricin pellet into Markov's leg (Carus 1998 and Tompkins 2004). In the same year the Bulgarian Secret Police also attempted to assassinate the Bulgarian defector Vladimir Kostov (Carus 1998). Ricin was also reportedly used in the Iran/Iraq war in the 1980's (Tompkins 2004). Members of the Patriot's council, an antigovernment extremist group, were arrested in 1991 for planning to kill a US Marshall by putting ricin on his car door handles (Carus 1998 and Mirarchi and Allswede 2006). In 1993 Dwayne Kuehi was convicted for possession and planned used of ricin to murder a building inspector, and in 1995 Debora Green, a medical doctor, was convicted of poisoning her husband with ricin (Carus 1998). Thomas Leahy was also convicted for possession and threatened use of ricin as a weapon (Carus 1998). Recipes for ricin preparation were found in Al Qaeda caves in Afghanistan in 2001, and traces of the toxin were detected at Al Qaeda sites that were likely used for making biological weapons (Tompkins 2004). Use of castor seeds to kill unwanted children has also been reported in the Republic of Malawi (Carus 1998). Six suspects were arrested after a ricin production laboratory was discovered in their apartment in Manchester, England in 2002. British police also found traces of ricin in two residences in the

London area where suspects were arrested for their involvement in a plan to use ricin to attack the Russian embassy in Chechnya (Mirarchi and Allswede 2006). In 2003, a series of letters and a vial of ricin were sent to the Department of Transportation (Tompkins 2004). The letters demanded that the sleep requirement for truck drivers remain at 8 hours, instead of being changed to a required 10 hours of rest between shifts. The letters threatened to dump ricin into water supplies if the demands for an 8-hour sleep requirement were not met. In 2004, ricin was also found in the mailroom that delivers mail to Senate Majority Leader Bill Frist's office (Hanson 2004 and Mirarchi and Allswede 2006). There are also numerous cases involving the threatened use of ricin without confirmation of possession of the toxin, or the confirmed possession of ricin without evidence of intent to commit a crime (Carus 1998).

The documented cases of actual or threatened use of ricin, along with the potency, wide availability, and easy preparation of the toxin demonstrate a forensic need to develop a typing method to link ricin evidence to a particular source. Therefore, it is necessary to characterize the genetic diversity present across *Ricinus communis* cultivated varieties from different geographic regions to develop a genotyping scheme that links castor bean evidence to a particular source, geographic region, or batch.

Forensic molecular tools to identify plant species and populations

The field of forensic science uses a variety of PCR-based typing systems followed by methods that detect either length or sequence variation to identify DNA polymorphisms in a variety of samples. DNA polymorphisms based on length include variable number tandem repeats (VNTR), short tandem repeats (STR), and simple

sequence repeats (SSR). These length-based analysis methods tend to allow for easier multiplex capability, statistical calculations, and less labor-intensive effort compared to sequence-based analyses. Sequence-based DNA polymorphisms, such as multiple locus sequence typing (MLST) and single nucleotide polymorphisms (SNP) are also targeted for forensic typing.

Specifically in the field of forensic botany, there have been evolving strategies and technologies employed to address plant evidence in criminal cases over the past twenty years. Historically, forensic scientists used plant anatomy, such as leaf morphology and tree growth ring patterns, to form physical matches and to identify plant species for botanical evidence (Coyle 2005). In fact, the first case to use plant systematics and plant anatomy for evaluating plant evidence in court was in 1932 when Charles Lindbergh's son was kidnapped from a second story room. The suspect used a wooden ladder to gain access to the nursery. Microscopic analysis of the wood-grain patterns from the ladder was used to tie the suspect to this crime (Coyle 2005). Classical plant DNA typing technologies emerged, and Random Amplified Polymorphic DNA Analysis (RAPD) became the molecular method of choice for botanical evidence analysis. This technique was used in a groundbreaking case in Phoenix, Arizona in 1992 (Coyle 2005). RAPD was used to link pods from the suspect's truck to a paloverde tree at the crime scene where the body of Denise Johnson was found (Coyle 2005). Although RAPD compares genetic profiles within a strain or species, it requires large quantities of high quality, high molecular weight DNA as a template, which is problematic for forensic DNA evidence. Nuclear ribosomal internal transcribed spacers 18S, ITS1, and ITS2, and chloroplast genes such as rbcL, atpB, and ndhF have been used in forensic botany to

identify plant species based on sequence polymorphisms (Dommelen 2002 and Coyle 2005). For intraspecies differentiation, sequence from chloroplast intergenic spacers and AFLP have been utilized, although AFLP can be problematic for challenging forensic field samples as it is particularly susceptible to inhibition from dirty samples, mixtures are problematic, and repeatability is challenging when working with degraded evidence. Among the DNA polymorphisms available to target, analysis and assay development of VNTRs are optimal because VNTRs enable rapid testing of multiple loci with high potential for discriminatory power due to the multiple alleles possible at each locus. In contrast, SNPs are typically bi-allelic, although three or four alleles are theoretically possible at a polymorphic site.

The type of forensic sample or evidence must also be considered when evaluating appropriate DNA methods to exploit the underlying genetic variation. Castor bean evidence would most likely be found in the form of plant material, seeds, bean mash, or crude ricin preparations, which have all been shown to contain DNA. In the case of crude ricin preparations, trace amounts of degraded DNA may be present. In targeting a particular type of DNA polymorphism to be used as part of a genotyping scheme for *Ricinus communis*, it is advantageous to target organelle DNA. Small circular organelle DNA is more resistant to degradation than linear, nuclear DNA, and is present in higher copy number than nuclear DNA. Thus it is a more appropriate target for genotyping challenging samples containing trace or degraded DNA. Plant mitochondrial DNA, also present as a circular molecule in high copy number, is known to undergo frequent rearrangements (i.e., recombination). Chloroplast DNA is more genetically stable, therefore, it is a more appropriate organelle target for development of genotyping assays.

Detection and characterization of *Ricinus communis* and its toxin

The *Ricinus communis* genome has not been sequenced in its entirety. Nucleotide sequence for different ricin genes, RCA (*Ricinus communis* agglutinin), a few ribosomal RNA genes and intergenic spacers, and a limited number of housekeeping genes for photosynthesis and plant metabolism are available through Genbank. The Institute for Genomic Research (TIGR) is also actively working to sequence the *Ricinus communis* chloroplast genome for several cultivars. Conversely, a multitude of protein-based detection assays for ricin exist to confirm the existence of the toxin in a variety of matrices. Historically, confirmation of intoxicating exposure to ricin has been carried out by enzyme-linked immunosorbent assay (ELISA) of blood or other body fluids. More recently, Huelseweh et al. (2006) published a protein microarray detection system to identify ricin and other biowarfare agents in samples from the environment, battlefields, and food to aid in biodefense and risk assessment processes. Lubelli et al. (2006) also present an immuno-polymerase chain reaction assay to detect ricin and other ribosome-inactivating proteins. These publications highlight a few of the many techniques available to detect ricin using antibody-based systems, but there are no widely available DNA-based genotyping assays for *Ricinus communis*. Furthermore, to date, there are no published phylogeographic studies of *Ricinus communis* cultivated varieties.

Phylogeographic studies of plants

Phylogeography aims to discover the role that historical factors have played in influencing the genetic structure of populations within a species by studying the history of individual genes distributed throughout varied geographical populations in order to construct genealogical trees. The concept of phylogeography was first introduced by Avise et al. (1987). Most phylogeographic studies targeted mitochondrial markers in animal systems while phylogeographic studies of plant species have been less common (Schaal et al. 1998). Schaal et al. (1998) suggest that phylogeographic studies are applicable and play an important role when considering plant evolution, but that plant phylogeographic research has often been challenging due to a lack of appropriate intraspecies genetic diversity. As Schaal et al. (1998) indicate, although mitochondrial DNA is targeted for phylogeographic studies in animals, plant mitochondrial DNA undergoes frequent rearrangements, intramolecular recombinations, and has a low nucleotide substitution rate making it unsuitable for the study of plant phylogeographic variation. Chloroplast DNA, in contrast, exhibits a much higher rate of nucleotide substitution, and the arrangement of chloroplast genes is highly conserved. Chloroplast DNA is thus a better target for phylogeographic studies in plants. Moreover, the existence of tRNA genes and genes for photosynthesis in a highly conserved arrangement allows the development of consensus oligonucleotides that can be used in PCR-based plant population phylogenetic studies. Schaal et al. (1998) assert that for the chloroplast genome, the highest mutation rate occurs in the single-copy regions as opposed to in the inverted repeats (Schaal et al. 1998). Moreover, Taberlet and Gielly (1991) also state that the noncoding regions in the single-copy portion of the chloroplast genome should

exhibit the highest variability because they are subject to less functional constraint than the coding regions (Schaal et al. 1998). The intergenic regions are easily targeted for PCR amplification and subsequent analyses by using universal oligonucleotides designed to amplify portions of the highly conserved flanking genes. These universal oligonucleotides were originally designed by alignment of complete cpDNA sequences from GenBank or Embl for *Nicotiana tabaccum* (tobacco), *Oryza sativa* (rice), and *Marchantia polymorpha* (liverwort) and identification of conserved regions (Petit et al. 1996). Smaller amplicons are readily characterized by direct sequencing of PCR products, and PCR RFLP can be used for comparative analysis of larger noncoding amplicons.

Shaw et al. (2005) published a study evaluating the value of using 21 noncoding chloroplast DNA regions for phylogenetic investigation of species-level relationships based on sequence comparisons from samples representing 10 evolutionarily diverse major plant groups. The non-coding regions were ranked according to the highest average potentially informative character value (the collective number of nucleotide substitutions, indels, and inversions) in order to ultimately highlight several noncoding chloroplast regions that are well suited for low-level (i.e., interspecies or population level) systematic studies (Shaw et al. 2005). Olsen and Schaal (1999) also published a phylogeographic study of another member of the family Euphorbiaceae, cassava (*Manihot esculenta)*, using the single-copy nuclear gene glyceraldehyde 3-phosphate dehydrogenase (G3pdh) (Olsen and Schaal 1999, Olsen and Schaal 2001, Olsen 2004 and Schaal and Olsen 2000). This study found high levels of sequence variability in the

noncoding portions of this gene when this locus was compared between cassava and its relatives.

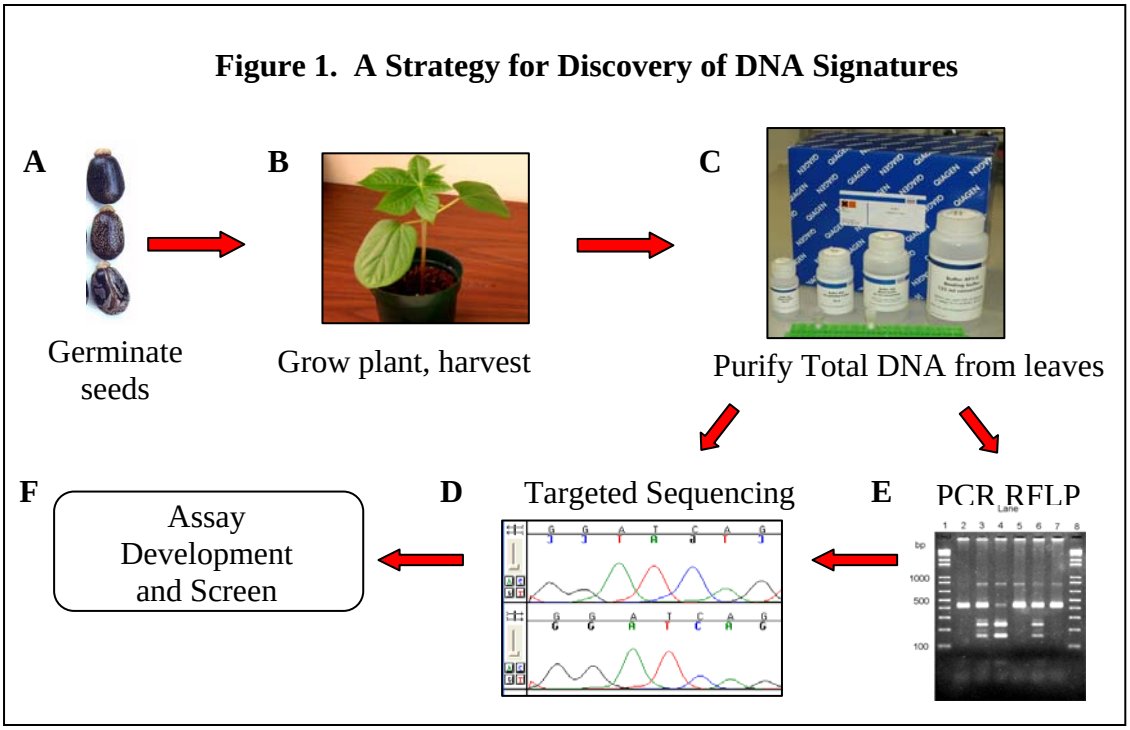
Study design to forensically characterize *Ricinus communis* cultivars

After consideration of the published literature and given 1) the limited available genetic data for *Ricinus communis*, 2) the conventional use of noncoding chloroplast regions in plant phylogeographic studies, 3) the existence of universal oligonucleotides targeting these regions, and 4) the advantage of using organelle DNA in forensic analyses, published universal oligonucleotides for noncoding cpDNA were investigated for the present phylogeographic study of *Ricinus communis* cultivated varieties.

In this study, multiple loci of chloroplast noncoding sequence data and a few nuclear noncoding regions were examined to identify DNA polymorphisms present among representatives from a geographically diverse panel of *Ricinus communis* cultivated varieties. The primary objectives for this research were 1) to successfully cultivate castor plants and extract sufficient yields of high quality DNA from an assortment of castor cultivated varieties, 2) to use PCR and sequencing to screen available universal oligos against a small panel of castor cultivars, 3) to identify DNA polymorphisms within the amplified regions, and 4) to evaluate these DNA polymorphisms as appropriate candidates for assay development (see Figure 1). Additional goals were to design, test and optimize assays targeting any DNA polymorphisms that were discovered and to rapidly screen many castor cultivars to determine the amount of diversity present at that particular locus. Ultimately, the goal of this study was to construct a phylogeographic tree representing the genetic relationships

present among *Ricinus communis* cultivars from diverse geographic regions. These research objectives were designed to test the hypothesis that cultivated varieties of *Ricinus communis* from various geographic regions can be distinguished from one another based on differences present at the genetic level. In addition, the present study sought to determine the amount of diversity present among *Ricinus communis* cultivars.

Figure 1. A Strategy for Discovery of DNA Signatures



Strategical flow chart for signature discovery process. A. Germinate different Castor bean cultivar seeds. **B.** Cultivate different Castor bean cultivars and harvest leaf tissue for DNA extraction. **C.** Purify total Castor bean DNA from leaves using the Qiagen DNeasy Plant Mini DNA Extraction kit. **D.** Sequence PCR amplicons of various noncoding regions on the chloroplast genome to identify DNA polymorphisms. **E.** Use restriction enzymes to detect differential digest patterns of noncoding chloroplast PCR amplicons among various Castor bean cultivars. Identify polymorphisms responsible for differential digests by targeted sequencing. **F.** Develop assays for DNA polymorphisms and screen additional Castor bean cultivars using these assays.

Materials and Methods

Growth of cultivated varieties

Sixty *R. communis* cultivars were studied for this research (see Table 1 and Figure 2). A diverse selection of seeds (Figure 3) was obtained from a USDA collection of 300 cultivars shared between Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory (LLNL) under the ownership of Dr. Paul Jackson. The seed coats of 5 seeds per cultivar were scarified with 120 grit sandpaper and placed in a damp paper towel in a plastic bag. Bags were placed in a humidified incubator at 37 °C and the seeds were left to germinate in the dark for 3-7 days. Upon germination, 3-5 seeds were potted using Miracle Grow Potting Mix and disposable 4-inch square pots. The germination efficiency was recorded for each propagated cultivar. Samples were exposed to 12 hours of wide-spectrum light per day using Gro Lux 40 watt lights, and germinated seeds and plants were watered to keep soil moist using tap water. As the cotyledons surfaced, the sprouting date was recorded. Plants were grown for an additional 2-3 weeks until plants had 2 or more substantial true leaves, at which time the leaf tissue was harvested for DNA extraction (see Figure 3). Excess leaf material was sealed in a plastic bag and archived at -80 °C.

DNA Extraction

100 mg of fresh leaf tissue was harvested for each cultivar DNA extraction. Two methods were employed to prepare leaf tissue for DNA extraction. Either 100 mg of leaf tissue was kept frozen under liquid nitrogen and ground to a powder using a ceramic mortar and pestle, or 100 mg of leaf tissue was frozen under liquid nitrogen and bead beat

at 38,000 RPM for 5 seconds using 2.3 mm stainless steel beads. The latter method was used to facilitate faster processing times for extraction of a large number of samples at the same time. Frozen ground leaf tissue was then extracted using the Qiagen DNeasy Plant Mini DNA Extraction Kit according to the kit protocol, with the following specifications: to include the optional 5 min lysate spin at 20,000 x *g* and two 100 µL elutions per cultivar. Quality and quantity of the extracted DNA was determined by electrophoresis through a 0.8 % agarose gel containing 0.5 µg/mL ethidium bromide followed by visualization under UV light.

Amplification

Universal primer sequences from a variety of literature sources (Petit et al. 1996, Taberlet and Gielly 1991, Shaw et al. 2005, Demesure et al. 1995, and Chiang et al. 1998) were ordered from Sigma Genosys and were used to amplify regions targeting the intergenic spacers of the chloroplast genome (Table 2). Briefly, 5 ng of castor bean DNA was used as a template in 50 µl PCR reactions using Clontech Advantage 2 PCR Reagents (buffer, dNTPs, and polymerase) and primers (1 µM final concentration each) using the PCR parameters recommended for each primer set by the references supplying the primer sequences. PCR products were electrophoresed through a 2 % agarose gel containing 0.5 µg/mL ethidium bromide and visualized under UV light. PCR products were cleaned up using the Qiagen QIAquick PCR purification kit, or the Edge Biosystems ExcelaPure 96-Well UF PCR Purification Kit and eluted in 52 uL of elution buffer. If DNA fragments other than those expected based on the primers used were present, PCR products were gel purified using the Qiagen Gel Extraction kit, and eluted in 52 µL of elution buffer.

Targeted Sequencing

PCR products from different cultivars were sequenced and compared to one another to identify DNA polymorphisms present among cultivars for any given intergenic region. Eight µL of purified PCR product was used as a template for sequencing reactions using the GenomeLab DTCS kit from Beckman Coulter and the appropriate primer at 1 µM final concentration, according to the following parameters: 96 °C (20 s), 50 °C (20 s), 60 °C (4 min) for 30 cycles. Sequencing reactions were purified according to the suggested protocol, analyzed on the Beckman Coulter CEQ 8000 under the lfr-a sequencing method, and the resulting sequences were analyzed using the default PCR product parameters. Sequences were imported into Sequencher software and assembled using 70-85 % minimum match percentage and 15-20 % minimum overlap.

RFLP

To minimize the amount of sequencing performed, Restriction Fragment Length Polymorphism (RFLP) was used to screen PCR products for differences between cultivars. This protocol streamlined the search for relevant DNA polymorphisms by screening a higher number of primer pairs and individual cultivars. Purified PCR products from different cultivars were digested with the restriction endonucleases *Mse*I, *Apo*I, *Ase*I, and *Dra*I. DNA fragment banding patterns were compared to one another following electrophoresis through 2-4 % agarose gels containing 0.5 µg/mL ethidium bromide and visualized under UV light. Digests of intergenic regions that revealed differential fragment patterns among cultivars were verified by sequencing of that PCR amplicon according to the procedure listed above for targeted sequencing.

Assay Design

Assays were designed to target the DNA polymorphisms discovered in the intergenic spacer between trnC and trnD (trnCD assay) and between trnS and trnT (C del assay). Primers were designed to flank the DNA polymorphisms using Primer 3 software (Rozen and Skaletsky 2000), and primer pairs were ordered from Proligo with a Beckman Coulter CEQ 8000-compatible fluorescent dye (D2 or D3) attached to the 5' end of the forward primer. D2 labeled primers pairs were used at a final concentration of 0.16 µM each in a 25 µL PCR reaction to amplify 2.5 ng of castor bean DNA using Clontech Advantage 2 PCR Reagents (buffer, dNTPs, and polymerase). The trnCD assay PCR was run with the following parameters: 95 °C (1 min) followed by 95 °C (30 s), 48 °C (30 s), 68 °C (60 s) for 30 cycles, and a final incubation at 68 °C for 7 min. The C del assay PCR used similar PCR cycling parameters with slight modifications: 95 °C (2 min) followed by 95 °C (30 s), 46 °C (30 s), 68 °C (60 s) for 30 cycles, and a final incubation at 68 °C for 7 min. PCR products were checked for the presence of amplicons by electrophoresis through a 2 % agarose gel containing 0.5 µg/mL ethidium bromide followed by visualization under UV light. Internal standards were developed to represent all alleles observed for the DNA polymorphisms. The standards were made by amplification as stated above, but using D3-labeled forward primers. Unknown PCR products and internal standards were diluted 1:8 using PCR-grade water. Two µL of the unknown diluted PCR reaction was added to 2 µL of each internal standard, 0.35 µL of Beckman Coulter GenomeLab DNA size standard 600, and 31.65 µL Beckman Coulter GenomeLab Sample Loading Solution for a final volume of 40 µL. Samples were

analyzed on the Beckman Coulter CEQ 8000 using the frag 4 + 5 fragment analysis method using default parameters.

Figure 2. Geographic distribution of *Ricinus communis* cultivar templates

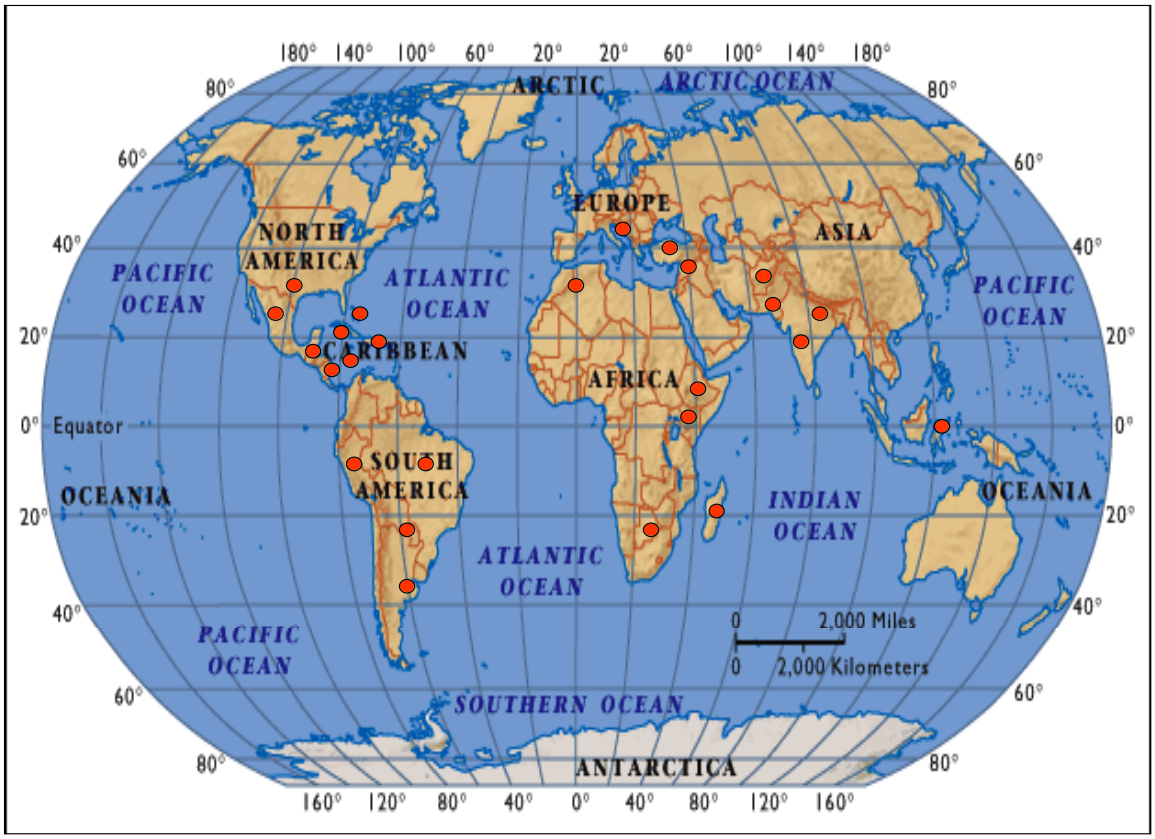
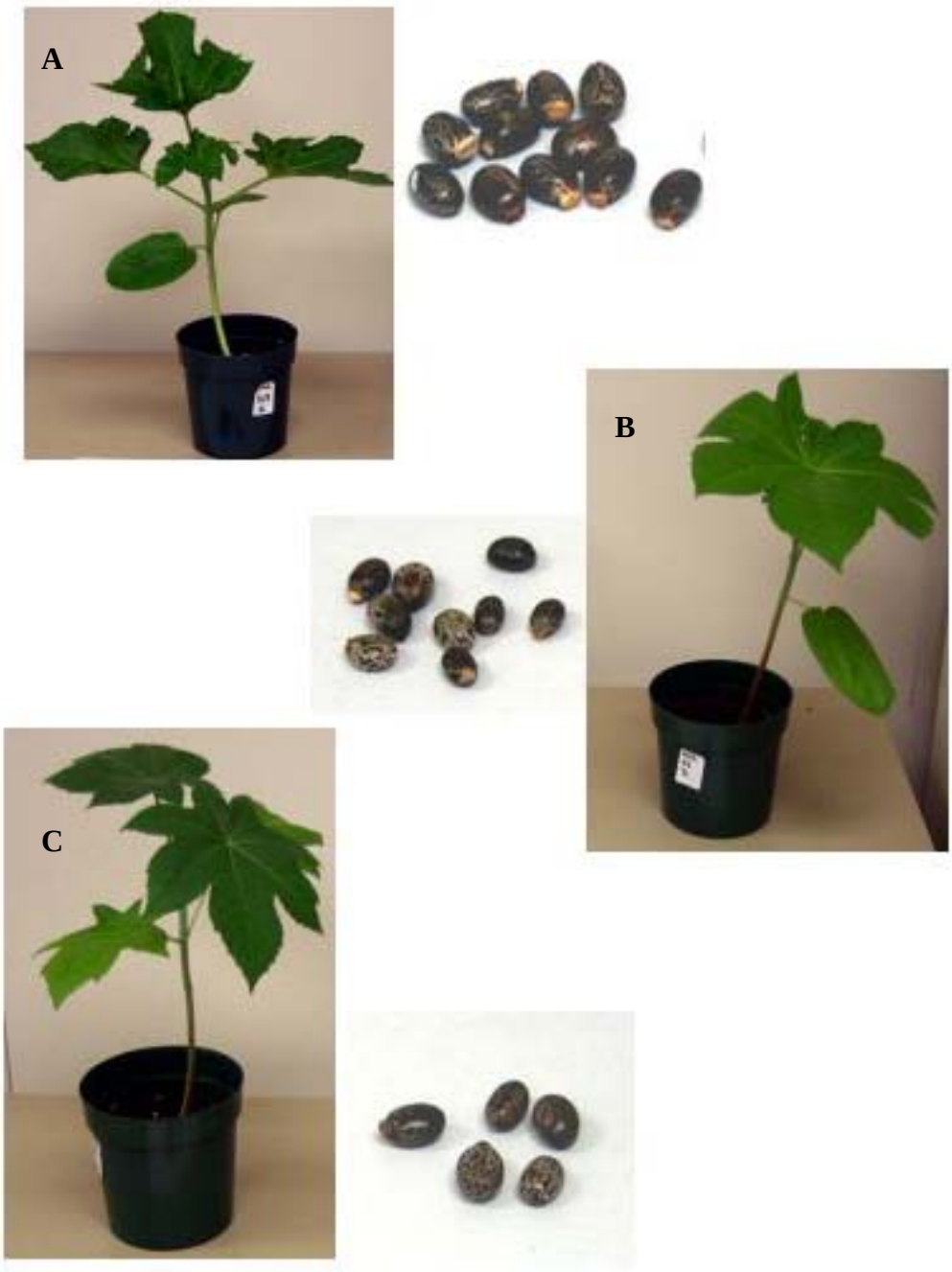


Figure 3. Photographs of a sampling of Castor bean plants and seeds cultivated for DNA Extraction.



Photographs of Castor bean seeds and plants cultivated for this study. A. Plant and seeds from Castor cultivar LL-163 (Texas). **B.** Plant and seeds from Castor cultivar LL-52 (Kenya). **C.** Plant and seeds from Castor cultivar LL-39 (Syria).

Table 1. Castor Cultivar DNA Templates

Item #	LLNL #	Accession #	Source Country	Plant Name
1	LL-1	PI 162 912	Paraguay	235
2	LL-2	PI 163 162	Brazil	Var 39
3	LL-4	PI 165 446	Mexico	Higuerilla
4	LL-7	PI 167 238	Turkey	Hindyagi Giza No
5	LL-8	PI 167 287	Turkey	481
6	LL-9	PI 167 288	Turkey	482
7	LL-10	PI 167 342	Turkey	536
8	LL-11	PI 170 682	Turkey	1393
9	LL-12	PI 170 684	Turkey	Kahverengi
10	LL-14	PI 170 686	Turkey	2334
11	LL-15	PI 173 090	Turkey	7569
12	LL-17	PI 173 795	Turkey	8750
13	LL-18	PI 173 946	India	Errand
14	LL-19	PI 173 947	India	Errand
15	LL-20	PI 173 948	India	Errand
16	LL-30	PI 179 729	India	No 10394 Harind
17	LL-39	PI 181 916	Syria	Hama 21
18	LL-44	PI 183 078	India	Divela
19	LL-46	PI 183 347	India	Divela
20	LL-47	PI 183 468	India	Erari
21	LL-48	PI 183 470	India	Andi
22	LL-51	PI 184 133	Yugoslavia, Serbia	No. 303
23	LL-52	PI 192 949	Kenya	Mauthner
24	LL-54	PI 195 811	Guatemala	2835
25	LL-55	PI 197 048	El Salvador	3027
26	LL-57	PI 201 830	Madagascar	Tige Blanche
27	LL-59	PI 202 667	India	HC 3
28	LL-61	PI 202 711	Brazil	1156
29	LL-63	PI 203 126	India, Delhi	TMV-1
30	LL-64	PI 203 128	India, Delhi	TMV-3
31	LL-65	PI 203 130	India, Delhi	EB 31
32	LL-66	PI 203 661	Paraguay	
33	LL-67	PI 204 321	India	Chittaharalu
34	LL-68	PI 204 322	India	Rosy Castor
35	LL-72	PI 206 515	Jamaica	
36	LL-74	PI 207 868	Peru	La Chocolate
37	LL-75	PI 208 464	Nepal	
38	LL-76	PI 208 689	Algeria	
39	LL-80	PI 208 840	Cuba	
40	LL-82	PI 208 842	Cuba	
41	LL-84	PI 209 132	Puerto Rico	
42	LL-85	PI 209 326	Virgin Islands (U.S.)	
43	LL-87	PI 209 622	Cuba	
44	LL-88	PI 212 115	Afghanistan	12957
45	LL-96	PI 215 775	Peru	TM 2957
46	LL-98	PI 217 539	Pakistan	13933
47	LL-102	PI 219 766	Botswana	Dwarf Castor
48	LL-103	PI 219 767	Argentina, Buenos Aires	No. 464
49	LL-104	PI 219 770	Argentina, Buenos Aires	Rouge de Fez
50	LL-122	PI 221 698	Indonesia, Java	No. 1
51	LL-123	PI 222 265	Iran	1407
52	LL-135	PI 229 785	Iran	15783
53	LL-142	PI 241 362	Brazil	No. 374
54	LL-147	PI 241 368	Brazil	No. 383
55	LL-150	PI 241 371	Brazil	No. 397
56	LL-157	PI 246 996	Bahamas	
57	LL-162	PI 248 390	India	K31
58	LL-163	PI 599 751	USA (TEXAS)	LYNN
59	LLC-1			Whatcom seed company
60	LLC-3	G6729		Mainstreet

LL: LLNL # from USDA Inventory

LLC: Commercial seeds ordered from internet

* Preliminary DNA panel

* Secondary DNA panel

Table 2. Castor Primer Record

Primer Name	Primer Sequence	Gene Region	Approx. Amplicon Size	Tier*	Rank**	Reference
AtpB-1	ACATCKARTACKGGACCAATAA	atpB-rbcL	1200			Chiang *et al.* 1998
rbcL	AACACCAGCTTTRAATCCAA	atpB-rbcL	1200			Chiang *et al.* 1998
GPDX7F	GATAGATTTGGAATTGTTGAGG	G3pdh	1200			Strand *et al.* 1997
GPDX9R	AAGCAATCCAGCCTTGG	G3pdh	1200			Strand *et al.* 1997
psaA.F	ACT TCT GGT TCC GGC GAA CGA A	psaA-trnS	3500			Demesure *et al.* 1995
trnS.2.R	AAC CAC TCG GCC ATC TCT CCT A	psaA-trnS	3500			Demesure *et al.* 1995
psbC.F	GGT CGT GAC CAA GAA ACC AC	psbC-trnS	1700			Demesure *et al.* 1995
trnS.1.R	GGT TCG AAT CCC TCT CTC TC	psbC-trnS	1700			Demesure *et al.* 1995
rpL16F71	GCT ATG CTT AGT GTG TGA CTC GTT G	rpL16	1000	2	7/21	Shaw *et al.* 2005
rpL16R1516	CCC TTC ATT CTT CCT CTA TGT TG	rpL16	1000	2	7/21	Shaw *et al.* 2005
trnC.GCA.R	CAC CCR GAT TYG AAC TGG GG	rpoB-trnC	1300	1	2/21	Shaw *et al.* 2005
rpoB	CKA CAA AAY CCY TCR AAT TG	rpoB-trnC	1300	1	2/21	Shaw *et al.* 2005
rpS16.F	AAA CGA TGT GGT ARA AAG CAA C	rpS16	850	2	6/21	Shaw *et al.* 2005
rpS16.R	AAC ATC WAT TGC AAS GAT TCG ATA	rpS16	850	2	6/21	Shaw *et al.* 2005
trnC.F	CCA GTT CAA ATC TGG GTG TC	trnC-trnD	3500	3	11/21	Demesure *et al.* 1995
trnD.R	GGG ATT GTA GTT CAA TTG GT	trnC-trnD	3500	2	10/21	Demesure *et al.* 1995
ycf6.R	GCC CAA GCR AGA CTT ACT ATA TCC AT	trnC-trnD	-	3	11/21	Shaw *et al.* 2005
ycf6F	ATG GAT ATA GTA AGT CTY GCT TGG GC	trnC-trnD	-	2	8/21	Shaw *et al.* 2005
psbM.R	ATG GAA GTA AAT ATT CTY GCA TTT ATT GCT	trnC-trnD	-	2	8/21	Shaw *et al.* 2005
psbM.F	AGC AAT AAA TGC RAG AAT ATT TAC TTC CAT	trnC-trnD	-	2	10/21	Shaw *et al.* 2005
trnD.F	ACC AAT TGA ACT ACA ATC CC	trnD-trnT	1800	1	1/21	Demesure *et al.* 1995
trnT.1.R	CTA CCA CTG AGT TAA AAG GG	trnD-trnT	1800	1	1/21	Demesure *et al.* 1995
trnE.UUC	5'-AGG ACA TCT CTC TTT CAA GGA G-3'	trnD-trnT	-	1	1/21	Shaw *et al.* 2005
trnY.GUA	5'-CCG AGC TGG ATT TGA ACC A-3'	trnD-trnT	-	1	1/21	Shaw *et al.* 2005
trnH.F	ACG GGA ATT GAA CCC GCG CA	trnH-trnK	2000			Demesure *et al.* 1995
trnK.ex1.R	CCG ACT AGT TCC GGG TTC GA	trnH-trnK	2000			Demesure *et al.* 1995
trnK.ex1.F	GGG TTG CCC GGG ACT CGA AC	trnH-trnK	2500			Demesure *et al.* 1995
trnK.ex2.R	CAA CGG TAG AGT ACT CGG CTT TTA	trnH-trnK	2500			Demesure *et al.* 1995
trnM.F	TGC TTT CAT ACG GCG GGA GT	trnM-rbcL	2900			Demesure *et al.* 1995
rbcL.R	GCT TTA GTC TCT GTT TGT GG	trnM-rbcL	2900			Demesure *et al.* 1995
trnS.1.F	GAG AGA GAG GGA TTC GAA CC	trnS-trnfM	1700	1	4/21	Demesure *et al.* 1995
trnfM.R	CAT AAC CTT GAG GTC ACG GG	trnS-trnfM	1700	1	4/21	Demesure *et al.* 1995
trnS.2.F	CGA GGG TTC GAA TCC CTC TC	trnS-trnT	1500	3	21/21	Demesure *et al.* 1995
trnT.2.R	AGA GCA TCG CAT TTG TAA TG	trnS-trnT	1500	3	21/21	Demesure *et al.* 1995
rpL20	TTTGTTCTACGTCTCCGAGC	rpL20-rpS12	800	3	12/21	Hamilton 1999
rpS12	GTCGAGGAACATGTACTAGG	rpL20-rpS12	800	3	12/21	Hamilton 1999
psbB	GTTTACTTTTGGGCATGCTTCG	psbB-psbF	800			Hamilton 1999
psbF	CGCAGTTCGTCTTGGACCAG	psbB-psbF	800			Hamilton 1999
trnHGUG	ACTGCCTTGATCCACTTGGC	trnH-psbA	500	3	13/21	Hamilton 1999
psbA	CGAAGCTCCATCTACAAATGG	trnH-psbA	500	3	13/21	Hamilton 1999
trnS.GCU	AGA TAG GGA TTC GAA CCC TCG GT	trnS-trnG	800	1	3/21	Shaw *et al.* 2005
3'trnG.UUC	GTA GCG GGA ATC GAA CCC GCA TC	trnG	700	2	9/21	Shaw *et al.* 2005
5'trnG2G	GCG GGT ATA GTT TAG TGG TAA AA	trnG	700	2	9/21	Shaw *et al.* 2005
5'trnG2S	TTT TAC CAC TAA ACT ATA CCC GC	trnS-trnG	800	1	3/21	Shaw *et al.* 2005
5'trnL.UAA.R (	TCT ACC GAT TTC GCC ATA TC	trnT-trnL	700	1	5/21	Shaw *et al.* 2005
trnT.UGU.2F	CAA ATG CGA TGC TCT AAC CT	trnT-trnL	700	1	5/21	Shaw *et al.* 2005
trnE.RC.UUC	CTCCTTGAAAGAGAGATGTCCT	trnD-trnT	-			Designed from seq data
trnT.1.R.M13	CAGGAAACAGCTATGACCCTACCACTGGAGTTAAA	trnD-trnT	-			Designed from seq data
trnT.1.R.674	AAGCGAAAGGTTTGATTTACC	trnD-trnT	-			Designed from seq data
trnY.int	GAATGAATTTAGAAAGAATGAATTTAG	trnD-trnT	-			Designed from seq data
trnCD.F(D2)	dye D2-GGGAAGAAGTGGACTTTAGGG	trnC-trnD	222, 230, 237			Designed from seq data
trnCD.F(D3)	dye D3-GGGAAGAAGTGGACTTTAGGG	trnC-trnD	222, 230, 237			Designed from seq data
trnCD.R	ACATTACTTCGCTCGAATCC	trnC-trnD	222, 230, 237			Designed from seq data
C del.F (D2)	dye D2-ACCTTCACCTAATTTATCAATGTT	trnS-trnT	143, 269			Designed from seq data
C del.F (D3)	dye D3-ACCTTCACCTAATTTATCAATGTT	trnS-trnT	143, 269			Designed from seq data
C del.R	TTGGATAATTTTGCCCCCTA	trnS-trnT	143, 269			Designed from seq data

* Shaw et al. (2005) primer groupings

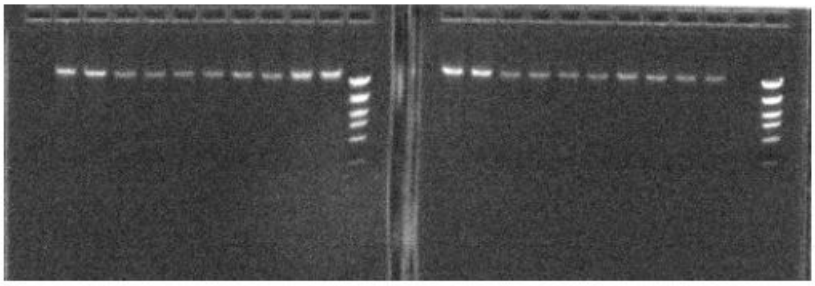
** Shaw et al. (2005) rankings for most valuable primers sets for polymorphism discovery

Results

Plant cultivation and DNA extraction

The germination efficiencies of the castor plants averaged approximately 80% (an average of 4 of 5 seeds germinated) after scarification and humidified incubation. Castor plant seeds required an average of 4-5 days to germinate and an average of 7-10 days to spread cotyledons after germinated seeds were potted (Table 3). DNA extracted from castor leaves using the Qiagen DNeasy Mini Plant Extraction kit was of suitable quality and quantity to facilitate downstream applications such as PCR, restriction enzyme digestion, and sequencing (Figure 4). The purified DNA showed no evidence of degradation as determined by inspection of samples analyzed on agarose gels, and average DNA yields were 1 µg/100 mg leaf tissue when the leaf tissue was disrupted under liquid nitrogen using a mortar and pestle. The average DNA yield doubled when using 2.3 mm stainless steel beads and a bead beater to disrupt leaf tissue that was frozen under liquid nitrogen.

Figure 4. DNA Extracted from Castor Bean Leaves



Total DNA Extracted from Castor bean leaves. All samples were extracted using the Qiagen DNeasy Plant Mini DNA Extraction kit. Purified DNA preparations are shown on 0.8 % agarose gel stained with ethidium bromide and are larger than the 10 kb band in the standard.

Table 3. Castor Bean Seed Germniation and Cultivation Plant Record

ID #	Accession#	Plantid	Name/ Origin	Date incubated	Date germinated	Ratio germinated	Date sprouted	# of plants
LL-1	PI 162912	235	Paraguay	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-2	PI 163162	Var. 39	Brazil	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-4	PI 165446	Higuerilla	Mexico	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-7	PI 167238	Hindyagi Giza No 3	Turkey	3/8/06	3/14/06	5 of 5	3/21/06	3
LL-8	PI 167287	481	Turkey	3/17/06	3/17/06-3/22/06	4 of 5	3/30/06	4
LL-9	PI 167288	482	Turkey	4/6/06	4/10/06	5 of 5	4/16/06	3
LL-10	PI 167342	536	Turkey	4/6/06	4/12/06	5 of 5	4/18/06	3
LL-11	PI 170682	1393	Turkey	4/6/06	4/10/06	4 of 5	4/16/06	3
LL-12	PI 170684	Kahverengi	Turkey	11/14/05	11/17/05	4 of 5	11/28/05	4
LL-14	PI 170686	2334	Turkey	4/6/06	4/10/06	4 of 5	4/16/06	3
LL-15	PI 173090	7569	Turkey	4/6/06	4/10/06	4 of 5	4/16/06	3
LL-17	PI 173795	8750	Turkey	4/6/06	4/10/06	5 of 5	4/16/06	3
LL-18	PI 173946		India	7/26/05	8/3/05	4 of 5	8/11/05	4
LL-19	PI 173 947	India	Errand	4/24/06	4/28/06	4 of 4	5/3/06	4
LL-20	PI 173 948	India	Errand	4/24/06	5/8/2006-5/11/06	2 of 4	5/20/06	2
LL-30	PI 179729	No. 10394 Harind	India	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-39	PI 181916	Hama 21	Syria	8/19/05	8/23/05	5 of 5	8/31/05	5
LL-44	PI 183078	Divela	India	5/8/06	5/11/06	5 of 5	5/20/06	3
LL-46	PI 183347	Divela	India	5/8/06	5/11/06	4 of 5	5/20/06	3
LL-47	PI 183468	Erari	India	5/8/06	5/11/06	4 of 5	5/20/06	3
LL-48	PI 183470	Andi	India	5/8/06	5/11/06	2 of 5	5/20/06	2
LL-51	PI 184133	No. 303	Yugoslavia, Serbia	5/11/06	5/15/06	5 of 5	5/22/06	3
LL-52	PI 192949	Mouthnek	Kenya	7/26/05	7/29/05	4 of 5	8/4/2005-8/9/05	4
LL-54	PI 195811	2835	Guatemala	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-55	PI 197048	3027	El Salvador	1/27/06	1/31/06	5 of 5	2/7/06	3
LL-57	PI 201830	Madagascar	Tige Blanche	8/19/05	8/23/05	4 of 5	8/30/05	4
LL-59	PI 202667	HC 3	India	5/11/06	5/15/06	3 of 5	5/22/06	3
LL-61	PI 202711	1156	Brazil	5/11/06	5/15/06	4 of 5	5/22/06	3
LL-63	PI 203126	TMV-1	India, Delhi	5/11/06	5/15/06	3 of 5	5/22/06	3
LL-64	PI 203128	TMV-3	India, Delhi	5/11/06	5/15/06	4 of 5	5/22/06	3
LL-65	PI 203130	EB 31	India, Delhi	5/11/06	5/15/06	3 of 5	5/22/06	3
LL-66	PI 203661	2147	Paraguay	11/14/05	11/17/05	3 of 5	11/28/05	3
LL-67	PI 204321	Chittaharalu	India	5/11/06	5/15/06	4 of 5	5/22/06	3
LL-68	PI 204322	Rosy Castor	India	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-72	PI 206515		Jamaica	2/13/06	2/16/06	4 of 5	2/21/06	3
LL-74	PI 207868	La Chocolate	Peru	7/26/05	7/29/05	5 of 5	8/4/2005-8/9/05	5
LL-75	PI 208464		Nepal	2/13/06	2/16/06	5 of 5	2/21/06	3
LL-76	PI 208689		Algeria	2/13/06	2/16/06	5 of 5	2/21/06	3
LL-80	PI 208840		Cuba	3/8/06	3/14/06	5 of 5	3/17/06	3
LL-82	PI 208842		Cuba	2/13/06	2/16/06	5 of 5	2/21/06	3
LL-84	PI 209132		Puerto Rico	3/8/06	3/14/06	5 of 5	3/21/06	3
LL-85	PI 209326		Virgin Islands	11/14/05	11/17/05-11/24/05	4 of 5	11/28/05	4
LL-87	PI 209622		Cuba	1/13/06	1/17/06	3 of 5	1/21/06	3
LL-88	PI 212115	12957	Afghanistan	10/25/05	10/27/05	4 of 5	11/2-11/7	4
LL-96	PI 215775	TM 2957	Peru	3/8/06	3/17/2006- 3/21/06	3 of 5	3/22/06	3
LL-98	PI 217539	13933	Pakistan	2/13/06	2/16/06	5 of 5	2/21/06	3
LL-102	PI 219766	Dwarf Castor	Botswana	1/27/06	1/31/06	4 of 5	2/7/06	3
LL-103	PI 219767	No. 464	Argentina, Buenos Aires	8/19/05	8/23/05	3 of 5	8/30/05	3
LL-104	PI 219770	ROUGE DE FEZ	Argentin, Buenos Aires	11/14/05	11/17/05	5 of 5	11/28/05	5
LL-122	PI 221698	No. 1	Indonesia, Java	7/26/05	8/4/05	2 of 5	8/15/05	2
LL-123	PI 222265	1407	Iran	7/26/05	8/4/05	1 of 5	8/15/05	1
LL-135	PI 229785	15783	Iran	3/17/06	3/21/06	4 of 5	3/30/06	4
LL-142	PI 241362	No. 374	Brazil	8/19/05	8/23/05	3 of 5	8/30/05	3
LL-147	PI 241368	No. 383	Brazil	10/25/05	10/27/05	3 of 5	11/2/05	3
LL-150	PI 241371	No. 397	Brazil	3/8/06	3/14/06	5 of 5	3/21/06	3
LL-157	PI 246996		bahamas	10/25/05	10/27/05	3 of 5	11/2-11/3	3
LL-162	PI 248390	K 31	India	11/14/05	12/1/05	3 of 5	12/1/05	3
LL-163	PI 599751	Lynn	US, Texas	7/26/05	8/1/05-8/9/05	3 of 5	8/9/2005-8/15/05	3
LLC-1		Red Castor	WhatcomSeed co	6/29/05	7/5/05-7/8/05	5 of 8	7/11/05-7/13/05	5
LLC-3	lot #G6729	spotted	Mainstreet	7/11/05	7/15/05-7/18/05	8 of 8	7/22/05-7/25/05	8

Targeted sequencing of chloroplast intergenic amplicons: The original strategy to investigate the diversity present among castor plant cultivated varieties was to use universal oligonucleotides to amplify primarily noncoding spacers in a small subset of castor DNAs (Table 1). Amplicons from the preliminary castor DNA panel (Table 1) were sequenced and compared to one another for differences in DNA sequence (Figure 1). Results from previous publications (Chiang et al. (1998), Strand et al. (1997), and Demesure et al. (1995)) supplied the universal primer sequences (Table 2) initially used for this study, and the results using these sources are presented below under the non-ranked loci results section. Based on the initial results and the data presented in Shaw (2005), the strategy changed to focus targeted sequencing on the most highly ranked universal loci (Table 2) from their study. Those results are presented below in the Shaw Tier 1 Ranked loci results section.

A. Non-ranked loci: Universal oligonucleotides with no known homology in *Ricinus communis.*

*atpB-rbc*L: The noncoding spacer between the ATP synthase CF1 beta chain gene and the ribulose 1,5-bisphosphate carboxylase/oxygenase large chain gene (Chiang et al. 1998) was investigated. Chiang et al. reported an evolutionary trend of increasing size of the noncoding region between *atpB* and *rbc*L upon comparison of liverworts, mosses, and vascular plants and suggested that this noncoding spacer may be useful for phylogenetic analyses at the population, subspecific, specific, and generic levels. Oligonucleotide sequences for atpB-1 and rbcL-1 provided by Chiang et al. (1998) were used to amplify this region in the preliminary Castor DNA panel. PCR products were

obtained, but the PCR conditions had to be optimized to minimize non-specific multiple fragment amplification for each DNA template. Successful amplification was achieved after optimization using the following PCR conditions; 95 °C (1 min), followed by 30 cycles of 95 °C (30 s), 51 °C (30 s), 68 °C (2 min), and an additional incubation at 68 °C (7 min) using an ABI 9700 thermal cycler and Clontech Advantage 2 PCR reagents. The resulting amplicons were approximately 1200 bp in length. Purified PCR products were sequenced using atpB-1 and rbcL-1 as sequencing primers, but this region yielded poor sequence data that could not be assembled or cleaned-up. Therefore this region was not further investigated.

G3pdh: Strand et al. (1997) examined the glyceraldehyde 3-phosphate dehydrogenase gene sequence to develop universal primer sets capable of targeting low copy-number nuclear genes and yielding increased information on the genetic diversity present in plant nuclear genomes. Olsen and Schaal also used this region for a species level phylogenetic study of the evolutionary origin of cassava (Olsen and Schaal 1999, Olsen and Schaal 2001, Olsen 2004 and Schaal and Olsen 2000). GPDX7F and GPDX9R oligonucleotides from Strand et al. (1997) were used to amplify this region in the preliminary DNA panel for *Ricinus communis*. Amplification consistently yielded at least 4 PCR products per DNA ranging from 800-2600 bp. Multiple amplicons per DNA were purified from each other by gel extraction and the amplicons were sequenced using GPDX7F and GPDX9R as sequencing primers. The 1300 bp amplicon yielded good quality sequencing data while all other amplicons yielded poor quality sequencing data that could not be analyzed. Forward and reverse reads were assembled together to give complete coverage for these amplicons, but no differences were detected in the sequence

assembly comparing the preliminary castor DNA panel. PCR RFLP was also tested on this locus (see PCR RFLP Results section below).

trnM-rbcL, psbC-trnS, trnH-trnK, psaA-trnS, trnS-trnT:

Oligonucleotide sequences for targeting the noncoding regions between the following chloroplast genes- *trnM-rbcL, psbC-trnS, trnH-trnK, psaA-trnS, trnS-trnT* were obtained from Demesure et al. (1995), who developed these universal primers to discover DNA polymorphisms present in land plants at different taxonomic levels for phylogenetic analyses.

*trnM-rbc*L: The noncoding region between the tRNA gene trnM and the ribulose 1,5-bisphosphate carboxylase/oxygenase large chain gene using trnM.F and rbcL.R oligonucleotides did not yield amplicons in any members of the preliminary castor DNA panel, and this region was not further investigated for the present study.

psbC-trnS: Oligonucleotides psbC.F and trnS.1.R were used to amplify the spacer between the photosystem II 44 kDa protein gene and the tRNA gene trnS. PCR of the preliminary castor DNA panel yielded amplicons of approximately 1700 bp in length, and psbC.F and trnS.1.R were used to generate successful sequencing data, but analysis of the psbC.F_trnS.1.R assembly did not yield any polymorphisms present among members of the preliminary castor DNA panel, and this region was not further investigated for this study.

trnH-trnK: Two sets of oligonucleotides (trnH.F and trnK.ex1.R, trnK.ex1.F and trnK.ex2.R) were used to amplify the noncoding spacer between the tRNA genes trnH and trnK in two amplicons. Both primer sets effectively amplified single amplicons

ranging in size from 2000 bp (trnH.F_trnK.ex1.R) to 2500 bp (trnK.ex1.F_trnK.ex2.R) in all members of the preliminary castor DNA panel, and the same oligonucleotides were used to successfully sequence both amplicons. The trnH.F_trnK.ex1.R assembly yielded 1 polymorphism, an A/T SNP (Table 4). The trnK.ex1.F_trnK.ex2.R assembly yielded 2 T/G SNPs (Table 4).

psaA-trnS: The spacer separating the genes for photosystem I P700 apoprotein A1 and the tRNA trnS were amplified using the oligonucleotides psaA.F and trnS.2.R in all members of the preliminary castor DNA panel, yielding a PCR product approximately 3500 bp in length. The same oligonucleotides were used to sequence this amplicon to reveal 2 A/G SNPs (Table 4).

trnS-trnT: The oligonucleotides trnS.2.F and trnT.2.R were used to target the intergenic region between tRNA genes trnS and trnT. This region yielded amplicons of approximately 1500 bp for all members of the preliminary castor DNA panel, and the PCR primers were used to sequence the amplicons. A 126 bp insertion/deletion (indel) was discovered approximately 100 bp from the 5' end of the trnS.2.F assembly. Based on this substantial polymorphism, additional castor DNAs were sequenced to target this locus. (See Table 4 and the Assay Development Results section for more details).

Table 4. Targeted Sequencing Results Summary

Primer Combinations	Gene Region	Approx. Amplicon Size (bp)	Reference	Shaw Tier	Shaw Rank	Targeted Sequence Results	Castor Cultivars	Polymorphism Reference Sequence
AtpB-1	atpB-rbcL	1200	Chiang *et al.* 1998	-	-	Poor quality sequence data	-	-
rbcL	atpB-rbcL	1200	Chiang *et al.* 1998	-	-		-	-
GPDX7F	G3pdh	1200	Strand *et al.* 1997	-	-	No differences observed	-	-
GPDX9R	G3pdh	1200	Strand *et al.* 1997	-	-		-	-
psaA.F	psaA-trnS	3500	Demesure *et al.* 1995	-	-	A	**1**, 74	(A/G)TGCAACCTAGT(G/A)TATTCCTATC
trnS.2.R	psaA-trnS	3500	Demesure *et al.* 1995	-	-	G	**3**, 52	
psaA.F	psaA-trnS	3500	Demesure *et al.* 1995	-	-	G	**1**, 74	(G/A)TATTCCTATCTTAATTAAAAGTTT
trnS.2.R	psaA-trnS	3500	Demesure *et al.* 1995	-	-	A	**3**, 52	
psbC.F	psbC-trnS	1700	Demesure *et al.* 1995	-	-	No differences observed	-	-
trnS.1.R	psbC-trnS	1700	Demesure *et al.* 1995	-	-		-	-
rpoB	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	A	**1**, 74, 12, 18, 39, 88, 103, 104, 147	(A/C)TGAATTCGGGATTAAATCTTTCGA
trnC.GCA.R	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	C	**3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163	
rpoB	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	T	**1**, 74, 12, 18, 39, 88, 103, 104, 147	(T/G)ATTTTTT(G/T)ATTTTTTTTTTAGCA
trnC.GCA.R	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	G	**3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163	
rpoB	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	T	**1**, 74, 18, 88, 103, 104, 147	(T/G)ATTTTTTTTTTAGCACTTAGCTTT
trnC.GCA.R	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	G	**3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163	
rpoB	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	G	**1**, 74, 12, 18, 103	(T/G)ATTTTTTTTTTAGCACTTAGCTTT
trnC.GCA.R	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	T	**3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163	
rpoB	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	AAAAAAAAAA	**1**, 74, 18, 39, 103	(A/-) AAAAAAAAAGCAAAAAAAAAGG
trnC.GCA.R	rpoB-trnC	1300	Shaw *et al.* 2005	1	2/21	AAAAAAAAA -	**3**, 52, 57, 85, 122, 123, 142, 162, 163	
trnD.F	trnD-trnT	1800	Demesure *et al.* 1995	1	1/21	Poor quality sequence data	-	-
trnT.1.R	trnD-trnT	1800	Demesure *et al.* 1995	1	1/21		-	-
trnH.F	trnH-trnK	2000	Demesure *et al.* 1995	-	-	A	**1**, 74	(A/T)TTTTCTTTTATAGAGAATTCGTGT
trnK.ex1.R	trnH-trnK	2000	Demesure *et al.* 1995	-	-	T	**3**, 52	
trnK.ex1.F	trnH-trnK	2500	Demesure *et al.* 1995	-	-	T	**1**, 74	(G/T)AATTCTTGGTTATTTGAGGTCTTG
trnK.ex2.R	trnH-trnK	2500	Demesure *et al.* 1995	-	-	G	**3**, 52	
trnK.ex1.F	trnH-trnK	2500	Demesure *et al.* 1995	-	-	T	**1**, 74	(G/T)TGTCGAGTGAATAAATGGATAGAC
trnK.ex2.R	trnH-trnK	2500	Demesure *et al.* 1995	-	-	G	**3**, 52	
trnM.F	trnM-rbcL	2900	Demesure *et al.* 1995	-	-	No amplification	-	-
rbcL.R	trnM-rbcL	2900	Demesure *et al.* 1995	-	-		-	-
trnS.1.F	trnS-trnfM	1700	Demesure *et al.* 1995	1	4/21	Poor quality sequence data	-	-
trnfM.R	trnS-trnfM	1700	Demesure *et al.* 1995	1	4/21		-	-
trnS.2.F	trnS-trnT	1500	Demesure *et al.* 1995	3	21/21	- 126 bp insertion	**1**, 74, 12, 18, 39, 88, 103, 104, 147	CCTTGATATAGATTCTCTATTCCTA(+/-1
trnT.2.R	trnS-trnT	1500	Demesure *et al.* 1995	3	21/21	+ 126 bp insertion	**3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163	ACTTAGGCGATAGGGGGCAAAATTA
trnS.GCU	trnS-trnG	800	Shaw *et al.* 2005	1	3/21	A	**1**, 74, 18, 88, 147	(---/AAA) GCCTTTTTCCCTCTTTTCTTTA
5'trnG2S	trnS-trnG	800	Shaw *et al.* 2005	1	3/21	C	**3**, 52, 57, 85, 123, 142, 157, 162, 163	
trnS.GCU	trnS-trnG	800	Shaw *et al.* 2005	1	3/21	AAAAAAAAAAAA	**1**, 74, 18, 104, 147	(A/C)AAGCCTTTTTCCCTCTTTTCTTTA
5'trnG2S	trnS-trnG	800	Shaw *et al.* 2005	1	3/21	AAAAAAAAA - - -	**3**, 52, 57, 85, 123, 142, 157, 162, 163	
3'trnG.UUC	trnG	700	Shaw *et al.* 2005	2	9/21	No differences observed	-	-
5'trnG2G	trnG	700	Shaw *et al.* 2005	2	9/21		-	-
5'trnL.UAA.R (TabB)	trnT-trnL	700	Shaw *et al.* 2005	1	5/21	No differences observed	-	-
trnT.UGU.2F	trnT-trnL	700	Shaw *et al.* 2005	1	5/21		-	-

B. Shaw Tier 1 Ranked Loci: Mid-way through this research project the Shaw et al. (2005) manuscript was published. The paper tested a multitude of primer sets targeting chloroplast noncoding spacers across three closely related species from each of 10 groups representing eight major phylogenetic lineages modified from APG II (2003) to sample different habits and life strategies (e.g., woody perennials, herbaceous perennials, and herbaceous annuals) (Shaw et al. 2005). The results of the Shaw et al. (2005) study ranked 21 regions according the most variable regions producing the highest number of nucleotide substitutions, indels, and inversions. The strategy for the present research for discovery of polymorphisms changed to focus on the 5 top ranking loci (Tier 1 regions).

trnT-trnL: The noncoding spacer between the tRNA genes trnT and trnL is a Tier 1 region, and it is ranked as the 5th most informative region in the Shaw et al. (2005) study. This spacer was amplified and sequenced successfully using the oligonucleotides trnT.UGU.2F and trnL.UAA.R for the preliminary and secondary castor DNA panels. different cultivars for this intergenic region, so subsequent investigation of this region was terminated.

trnS-trnfM: This noncoding region is ranked as the 4th most potentially informative region in the Shaw et al. (2005) study. The oligonucleotides trnS.1.F and trnfM.R were used to amplify and sequence the intergenic region between the tRNA genes trnS and trnfM in the preliminary and secondary castor DNA panels. Upon initial inspection of the trnfM.R assembly, it appeared that cultivars LLC-3, LL-52, and LL-162 showed 4 additional adenine bases inserted into a region approximately 105-135 bp from the 5' end of the assembly, but closer inspection of the chromatograms revealed poor

quality peak heights in this region that could not be resolved with additional sequencing reads, so these differences were not considered significant. Additional DNA polymorphisms were not detected upon comparison of the sequence data for different cultivars for this intergenic region, so subsequent investigation of this region was not continued.

trnS-trnG: The oligonucleotides trnS.GCU and 3'trnG.UUC were used to amplify the noncoding spacer between the tRNA genes trnS and trnG in the preliminary and secondary castor DNA panels. Although the intergenic region spanning from trnS to the 5' end of trnG (sequenced with the oligonucleotides trnS.GCU and 5'trnG2S) was ranked as the 3rd most potentially informative region, the second portion of this amplicon, trnG, was also sequenced as it was originally thought that this region was a part of the Tier 1, 3rd most potentially informative region, when in reality, trnG (sequenced with 5'trnG2G and 3'trnG.UUC) was ranked in Tier 2 as the 9th most potentially informative region. Two DNA polymorphisms including 1 A/C SNP and an SSR of either 9 or 12 adenines were detected in the trnS to the 5' end of trnG region, while no additional polymorphisms were discovered in the trnG region (Table 4).

rpoB-trnC: The second highest-ranking potentially informative region, the noncoding spacer between the RNA polymerase beta chain gene and the tRNA gene trnC was successfully amplified and sequenced in the preliminary and secondary castor DNA panels using the oligonucleotides rpoB and trnD.GCA.R . The amplicons were approximately 1200 bp in length and were purified by gel extraction to remove minimal background amplification observed on the agarose gel. Five DNA polymorphisms were observed upon comparison of the sequencing assembly: an A/C SNP, three T/G SNPs,

and a SSR of either 9 or 10 adenines present (Table 4).

trnD-trnT: Several oligonucleotide combinations were used for amplification and sequencing of the top ranking intergenic region between the tRNA genes trnD and trnT. Although Shaw et al. (2005) states that this spacer amplified easily for most taxa using the primers trnD.F and trnT.1.R, several attempts to amplify this spacer in *Ricinus communis* using these oligonucleotides yielded significant primer dimers in all templates, but no target PCR products. Shaw et al. (2005) used internal oligonucleotides (trnE.UUC and trnY.GUA) to sequence across the trnD-trnT spacer, so these sequencing oligonucleotides were used to amplify this intergenic region in our castor DNAs. The oligonucleotide combination trnD.F and trnE.UUC was used to amplify a 600 bp PCR product, and trnY.GUA and trnT.1.R were used to amplify a 1600-2000 bp PCR product in the preliminary castor DNA panel. The trnD-trnE amplicons exhibited no primer dimers, and only a slight background PCR product approximately 1200 bp in length, which was extremely minimal in intensity when compared with the target 600 bp PCR product, so these amplicons were purified using the Qiaquick PCR clean-up column. The trnY-trnT amplicons showed slight primer dimers, and some background multiple fragments, but yielded a target PCR product of either 1600 bp (for LLC-1 and LL-74) or 2000 bp (for LLC-3 and LL-52). These amplicons were purified by gel extraction. Sequencing of these amplicons was difficult, because they are extremely AT rich consisting of extensive poly-A/T repeats (a consistent finding with Shaw et al. 2005). The trnD-trnE amplicon was sequenced most successfully, but no substantial polymorphisms were detected. The trnY-trnT region was very difficult to sequence, especially at the trnT.1.R end. TrnT.1.R reads that were of sufficient quality to be

analyzed were assembled into 2 contigs (contigs 60 and 61), indicating a major difference at this site. There appeared to be a large rearrangement in this region, but sequencing data was not of high enough quality to identify the polymorphisms present with reasonable confidence. Although all DNAs were not successfully sequenced for these amplicons, and sequencing data was of poor quality, the differential size of PCR products for the trnY-trnT region could be used to score polymorphisms among different castor cultivars (Table 5).

Additional oligonucleotides were designed with the intention of improving sequence data quality and the amount of sequence data for the trnY-trnT region. trnE.RC.UUC was designed as an alternative to trnY.GUA by making the reverse complement of trnE.UUC. Also, using a strategy from (Ghedin et al 2005), a M13 sequence tag was added to the trnT.1.R oligonucleotide to facilitate sequencing from this difficult end. The trnE.RC.UUC and trnT.1.R.M13 amplicons resulted in extensive multiple amplicons that were too difficult to purify via gel extraction because an appropriate target PCR product could not be discerned on the agarose gel.

The subsequent strategy tested for DNA polymorphisms using PCR RFLP (see PCR RFLP Results section below), which yielded a differential digest. Additional oligonucleotides were designed to target the trnY-trnT region. The oligonucleotide trnT.1.R.674 was designed to target the cultivars that assembled into Contig 60, a 674 bp assembly. An analogous oligonucleotide design was attempted for Contig 61, but the primer design software could not find an adequate candidate oligonucleotide for this 386 bp consensus sequence. The oligonucleotide trnY.internal was also designed to amplify the trnY-trnT region in combination with trnT.1.R.674. Single fragment, robust target

amplicons with minimal background resulted from amplifications with these oligonucleotides, but subsequent sequencing with these oligonucleotides did not yield a significant improvement in sequencing read length or sequence quality, and sequencing strategies to identify and verify the polymorphisms present in this region were discontinued.

Table 5. Region trnD-trnT PCR Results using oligonucleotides trnY.GUA and trnT.1.R

DNA Template	Amplicon Size (bp)
LLC-1	1600
LLC-3	2000
LL-18	1600
LL-39	1600
LL-52	2000
LL-57	2000
LL-74	1600
LL-103	1600
LL-122	2000
LL-123	2000
LL-142	2000
LL-163	2000

PCR Restriction Fragment Length Polymorphism: PCR amplification of noncoding cpDNA spacers was followed by digestion of amplicons with restriction enzymes possessing AT rich recognition sequences to explore DNA polymorphisms present in loci that were difficult to amplify and sequence using the targeted sequencing approach, in addition to loci from Hamilton (2000), and from Shaw (2005) Tier 2 regions (see Table 6 for all PCR RFLP Results).

A. PCR RFLP of targeted sequencing problematic primer sets

G3pdh: The nuclear gene glyceraldehyde 3-phosphate dehydrogenase was

amplified in the preliminary and secondary castor DNA panels using oligonucleotides GPDX7F and GPDX9R and yielded four amplicons per template. *Mse*I successfully digested the purified amplicons, but no differential banding patterns were observed upon analysis of digests of different templates on 4% agarose gels.

trnD-trnT: The noncoding spacer between the tRNA genes trnD and trnT was amplified in the preliminary and secondary castor DNA panels using oligonucleotides trnY.GUA and trnT.1.R and yielded multiple amplicons per template. Mse I successfully digested the purified amplicons, and differential restriction patterns were observed upon analysis of digests of different templates on 4% agarose gels.

B. PCR RFLP of Hamilton loci

psbA-trnH: Oligonucleotides psbA and trnH.GUG were used to amplify the spacer between the gene encoding photosystem II protein D1 and the tRNA gene trnH in the preliminary and secondary castor DNA panels. Purified amplicons were approximately 600bp in length and digested with *Mse*I, *Apo*I, and *Sau*3A. *Mse*I digests yielded no differences in restriction patterns between different amplicons, and *Apo*I, and *Sau*3A did not cut the psbA-trnH.GUG amplicons.

rpl20-rps12: Oligonucleotides rpL20 and rpS12 were used to amplify the noncoding region between the genes for ribosomal proteins L20 and S12 in the preliminary and secondary castor DNA panels in addition to LL-55, LL-75, LL-76, LL-87, LL-98, and LL-102 DNA templates. Purified amplicons were approximately 800 bp in length and were successfully digested with *Mse*I. No differential restriction patterns were observed among different template digests. A subset of the amplicons (LL-1, LL-

52, LL-74, LL-39, LL-88, and LL-102) was also digested separately with *Apo*I, *Ase*I, and *Dra*I. No differences in fragment pattern were observed for these combinations of templates and restriction enzymes.

psbB-psbF: Oligonucleotides psbB and psbF were used to amplify the spacer between the cytochrome b559 beta chain gene and the gene encoding photosystem II 47 kDa protein in the preliminary and secondary castor DNA panels in addition to LL-55, LL-75, LL-76, LL-87, LL-98, and LL-102 DNA templates. Purified amplicons were approximately 600 bp in length and were digested with *Mse*I, but no differential restriction patterns were observed upon comparison of digests of different DNA templates. A subset of the amplicons (LL-1, LL-52, LL-74, LL-39, LL-88, and LL-102) was also digested separately with *Apo*I, *Ase*I, and *Dra*I. Restrictions with *Apo*I and *Ase*I yielded no differences in restriction banding patterns, and *Dra*I did not cut any of the psbB-psbF amplicons.

C. PCR RFLP of Shaw Tier 2 Regions

rpL16: The intron sequence for the ribosomal protein L16 gene was amplified using the oligonucleotides rpL16F71 and rpL16R1516 for the preliminary and secondary castor DNA panels. The amplicons were approximately 1000 bp in length, and purified PCR products were successfully digested with *Mse*I. No differences in restriction patterns were observed upon comparison of digests from different template amplicons.

rpS16: The intron sequence for the ribosomal protein S16 gene was amplified using the oligonucleotides rpS16.F and rpS16.R for the preliminary and secondary castor DNA panels. Resultant amplicons were approximately 850 bp in length, and purified

PCR products were successfully digested with *Mse*I to yield differential restriction patterns for different amplicon templates. The rpS16 region was targeted for sequencing to identify the type of polymorphism present to account for the RFLP patterns, but sequencing with rpS16.F and rpS16.R yielded poor quality sequencing data that could not be assembled and analyzed.

trnC-trnD: The intergenic region between tRNA genes trnC and trnD was amplified using the oligonucleotides trnC.F and trnD.R. The amplicons were approximately 3500 bp in length and purified amplicons were digested with *Mse*I. Upon comparison of different amplicon template digests, differential restriction fragment patterns were observed. Oligonucleotides ycf6.F and psbM.R were also used to amplify a portion of the trnC.F-trnD.R amplicon. The ycf6.F-psbM.R PCR product was approximately 1000 bp in length, and purified amplicons were also digested with *Mse*I, and yielded restriction fragment patterns that corresponded with the differences observed upon digestion of the trnC.F-trnD.R region. The trnC-trnD region (particularly between the oligonucleotides ycf6.F and psbM.R) was selected for targeted sequencing to identify the polymorphism responsible for the differential restriction patterns. Oligonucleotides trnC.F and trnD.R were used to amplify this region in the preliminary castor DNA panel, and ycf6.F and psbM.R were used as sequencing oligonucleotides. Analysis of these sequencing data revealed a 7 bp indel that appeared to be part of a VNTR. Cultivars possessed 1 or 2 copies of this 7 bp imperfect repeat, and although the polymorphism did not exist within an *Mse*I recognition site, the number of 7 bp repeats altered the length of one of the fragments generated by the *Mse*I digestion.

Table 6. PCR RFLP Results Summary

Gene Region	Primer Combination	Approx. Amplicon Size	Reference	Tier	Rank	Enzyme(s)	Enzyme(s) Cut Amplicons	Differential Digest	Subsequent Research
G3pdh	GPDX7F, GPDX9R	1200	Strand *et al.* 1997			Mse I	+	-	none
psbA-trnH	psbA, trnHGUG	500	Hamilton 1999	3	13/21	Mse I, Apo I, Sau3A	+ (except Apo I)	-	none
trnD-trnT	trnY.GUA, trnT.1.R	1600, 1800	Demesure *et al.* 1995	1	1/21	Mse I	+	+	sequenced
rpl20-rps12	rpL20, rpS12	800	Hamilton 1999	3	12/21	Mse I, Apo I, Ase I, Dra I	+	-	none
psbB-psbF	psbB, psbF	800	Hamilton 1999			Mse I, Apo I, Ase I, Dra I	+ (except Dra I)	-	none
rpL16	rpL16F71, rpL16R1516	1000	Shaw *et al.* 2005	2	7/21	Mse I	+	-	none
rpS16	rpS16.F, rpS16.R	850	Shaw *et al.* 2005	2	6/21	Mse I	+	+	sequenced
trnC-trnD	trnC.F, trnD.R	3500	Demesure *et al.* 1995	3	11/21	Mse I	+	+	sequenced, assay designed
trnC-trnD	ycf6.F, psbM.R	1000	Shaw *et al.* 2005	2	8/21	Mse I	+	+	sequenced, assay designed

Assay Development: The trnS-trnT region that yielded the 126 bp indel, and the spacer between trnC-trnD that yielded the 7 bp VNTR were selected as optimal candidates for assay development.

A. trnS-trnT Assay Development: C del Assay

Primer design. Oligonucleotides were designed using Primer 3 software (Rozen and Skaletsky 2000) to target the 126 bp indel located between trnS and trnT with a PCR-based fragment analysis assay for the CEQ 8000. The amplicons were designed to be either 143 bp or 269 bp in length.

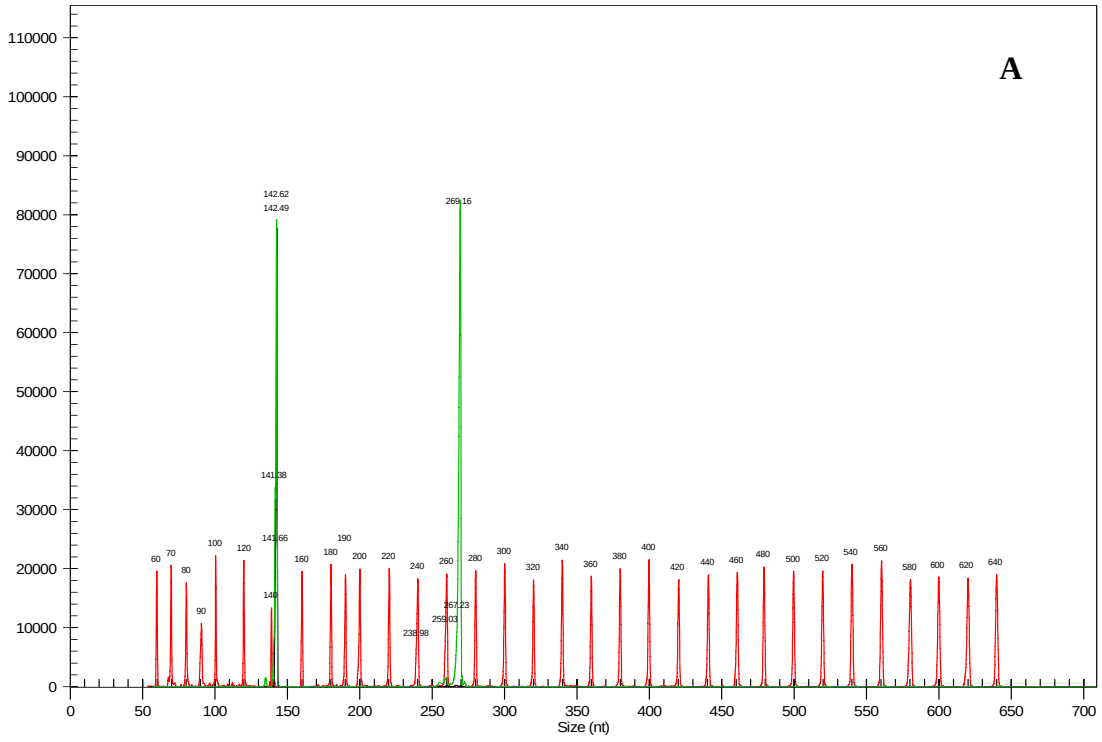
Amplification. The preliminary castor DNA panel was amplified using the oligonucleotides C del F (D2) and C del R. Similar amounts of PCR products in the expected approximate size range (143 bp-269 bp) were observed on 2% agarose gels with no background amplification or primer dimers present.

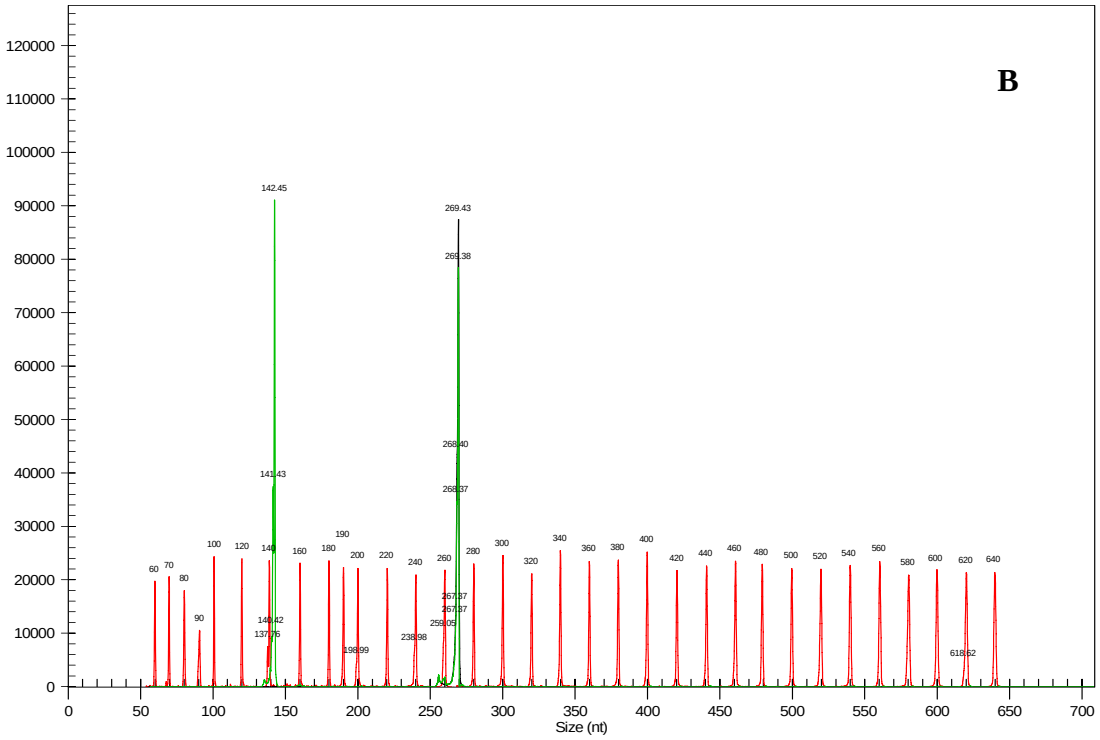
Internal standards. LLC-1 and LLC-3 were selected as templates for development of internal standards to be run with each sample on the capillary electrophoresis instrument. Standard PCR amplification and PCR RFLP verified that the LLC-1D internal standard amplicon would be 143 bp, and the LLC-3 internal standard amplicon would be 269 bp. The oligonucleotides Cdel.F (D3) and Cdel.R were used to amplify the internal standards. Approximate amounts of amplicons were the same for both templates, and there was no primer dimer or non-specific amplification observed upon analysis on 2% agarose gels.

Results. C del assays run on the preliminary castor DNA panel confirmed the differences observed previously using standard PCR and PCR RFLP. Sixty castor

cultivars were screened for the indel using the C del assay (Figure 5 and Table 7). Twenty-three cultivars contained the deletion at the trnS-trnT intergenic spacer, while 37 cultivars possessed the 126 bp insertion at this location. Triplicate runs for each template were consistent with average size fragments of either 142.66 bp or 269.50 bp with average standard deviations for triplicate runs of 0.06 bp or 0.10 bp, respectively. Internal standards were also consistent between runs with average fragment sizes of 142.54 bp or 269.23 bp with standard deviations of 0.13 bp and 2.33 bp, respectively.

Figure 5. C del Assay Fragment Analysis Chromatograms





C del Fragment Analysis Assay (Figure 5 continued). A. Castor bean cultivar LL-104 with a fragment size of 143 bp. **B.** Castor bean cultivar LL-72 with a fragment size of 269 bp. Internal standards are shown in green at 143 bp and 269 bp. Size standard is shown in red.

Table 7. C Del CEQ Fragment Analysis Assay Results

Cultivar	126 bp Insertion	Cultivar	126 bp Insertion	Cultivar	126 bp Insertion
LL-1	+	LL-48	-	LL-84	+
LL-2	+	LL-51	-	LL-85	+
LL-4	-	LL-52	+	LL-87	+
LL-7	-	LL-54	+	LL-88	+
LL-8	-	LL-55	+	LL-96	+
LL-9	-	LL-57	+	LL-98	+
LL-10	-	LL-59	+	LL-102	+
LL-11	-	LL-61	+	LL-103	-
LL-12	-	LL-63	+	LL-104	-
LL-14	-	LL-64	-	LL-122	+
LL-15	+	LL-65	+	LL-123	+
LL-17	+	LL-66	+	LL-135	-
LL-18	-	LL-67	+	LL-142	+
LL-19	+	LL-68	+	LL-147	-
LL-20	+	LL-72	+	LL-150	+
LL-30	-	LL-74	-	LL-157	+
LL-39	+	LL-75	+	LL-162	+
LL-44	-	LL-76	-	LL-163	+
LL-46	-	LL-80	+	LLC-1	-
LL-47	-	LL-82	+	LLC-3	+

B. trnC-trnD Assay Development: trnCD Assay

Primer design. Oligonucleotides were designed to target the 7 bp VNTR with a PCR based fragment analysis assay for the CEQ 8000. These amplicons were designed to be 230 bp if one copy of the 7 bp repeat was present, or 237 bp if two copies of the 7 bp repeat were present.

Amplification. DNA from the preliminary castor panel was amplified using the oligonucleotides trnCD.F (D2) and trnCD.R. Analysis of amplicons on 2% agarose gels showed similar amounts of the PCR products for different DNAs, amplicons in the correct size range at approximately 200 bp, and no primer dimers or background amplification present.

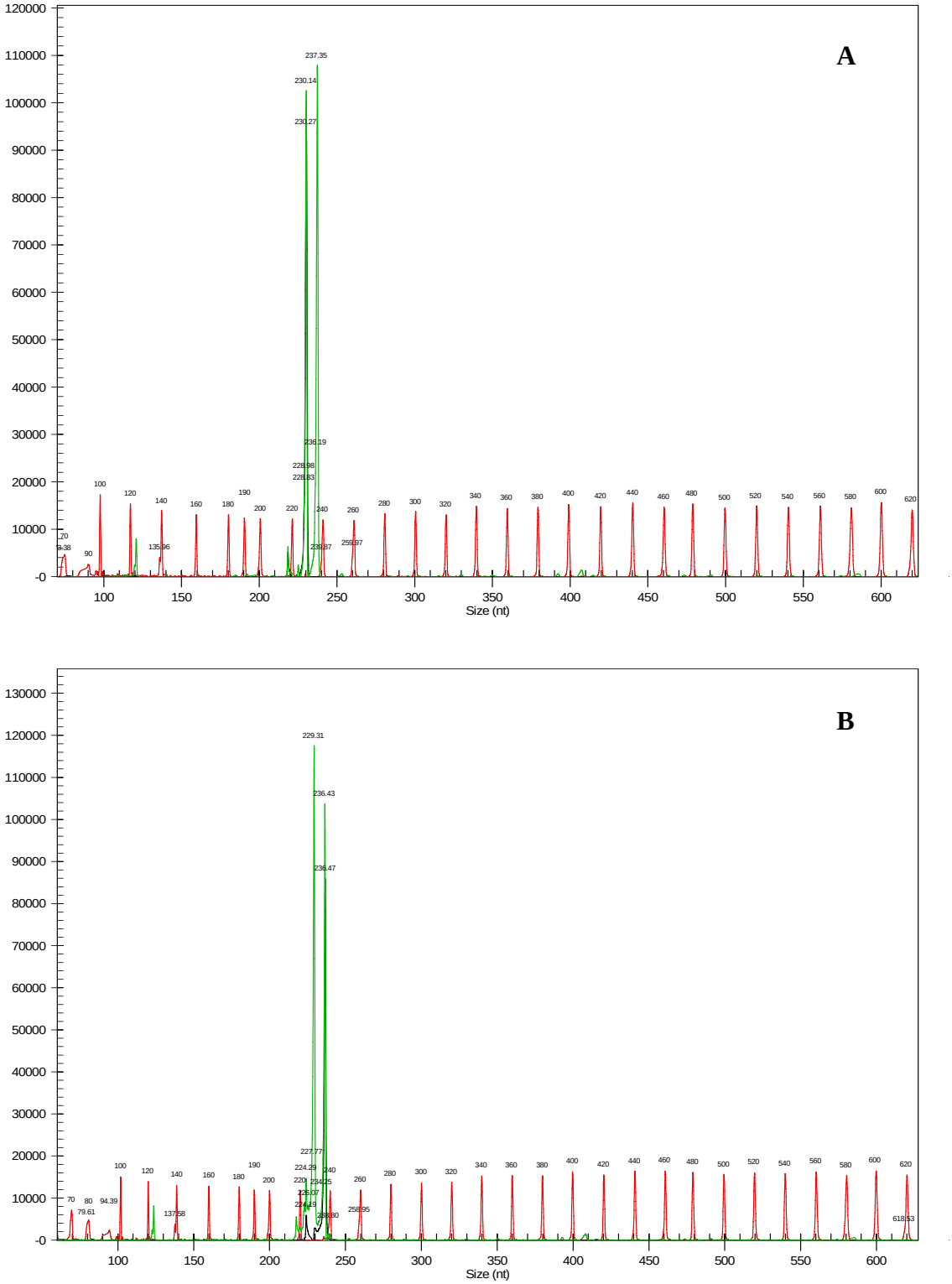
Internal Standards. LLC-1 and LLC-3 were selected as templates for development of internal standards to be run with each sample on the capillary electrophoresis instrument. Preliminary sequencing data showed that LLC-1 possessed 1 copy of the VNTR, while LLC-3 possessed two copies of the 7 bp repeat. trnCD.F (D3) and trnCD.R were used to amplify the internal standards. Approximate amounts of the amplicons were the same for both templates, and there were no primer dimers or non-specific amplification observed.

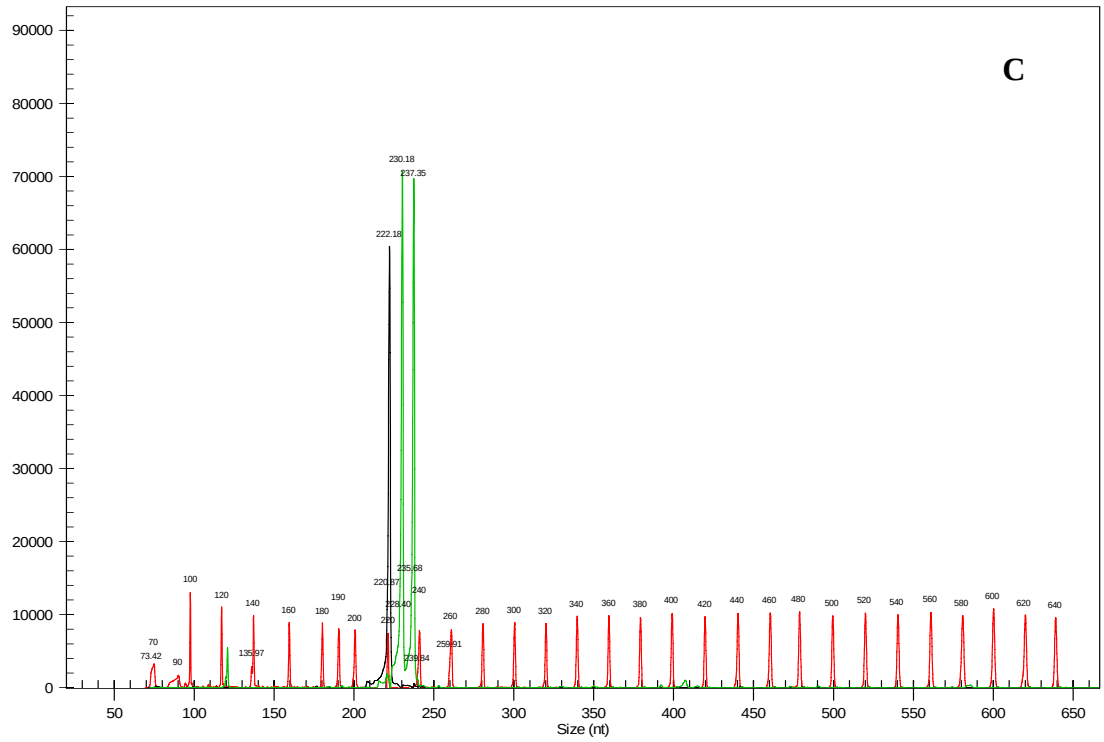
Results. Assays run on the preliminary castor DNA panel confirmed the polymorphisms discovered in the targeted sequencing results. LLC-1 and LL-74 showed peaks at 230 bp, and LLC-3 and LL-52 showed peaks at 237 bp. The trnCD assay was tested on 56 additional DNAs (Figure 6 and Table 8). 23 cultivars possessed 1 repeat of the 7 bp VNTR, while 35 cultivars possessed 2 repeats of the 7 bp VNTR. Triplicate runs for each template were consistent with average sizes of either 229.55 bp or 236.88

bp observed with average standard deviations for triplicate runs of 0.38 bp and 0.26 bp, respectively. Internal standards were also consistent between runs with average sizes of either 229.57 bp or 236.70 bp with standard deviations of either 0.54 bp or 0.58 bp, respectively. The assay appeared to work well, but there were unexpected results for two DNA templates.

Cultivars LL-39 and LL-88 yielded fragment sizes of approximately 222 bp when screened with the CD Assay on the CEQ (Figure 6 and Table 8). The average size observed for this allele was 221.56 bp with an average standard deviation of 0.45 bp. These templates were targeted for sequencing to determine the sequence polymorphism responsible for the third allele observed using the CD Assay. Oligonucleotides trnC.F and trnD.R were used to amplify this region in LL-39 and LL-88, and the ycf6.F oligonucleotide was used to sequence the purified PCR product. Sequence analysis revealed an 8 bp deletion at a second locus located 63 bp downstream from the 7 bp VNTR (Figure 7). Internal standards were also developed to represent this third allele, and the 222 bp internal standard will be run with subsequent CD Assays.

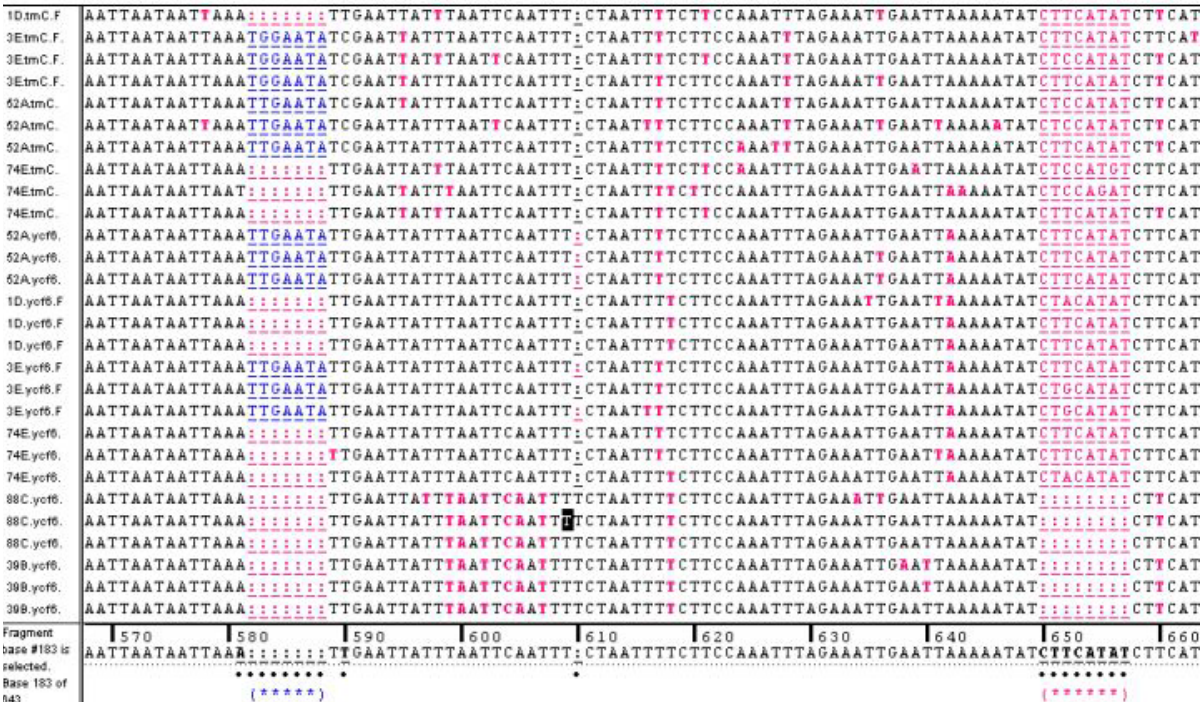
Figure 6. trnCD Assay Fragment Analysis Chromatograms





trnCD Fragment Analysis Assay (Figure 6 continued). A. Castor cultivar LL-135 with a fragment size of 230 bp. **B.** Castor cultivar LL-68 with a fragment size of 237 bp. **C.** Castor cultivar LL-88 with a fragment size of 222 bp. Internal standard peaks are shown in green at 230 bp and 237 bp. Internal size standard peaks are shown in red.

Figure 7. Sequence Assembly for verification of the 222bp allele in the trnCD Assay



Sequence identification of the polymorphism present in Castor cultivars with the 222 bp allele in the trnCD Fragment Analysis Assay. Castor cultivars LL-39 and LL-88 have an additional 8 bp deletion approximately 60 bp downstream from the 7 bp VNTR targeted in the trnCD Assay.

Table 8. trnCD CEQ Fragment Analysis Assay Results

Cultivar	D2 Peak Interpreted (bp)	Cultivar	D2 Peak Interpreted (bp)	Cultivar	D2 Peak Interpreted (bp)
LL-1	**237**	LL-48	230	**LL-84**	**237**
LL-2	**237**	LL-51	230	**LL-85**	**237**
LL-4	230	LL-52	237	**LL-87**	**237**
LL-7	230	**LL-54**	**237**	LL-88	222
LL-8	230	**LL-55**	**237**	**LL-96**	**237**
LL-9	230	**LL-57**	**237**	**LL-98**	**237**
LL-10	230	**LL-59**	**237**	**LL-102**	**237**
LL-11	230	**LL-61**	**237**	LL-103	230
LL-12	230	**LL-63**	**237**	LL-104	230
LL-14	230	LL-64	230	**LL-122**	**237**
LL-15	**237**	**LL-65**	**237**	**LL-123**	**237**
LL-17	**237**	**LL-66**	**237**	LL-135	230
LL-18	230	**LL-67**	**237**	**LL-142**	**237**
LL-19	**237**	**LL-68**	**237**	LL-147	230
LL-20	**237**	**LL-72**	**237**	**LL-150**	**237**
LL-30	230	LL-74	230	**LL-157**	**237**
LL-39	222	**LL-75**	**237**	**LL-162**	**237**
LL-44	230	LL-76	230	**LL-163**	**237**
LL-46	230	**LL-80**	**237**	LLC-1	230
LL-47	230	**LL-82**	**237**	**LLC-3**	**237**

Discussion

The present research effort accomplished many of the original research objectives, but continued research will be required to fully investigate and discover the genetic diversity present among *Ricinus communis* cultivars.

Plant cultivation and DNA extraction

The research objectives of successful plant cultivation and DNA extraction were achieved in this project. Consistent quantities of high quality DNA were easily extracted from young castor plants after germination and growth taking 3-4 weeks. The seed germination efficiency was sufficient for this work, but fungal growth on the scarified seeds was commonly observed and seems to be a pervasive problem. Prior surface sterilization with 10% (v/v) commercial bleach should reduce this problem significantly. Germination efficiencies could also be improved by optimized scarification and incubation conditions (Ellis et al. 1985). Although the time frame for growth and harvest of leaves for DNA extraction was acceptable for the current study, use of this cultivation and DNA extraction method would not be sufficient for genotyping castor seed evidence in a forensic setting in a timely manner. Ideally, DNA extraction directly from seeds would allow for quick genotyping of this evidence. DNA extraction directly from seeds has proven to be challenging in the past due to high levels of oil present in the seed. The resultant DNA was typically degraded and of poor enough quality to inhibit PCR or digestions with restriction enzymes. Future methods that quickly extract consistent quantities of high quality DNA directly from castor seeds and ricin preparations would be

extremely useful for genotyping forensic castor evidence and for streamlining the process to obtain genetic material for castor research.

Targeted sequencing approach

The targeted sequencing approach to investigate the diversity present among *Ricinus communis* cultivars was an appropriate preliminary strategy for this research considering the limited availability of background information regarding the diversity present among cultivars. This approach yielded comparable numbers and types of polymorphisms regardless of whether the investigated loci were ranked according to the Shaw et al. (2005) manuscript (Table 9)**.** The non-ranked loci yielded 5 SNPs and a 126 bp indel. The low copy nuclear gene G3PDH did not yield any differences among cultivars. Investigation of the top ranked Shaw et al. (2005) loci also yielded 5 SNPs and 2 SSRs. In addition, the D/T locus appeared to contain a large rearrangement, but this could not be verified with consistent, good quality sequencing data. Although the targeted sequencing approach yielded some polymorphisms, this system proved to be too time consuming and costly to continue once the low level of diversity present among cultivars was discovered.

Table 9. Summary of Mutations Discovered and Cultivar Groups

Gene Region	Oligonucleotide	Polymorphism Observed	Cultivar Groups
trnH-trnK	trnH.F trnK.ex1.R	A T	**1**, 74 **3**, 52
trnH-trnK	trnK.ex1.F trnK.ex2.R	T G	**1**, 74 **3**, 52
trnH-trnK	trnK.ex1.F trnK.ex2.R	T G	**1**, 74 **3**, 52
psaA-trnS	psaA.F trnS.2.R	A G	**1**, 74 **3**, 52
psaA-trnS	psaA.F trnS.2.R	G A	**1**, 74 **3**, 52
trnS-trnG	trnS.GCU 5'trnG2S	A C	**1**, 74, 18, 147, 88 **3**, 52, 57, 85, 123, 142, 157, 162, 163
trnS-trnG	trnS.GCU 5'trnG2S	AAAAAAAAAAAA AAAAAAAAA - - -	**1**, 74, 18, 104, 147 **3**, 52, 57, 85, 123, 142, 157, 162, 163
rpoB-trnC	rpoB trnC.GCA.R	A C	**1**, 74, 12, 18, 39, 103, 104, 147, 88 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163
rpoB-trnC	rpoB trnC.GCA.R	T G	**1**, 74, 12, 18, 39, 103, 104, 147, 88 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163
rpoB-trnC	rpoB trnC.GCA.R	T G	**1**, 74, 18, 103, 104, 147, 88 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163
rpoB-trnC	rpoB trnC.GCA.R	G T	**1**, 74, 12, 18, 39, 103 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163
rpoB-trnC	rpoB trnC.GCA.R	AAAAAAAAAA AAAAAAAAA -	**1**, 74, 18, 39, 103 **3**, 52, 57, 85, 122, 123, 142, 162, 163
trnD-trnT	trnY.GUA trnT.1.R	Differential Restriction Digest	39, 88 **1**, 74, 12, 18, 103, 104, 147, **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163
rpS16	rpS16.F rpS16.F	Differential Restriction Digest	**1**, 74, 12, 39, 103, 104, 147, 88 **3**, 52, 66, 85, 122, 123, 142, 157, 162, 163
trnS-trnT	C Del.F (D2) C Del.R	- 126 bp insertion - 126 bp insertion + 126 bp insertion + 126 bp insertion + 126 bp insertion	**1**, 74, 12, 18, 103, 104, 147, 4, 7, 8, 9, 10, 11, 14, 30, 44, 46, 47, 48, 51, 64, 76, 135 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163 1, 2, 15, 17, 19, 20, 39, 54, 55, 59, 61, 63, 65 ,67, 68, 72, 75, 80, 82, 84, 87, 96, 98, 102, 150, 88
trnC-trnD	trnCD. F (D2) trnCD. R	222 bp 230 bp 230 bp 237 bp 237 bp 237 bp	39, 88 **1**, 74, 12, 18, 103, 104, 147, 4, 7, 8, 9, 10, 11, 14, 30, 44, 46, 47, 48, 51, 64, 76, 135 **3**, 52, 57, 66, 85, 122, 123, 142, 157, 162, 163 1, 2, 15, 17, 19, 20, 54, 55, 59, 61, 63, 65 ,67, 68, 72, 75, 80, 82, 84, 87, 96, 98, 102, 150,

PCR RFLP approach

The PCR RFLP approach proved to be more cost effective and efficient for polymorphism discovery. PCR RFLP verified the diversity present in the trnD/T region, and differential digests in the trnCD region ultimately led to assay development for the VNTR present in this region, proving to be the highest value polymorphism discovered in this project (Table 9). The combined targeted sequencing and PCR RFLP effort yielded 10 SNPs, 3 indels, 1 VNTR, and 2 differential digests that could not be verified with sequencing (Table 9). An unexpected result proved to be that there does not appear to be a positive correlation between the top ranked Shaw et al. (2005) markers yielding higher levels of polymorphisms in *Ricinus communis* cultivars. This is surprising considering the diverse range of species tested in the Shaw et al. (2005) study.

Assay Development

Both the Cdel and the trnCD PCR-based fragment analysis assays show promise for utility in a research setting and for the field of forensics. Both assays target small chloroplast amplicons, which are extremely valuable for genotyping challenging forensics samples that may contain degraded DNA. The assays are rapid, with the potential to be high throughput, and the assay results are highly consistent with excellent repeatability. The C del assay sorts the present castor DNA sampling into 2 groups, which is not unexpected considering the same trend was observed for the majority of the castor polymorphisms. The C del assay provided a method of rapidly screening a large number of castor cultivars, and the assay may still prove useful as a preliminary screen for forensic samples. The trnCD assay performed similarly to the C del assay but,

unexpectedly, sorted the castor cultivar collection into three groups. All other markers only generated two distinguishable haplotypes.

Although only two polymorphisms were developed into assays, all polymorphisms discovered in this study may still prove useful for elucidation of castor genetic diversity. Many SNPs were only tested against the preliminary castor panel representing 4 individuals due to time constraints. These SNPs may sort the castor representatives into different groupings, depending on the amount and type of new castor cultivars tested in the future.

Genetic Diversity of *Ricinus communis* cultivated varieties

The results of the investigation of *Ricinus communis* non-coding spacers reveal that most of the plant universal oligonucleotides successfully amplified these regions in *Ricinus communis* cultivars, but the results indicate that the chloroplast sequence is highly conserved among cultivars. All of the polymorphisms discovered in this study, except for the trnCD assay, sort the castor plants into the same two groups, providing very little resolution for genotyping forensic castor samples or constructing a detailed phylogeographic tree. Although very little genetic diversity was discovered in the present study, it is promising that some phenotypic variation, such as leaf shape and size, stem color, and seed pod texture, was observed throughout cultivation of different castor cultivars. The lack of overall genetic variation among the *R. communis* cultivars reflected by the grouping of all castor cultivars tested into only two clades may support a previously proposed hypothesis that all castor varieties originated from a limited genetic pool of Old World plants from either Africa or Asia. If castor cultivars propagated and

traded by humans for their oil production originated from a limited sampling of Old World wild plants, this would account for the lack of genetic diversity observed in plants collected from diverse geographic locations throughout the world. The USDA panel used in this study may be biased by over-representing cultivars that were introduced in diverse regions of the world because of human propagation rather than sampling wild plants where *R. communis* is known to be endemic. The samples tested in this study, however, are representative of seeds and plants that are readily accessible to individuals who might be interested in acquiring them for nefarious purposes.

A continued research effort is certainly needed to fully elucidate the genetic diversity present among *Ricinus communis* cultivars. Existing assays should be optimized for real world conditions and tested on castor bean material found in a variety of forms (e.g., plants, seeds, ricin preparations). Of course, discovery of additional markers that segregate the cultivars into different groups is also necessary, and possible subdivision of the *R. communis* species into two sub-species should also be addressed once a more complete characterization of the genetic diversity among cultivars is accomplished. Low-copy and single copy nuclear genes may also prove to be useful targets for investigation of DNA polymorphisms among castor cultivars. As more *R. communis* genetic sequence is accessible through Genbank, investigation of these low copy nuclear regions will become possible. In addition, AFLP comparison of castor cultivars should provide a whole genome approach to identification of DNA polymorphisms present between highly similar castor cultivar genomes, and forensically appropriate assays could be developed to target the discovered DNA polymorphisms (this work is in progress at Lawrence Livermore National Laboratory). In short, determination

of the genetic diversity present among *Ricinus communis* cultivated varieties remains a valuable objective for the forensic community, and although the present study proved to be an important preliminary sampling of the castor diversity, future research using alternate experimental strategies is necessary to gain an understanding of the level of diversity present in *Ricinus communis*, and to exploit this diversity for forensic and biodefense purposes.

References

Armstrong, W. (March/April 1999). "Noteworthy Plants. Castor Bean: A Plant named after a tick." from http://waynesword.palomar.edu/plmar99.htm#intro.

Avise, J.C., Arnold, J., Ball, R.M, Bermingham, E., Lamb, T., Neigel, J.E. Reeb, C.A. and Saunders, N.C. (1987). "Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics." Ann. Rev. Ecol. Syst. 18: 489-522.

Brown, D. (2005). "Cornell University Poisonous Plants Informational Database. Ricinus communis." from http://www.ansci.cornell.edu/plants/castorbean.html#description.

Carus, W. Seth (1998). "Bioterrorism and Biocrimes. The Illicit Use of Biological Agents Since 1900. Center for Counterproliferation Research, National Defense University, Washington, D.C.

CDC. (2004). "Center for Disease Control Facts Sheet: Ricin." from http://www.bt.cdc.gov/agent/ricin/facts.asp.

Chiang, T.-Y., Schaal, B.A. and Peng, C.-I. (1998). "Universal primers for amplification and sequencing a noncoding spacer between the atpB and rpcL genes of chloroplast DNA." Bot. Bull. Acad. Sin. 39: 245-250.

Coyle, H. M. Forensic Botany. Principles and applications to criminal casework. CRC press, 2005.

Demesure, B., Sodzi, N. and Petit, R.J. (1995). "A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants." Mol Ecol 4(1): 129-131.

Dommelen, J. (2002). "Forensic Botany: Molecular Biology." from http://myweb.dal.ca/jvandomm/forensicbotany/molecularbiology.html.

Duke, J.A. (1998). " *Ricinus communis*." from Purdue University New Crop Resource Online Program. http://www.hort.purdue.edu/newcrop/duke_energy/Ricinus_communis.html.

Ellis, R., Hong, T. and Roberts, E. (1985). "Euphorbiaceae. Handbook of Seed Technology for Genebanks - Volume II. Compendium of Specific Germination Information and Test Recommendations." from http://www.ipgri.cgiar.org/publications/HTMLPublications/52/ch22.htm.

Franz, D. and Jaax, N (1997). "Ricin Toxin. Textbook of Military Medicine: Medical Aspects of Chemical and Biological Warfare." from http://www.nbc-

med.org/SiteContent/HomePage/WhatsNew/MedAspects/Ch-32electrv699.pdf#search='Ricin%20toxin%20Chapter%2032'.

Elodie Ghedin, E., Sengamalay, N.A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D.J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y., St George, K., Taylor, J., Lipman, D.J., Fraser, C.M., Taubenberger, J.K., and Salzberg1,S.L. (2005). "Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution." Nature 437(7062): 1162-1166.

Gielly, L. and P. Taberlet (1994). "The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences." Mol Biol Evol 11(5): 769-777.

Griffiths, G. D., Leek, M. D., Gee, D. J. (1987). The toxic plant proteins ricin and abrin induce apoptotic changes in mammalian lymphoid tissues and intestine. J. Pathol. 151: 221-229.

Hamilton, M. B. (1999). "Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation." Mol Ecol 8(3): 521-523.

Hanson, D. (March/April 2004). "Ricin Toxin: Law Enforcement's New Challenge. Police and Security News." from http://www.policeandsecuritynews.com/marApr04/RicinToxin.htm.

Huelseweh, B., Ehricht, R., and Marschall, H.J. (2006). "A simple and rapid protein array based method for the simultaneous detection of biowarfare agents." Proteomics 6(10): 2972-2981.

Lubelli, C., Chatgilialoglu, A., Boloognesi, A., Strocchi, P., Colombatti, M., and Stirpe, F.(2006). "Detection of ricin and other ribosome-inactivating proteins by an immuno-polymerase chain reaction assay." Anal Biochem. 355(1): 102-109.

Mirarchi, F. and Allswede, M. (2006). "Emergency Medicine. Chemical, Biological, Radiological, Nuclear And Explosives: Ricin." from http://www.emedicine.com/emerg/topic889.htm.

Morris, R. (1996-2003). "Plants for a Future. Ricinus communis." from http://www.ibiblio.org/pfaf/cgi-bin/arr_html?Ricinus+communis.

Olsen, K. and B. Schaal (2001). "Microsatellite variation in cassava (Manihot esculenta, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication." Am J Bot 88(1): 131-142.

Olsen, K. M. (2004). "SNPs, SSRs and inferences on cassava's origin." Plant Mol Biol 56(4): 517-526.

Olsen, K. M. and B. A. Schaal (1999). "Evidence on the origin of cassava: phylogeography of Manihot esculenta." Proc Natl Acad Sci U S A 96(10): 5586-5591.

Petit, R. J., Demesure, B., Dumolin-Lapegue, S. (1996). "Molecular tools for screening biodiversity: Plants and Animals." from http://www.pierroton.inra.fr/genetics/labo/cpdna.html.

Rozen S, Skaletsky H (2000) *Primer3 on the WWW for general users and for biologist programmers.* In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386.

Schaal, B. A., Hayworth, A., Olsen, K. M., Raushcher, J. T., Smith, W. A. (1998). "Phylogenetic studies in plants: problems and prospects." Mol. Ecol. 7: 465-474.

Schaal, B. A. and Olsen, K.M. (2000). "Gene genealogies and population variation in plants." Proc Natl Acad Sci U S A 97(13): 7024-7029.

Shamns, K. (2005). "Handbook of Vegetable Oils and Fats." from http://www.fediol.be/2/index.php

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C.,Winder, C. T., Schilling, E. E., Small, R. L. (2005). "The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analyses." Am. J. Bot. 92(1): 142-166.

Strand, A. E., Leebens-Mack, J. and Milligan, B.G (1997). "Nuclear DNA-based markers for plant evolutionary biology." Mol Ecol 6(2): 113-118.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). "Universal primers for amplification of three non-coding regions of chloroplast DNA." Plant Mol Biol 17(5): 1105-1109.

Tompkins, A. (2004). "Poynter Online Special Edition: Ricin." from http://www.poynter.org/column.asp?id=2&aid=60383.

Williams, D. (1995). "Castor Oil. Natural Protection from Deadly Viruses." from http://www.industryinet.com/~ruby/castoroil.html.