DNA Barcoding in a Crop Genebank: The *Capsicum annuum* Species Complex

Robert L. Jarret[*]

USDA/ARS/PGRCU, 1109 Experiment Street, Griffin, GA 30223

Abstract: Variability within eight cpDNA introns including *trnS-trnfM*, *trnL-trnT*, *trnH-psbA*, *trnF-trnL*, *trnD-trnT*, *trnC-rpoB*, *rps*16 and *matK*, and the nuclear *waxy* introns was examined in seven species of *Capsicum* (*C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, *C. pubescens*, *C. chacoense* and *C. rhomboideum*) in order to evaluate the feasibility of utilizing these loci for DNA barcoding within the *C. annuum* complex. Numerous insertions/deletions (indels) and substitutions were detected in all cpDNA introns. However, none was sufficient to differentiate the individual members of the *C. annuum* complex (*C. annuum*, *C. chinense* and *C. frutescens*). Variation within *trnL-trnT*, *trnF-trnL* and *trnH-psbA* enabled the differentiation of the complex from the other taxa examined. In contrast, single base indels and substitutions within the *waxy* introns permitted the differentiation of all taxa within the plant materials examined. The use of *trnH-psbA* or *trnL-trnT*, and the *waxy* introns is proposed for barcoding members of the *C. annuum* complex.

INTRODUCTION

Crop genebanks provide research materials (plant germplasm) that are essential for the study of crop domestication/evolution, as well as for crop improvement, archeological, ecological, botanical, etc. research worldwide [1]. Proper identification of plant materials in genebanks is an essential activity in those units. The downstream consequences of the distribution and/or use of misidentified plant materials in a research program, and the subsequent adverse effects on the quality of the literature that would result from its use, can be significant. Many crop genebanks contain large numbers (hundreds or thousands) of accessions within specific genera or of individual species [2]. While most of these are readily classified, accurate identification of even extensively studied crop species and their relatives can occasionally be problematic. Such is the case with the common garden pepper *Capsicum annuum* L. and its domesticated relatives *C. chinense* Jacq. and *C. frutescens* L.

Capsicum annuum is one of five cultivated species in the genus, the others being *C. baccatum* L., *C. chinense*, *C. frutescens* and *C. pubescens* Ruiz and Pav. While *C. pubescens* and *C. baccatum* are quite distinct morphologically, *C. annuum*, *C. chinense* and *C. frutescens* show evidence of parallel evolution for a variety of plant and fruit morphological characteristics as a result of similar regimens of human selection [3]. The near continuous overlap in morphological characteristics among these three species led various authors to recognize them as a species complex [4]. Within the complex, taxa are differentiated from one another based primarily on differences in corolla color, the presence or absence of a calyx constriction and the occurrence of multiple pedicels/node [4, 5]. However, despite the development of numerous keys [4-7] that include these and other traits, the scoring of some characteristics

remains ambiguous as their occurrence/expression is variable within species, and misidentifications can occur. Ease and accuracy in the identification of plant materials are functions of the knowledge and experience of the individual making those identifications. However, the failure to develop a single identification key that unambiguously identifies all the members of this species complex suggests that perhaps an alternative or complementary approach to identification might enhance the accuracy of the process.

DNA barcoding is a natural extension of the use of various DNA markers for purposes of 'fingerprinting' and molecular systematics, and the development of publicly accessible and searchable DNA sequences databases. The concept of 'barcoding,' that was initially based on the exploitation of variability within the mitochondrial *Cytochrome c oxidase* (*cox* I) gene, appears to be quite effective for identifying animals, insects, and other life forms [8-12]. However, this locus has been found to be less variable in plants [13] and subsequent studies have focused on the evaluation of alternative loci for use specifically with Angiosperms [13-16]. Many of the loci evaluated to date as potential substitutes for *cox* I have been cpDNA introns and these have been utilized with some success. Shaw *et al.* [17] examined a wide range of cpDNA introns for use in low-level systematic studies and identified those introns that provided a greater than average number of potentially informative characters (PICs) of potential use in defining systematic relationships. Several of these have been suggested as suitable for use as DNA barcodes [8-10, 13, 14, 16]. However, the use of DNA barcoding for species identification is not universally accepted and the strengths and weaknesses of it have been discussed by others [8, 10, 18-20].

The present report describes the evaluation of variability within entire or portions of eight cpDNA introns and the nuclear *waxy* introns as candidates for barcoding members of the *C. annum* complex.

*Address correspondence to this author at the USDA/ARS/PGRCU, 1109 Experiment Street, Griffin, GA 30223; Tel: 770 228 7303; Fax: 770 228 7282; E-mail: bob.jarret@ars.usda.gov

MATERIALS AND METHODOLOGY

Plant Materials

All plant materials utilized in this study were obtained from the S-9 genebank in Griffin, GA [2]. Seeds were treated with 10% trisodium phosphate, rinsed and sown in peat pots containing a standard peat-perlite potting mix in a greenhouse in March 2007. Plants were transferred to the field in Byron, GA in May 2007.

A panel consisting of 15 accessions (hereafter referred to as Group I) was selected for initial studies to evaluate variability within eight cpDNA introns and the nuclear *waxy* introns for purposes of differentiating taxa. Plant materials examined and their genebank inventory numbers [Plant Introduction (PI) or GRIF] included representative examples of the five cultivated *Capsicum* species specifically; *C. annuum* var. *annuum* (PI592837 and PI586661), *C. frutescens* (PI360724, PI439512, and PI585257), *C. chinense* (PI315008 and PI438638), *C. baccatum* var. *pendulum* (Willd.) Eshbaugh (PI260564), *C. baccatum* var. *baccatum* L. (PI439360) and *C. pubescens* (GRIF1613 and PI593635). In addition, Group I contained *C. chacoense* Hunz. (PI260435 and PI260436) a non-cultivated species previously shown to be closely related to but distinct from other cultivated species of *Capsicum* [4], and *C. rhomboideum* (Dunal) Kuntze (GRIF16136) a distantly related taxa intermittently assigned to *Witheringia* but currently classified as *Capsicum* [21]. Digital images of plants, flowers, and fruits of all genotypes as well as information on their origin and data on their morphological characteristics can be viewed at www.ars-grin.gov.

Additional plant materials (Group II) were examined in order to further evaluate variability both within and among members of the *C. annuum* complex and to validate the efficacy of markers identified in Group I in differentiating species. Group II contained 19 accessions of *C. annuum* var. *annuum* (GRIF9161 and PI nos. 215743, 267730, 281387, 281389, 329246, 410407, 438565, 438663, 441628, 547069, 585238, 586661, 586668, 586678, 592815, 592816 and 593562, 593587), 14 accessions of *Capsicum annuum* var. *glabriusculum* (Dunal) Heiser & Pickersgill (PI nos. 224405, 381626, 406948, 439325, 511885, 511886, 511887, 593491, 593574, 593644, 631135, 631136, 631141 and 632932), 10 accessions of *C. frutescens* (GRIF9301, GRIF9312 and PI nos. 195299, 241676, 368084, 439515, 441653, 585254, 586675 and 639661) and 10 accessions of *C. chinense* (GRIF9271, GRIF9324, GRIF9328 and PI nos. 224443, 238051, 257136, 257156, 438643, 439433 and 594139). These materials were selected at random from the genebank inventory.

DNA Isolation, PCR, and DNA Sequencing

Immature leaf tissue was harvested from individual plants in the field or greenhouse and DNA extracted using the DNeasy Plant Mini-kit (Qiagen). Since only a single accession of *C. rhomboideum* was available, DNA was isolated from two plants selected at random from within a population of plants of this species that was established in the greenhouse. DNA extracts were diluted in Tris-HCl/EDTA (TE) and stored at -20°C until utilized.

PCR Amplification and Sequencing

A total of eight cpDNA introns and one nuclear DNA intron were examined. These introns were selected based on their ability to provide potentially informative characters (PICs) [17], ease of amplification, or previously documented use as DNA barcodes [8, 13, 14, 16, 22]. Primers for amplification of introns *rps16, rpoB-trnC, trnH-psbA, trnD-trnT, trnS-trnfM, trnL-trnT, trnF-trnL* were as described by Shaw *et al.* [17]. The *matK* intron was amplified using primers 390F and 1326R [17]. The nuclear *waxy* introns were amplified using the primers reported by Walsh and Hoot [23]. Primer sequences and annealing temperatures are noted in Table **1**.

PCR reactions contained: 1X PCR Buffer II (Gibco/BRL), 3.0mM MgCl$_2$, 15 - 30 ng of template DNA, 25 pmol of each primer, and one unit of Taq polymerase (Gibco/BRL) in 50ul reactions. Samples were subjected to the PCR as follows; initial denaturation for 3 min at 95°C, 35 cycles of; 94°C - 2 min, annealing temperature - 1 min, extension at 72°C - 2 min, followed by a final extension at 72°C for 10 min. PCR products were separated and visualized on 1.0 % agarose gels (TAE) containing 0.1% ethidium bromide, and cleaned using the QIAquick PCR Purification Kit (Qiagen). PCR products were cleaned using Qiagen's QIAquick PCR Product Purification Kit (Valencia, CA) and stored at -20 $^{\circ}$C for up to 1 week prior to sequencing.

DNA Sequencing and Data Analysis

PCR products were sequenced by Sequetech (Mountain View, CA) or the University of Georgia's Integrated Biotech Laboratories (IBL) Sequencing and Synthesis Facility (Athens, GA). Contigs were assembled using Sequencher version 4.8. All sequence data were visually checked against the output chromatograms, trimmed and the sequences edited, as required. In the event of an ambiguity, the PCR products were re-amplified and re-sequenced

Kress and Erickson [14] reported that (with *trnH-psbA*) for purposes of barcode identification, no significant difference was observed between the uses of partial sequences, versus complete sequences. Hence, in order to minimize the bias associated with the variable sequence lengths [17], and reduce the potential costs associated with the use of this technology for the processing of large numbers of samples, sequences were trimmed to 650bp, as high quality reads of this length were readily accomplished in a single sequencing reaction. The shorter *trnH-psbA* was analyzed in its entirety. Sequences were aligned using ClustalW and analyzed visually and using the Unweighted Paired-Group Method with Arithmetic Averages (UPGMA) *via* MEGA 4.0. Distances are expressed as percent disimilarity. Complete data sets are available from the author. Sequence data have been submitted to GenBank.

RESULTS

All intron sequences amplified readily and reproducibly with slight modifications to previously published annealing temperatures.

Table 1.   **PCR Primers and Annealing Temperatures Utilized for Amplification of cpDNA Introns and the Nuclear *Waxy* Introns from *Capsicum* spp.**

| Locus	Forward Primer/Reverse Primer[1]	Annealing Temperature
trnS-trnfM	GAG AGA GAG GGA TTC GAA CC/ CAT AAC CTT GAG GTC ACG GG	50°C
trnL-trnT	CAA ATG CGA TGC TCT AAC CT/ TCT ACC GAT TTC GCC ATA TC	54°C
trnF-trnL	CAA AAT CGTG TAG ACG CTA CG/ ATT TGA GGT GAC ACT GAG	52°C
trnD-trnT	ACC AAT TGA ACT ACA ATC CC/ CTA CCA CTG AGT TAA AAG GG	58°C
trnC-rpoB	CAC CCR GAT TYG AAC TGG GG/ CKA CAA AAY CCY YCR AAT TG	55°C
*rps*16	AAA CGA TGT GGT ARA AAG CAA C/ AAC ATC WAT TGC AAS GAT TCG ATA	55°C
matK	CGA TCT ATT CAT TCA ATA TTT C/ TCT AGC ACA CGA AAG TCG AAG T	55°C
trnH-psbA	CGC GCA TGG TGG ATT CAC AAT CC/ GTT ATG CAT GAA CGT AAT GCT C	50°C
waxy	TGC CTG GGA TAC AAG CAT TA/ AAC TGG AGG CAA TGT GAA AT	56°C
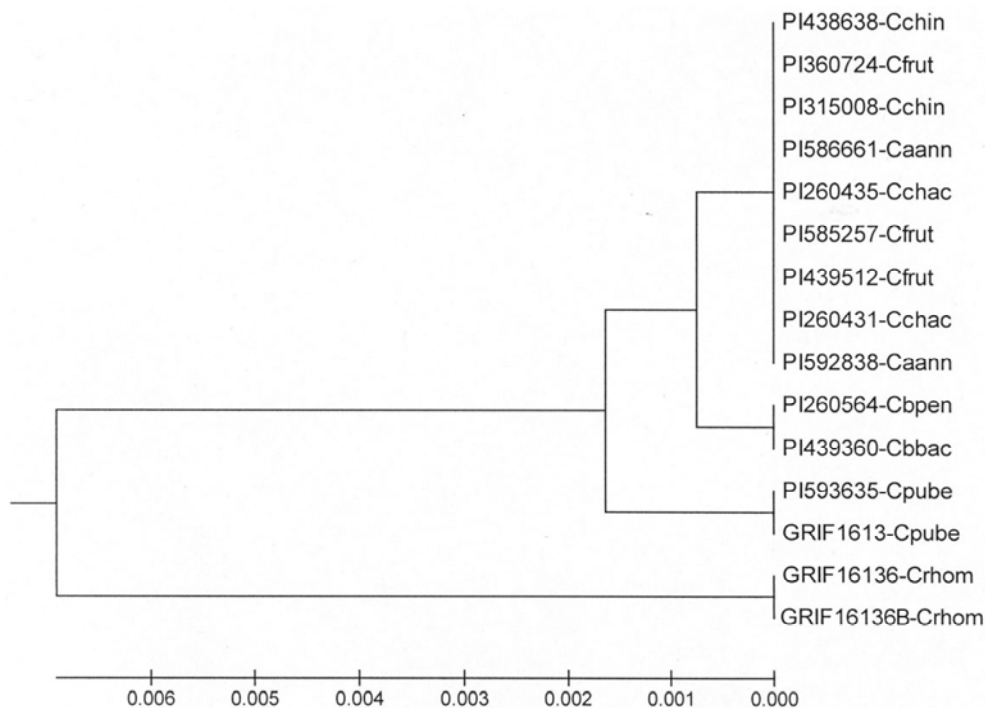
Fig. (1). UPGMA analysis of *trnC-rpoB* intron sequence data from seven *Capsicum* spp.

cpDNA Introns (with the Exception of *trnH-psbA*)

The cpDNA introns contained variable numbers of single bp substitutions and varying numbers of insertions/deletions (indels) that were variable in length (Table **2**). No intraspecific variability was observed for the occurrence of substitutions or the occurrence or length of indels, in Group I. Many single bp substitutions (~72%) and indels (~68%) were unique to *C. rhomboideum*. *Capsicum rhomboideum* contained the longest (71 bp) indel (*trnF-trnL*) in addition to the largest number of indels observed among all taxa and introns.

A UPGMA analysis of the individual intron-sequence data sets facilitated the visualization of the extent to which variability within each intron resolved individual taxa (Fig. **1**). Variation in bp substitutions within only *trnL-trnT* and *trnF-trnL* separated the *C. annuum* complex from the other taxa, and the other taxa from each other. Other data sets produced dissimilar results. The *trnD-trnT* data grouped C. *chacoense* with the *C. annuum* complex members, and *C. baccatum* and *C. pubescens* with each other. The *trnC-rpoB* data also grouped C. *chacoense* with the *C. annuum* complex members, but separated the other taxa from each other. Analysis of *rpS16* sequences separated the *C. annuum* complex members from the other taxa, but grouped *C. pubescens* with *C. baccatum*.

trnH-psbA Intron

This intron has been found to be highly variable among plant taxa and recommended as a candidate for a universal barcode in plants [13-15], and so was treated separately.

Group I

Alignment of the *trnH-psbA* intron sequences from the Group I accessions revealed the presence of a single poly-T microsatellite (T_{9-10}), 3 indels, and a total of 13 single base-pair substitutions (Table **2**). Within the plant material examined, the poly-T microsatellite was uniformly 10 bp in length, except for *C. pubescens* (9 bp). One single bp insertion was found only in *C. rhomboideum* and one single bp deletion was unique to *C. pubescens*. A 7 bp insertion was present only in *C. annuum*. A 13 bp insertion was unique to *C. chinense* and *C. frutescens*. Analysis of this data set, following the removal of the microsatellite and all indels, resulted in a cladogram with the delineation of species presented Fig. (**2**). Members of the *C. annuum* complex were unresolved, but members of the complex were separated (as a group) from the other taxa.

Group II

Analysis of the additional 53 accessions of the *C. annuum* complex members alone revealed the absence of any additional substitutions within the sequence data set, with one exception. The poly-T microsatellite was uniformly 10 bp in length – except for *C. annuum* var. *annuum* accessions GRIF 9161 PI267730 (T_{11}), *C. annuum* var. *annuum* accessions PI 281387 and PI 281389 (T_{12}), and *C. annuum* var. *glabriusculum* accessions PI 511886 and PI 511887 (T_{11}).

Accessions of *C. annuum* in Group I were differentiated from *C. chinense* and *C. frutescens* based on the presence or absence of a 7 or a 13 bp indel. However, this was not the case when the dataset was extended to include the Group II plant materials. The 7 bp insertion noted previously was absent in all *C. chinense* and *C. frutescens* accessions, as observed in Group I. However, it was also absent in 15 of 35 *C. annuum* (5 out of 19 *C. a.* var. *annuum* and 10 out of 16 *C. a.* var. *glabrisusculum*). In a similar fashion, the 13 bp insertion that was unique to *C. chinense* and *C. frutescens* in Group I materials was present in all Group II accessions of these spp. However, it was also present in 7 accessions of *C. a.* var. *glabriusculum* (PI nos. 406948, 439325, 593574, 593644, 631136, 631141 and 632932), and two accessions of *C. a.* var. *annuum* (GRIF 9161 and PI 267730). Four accessions of *Capsicum annuum* (PI 281387, 281389, 511886 and 511887) from Mexico lacked both insertions.

The Nuclear *Waxy* Introns

The *waxy* introns were examined by Walsh and Hoot [23] in their study of systematic relationships of cultivated and wild *Capsicum* spp. In that study, a series of indels and unique substitutions were identified - although these were not utilized to examine intraspecific diversity or evaluated for use as DNA barcodes. We opted to test the robustness of
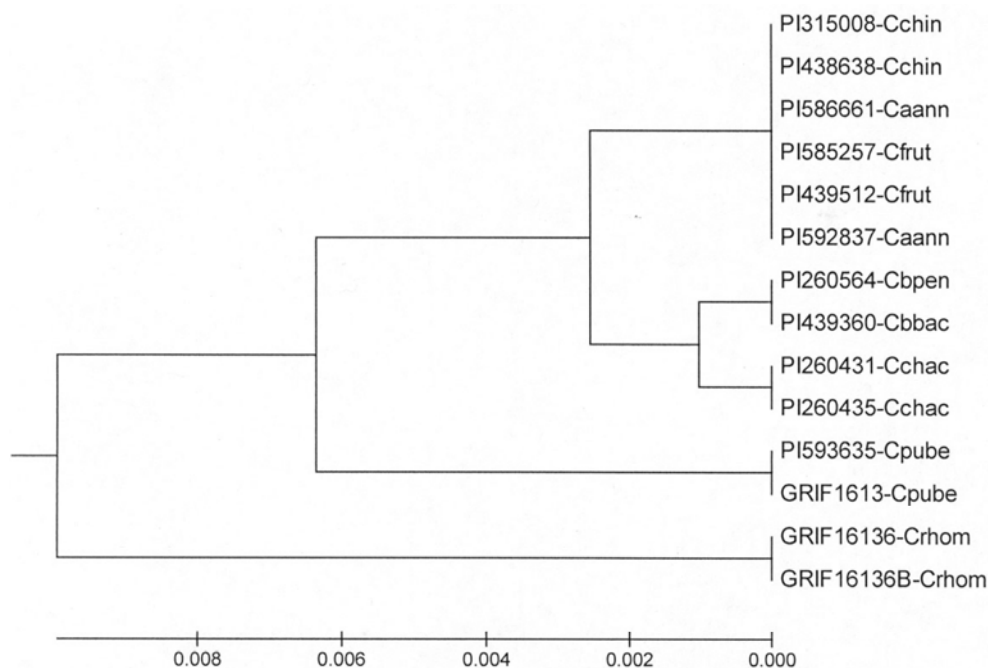

Fig. (2). UPGMA analysis of *trnH-psbA* intron sequence data from seven *Capsicum* spp.

Table 2.   Unique Substitutions and Indels among 15 Accessions (6 Species) of *Capsicum* Associated with cpDNA Introns and the Nuclear *Waxy* Intron

Locus	Substitutions	Indels[1]
trnS-trnfM	5	1 / 7 bp
trnL-trnT	5	1 / 5 bp 1 / 6 bp 1 / 12 bp
trnH-psbA	12	1 / 1 bp 1 / 7 bp 1 / 13 bp
trnF-trnL	14	1 / 6 bp 1 / 19 bp 1 / 71 bp
trnD-trnT	6	2 / 5 bp 1 / 1 -1, -15 or -25 bp
trnC-rpoB	11	1 / 7 bp 1 / 19bp
*rps*16	9	1 / 1 bp 1 / 5 bp
matK	8	0
waxy	73	8 / 6 - 1 bp, 1 - 1 or 2 bp, 1 - 12 bp

[1]Number of occurrences/length in bp.

those markers for purposes of barcoding the members of the *C. annuum* complex.

Group I

Seventy-three substitutions and 8 indels were detected in the Group I data set (Table **2**). Among these, 54 substitutions and 4 indels were unique to *C. rhomboideum*. The largest observed indel at this locus (12 bp) was also unique to *C. rhomboideum*. Other apparently unique markers included; *C. chacoense* - 5, *C. baccatum* - 4, *C. baccatum* var. *pendulum* - 2, *C. pubescens* - 3, *C. annuum* - 2, and *C. chinense* and *C. frutescens* - 1 each. The results of a UMGMA analysis of these data are presented in Fig. (**3**). As indicated, the members of the *C. annuum* complex were separated from each other, and from the other taxa analyzed.

Group II

A total of 7 markers were identified in the Group II data set. Markers unique to specific taxa within the group included; *C. annuum* – 2, *C. annuum* var. *annuum* – 1, *C. frutescens* - 1 and *C. chinense* - 1. Other markers were shared among two or more taxa. Of these unique markers, 2 were single bp indels, and the remainder were substitutions. The indels and substitutions that were unique to *C. annuum*, *C. chinense* and *C. frutescens* in the Group I materials were similarly unique to these species in the Group II materials.

DISCUSSION AND CONCLUSION

The plant materials contained within crop genebanks, and the uses for DNA barcoding to characterize the materials

within them are somewhat distinct from, for example, the application of this technology to a broad-based floristic or total biodiversity survey. Crop genebanks contain a limited number of plant families each containing one or more genera and are typically assembled around a single crop species (i.e. *C. annuum*). Other cultivated species of lesser importance (i.e. *C. frutescens*, *C. pubescens*, etc.) and wild crop-related species (*C. chacoense*, *C. rhomboideum*) and genera are included with the assumption that all of these have obvious or intrinsic value. Hence, the number of taxa to be subjected to barcoding in a crop genebank, or in a specific crop germplasm collection, is relatively restricted and most materials have already been confidently identified to the level of genus.

An inherent goal of most genebanks is to maximize the genetic diversity present within collections. This results in the intentional seeking out of morphological, physiological, ecological and genetic variants. Variants might be unique for characteristics such as plant structure, flower color, various adaptive traits, unique fruit characteristics, adaptation to geographical extremes, ploidy level, etc. One or more of the traits exhibited by these variants can confound classification efforts through the introduction of a characteristic that does not exist within established keys, or that influences the expression of one that does. These large assemblages of plant materials are narrow from a taxonomic standpoint, and yet broad in terms of the extent of their intraspecific (and intrageneric) morphological diversity. As collections of individual species increase in size and complexity, so do opportunities for mis-identifications as materials near the
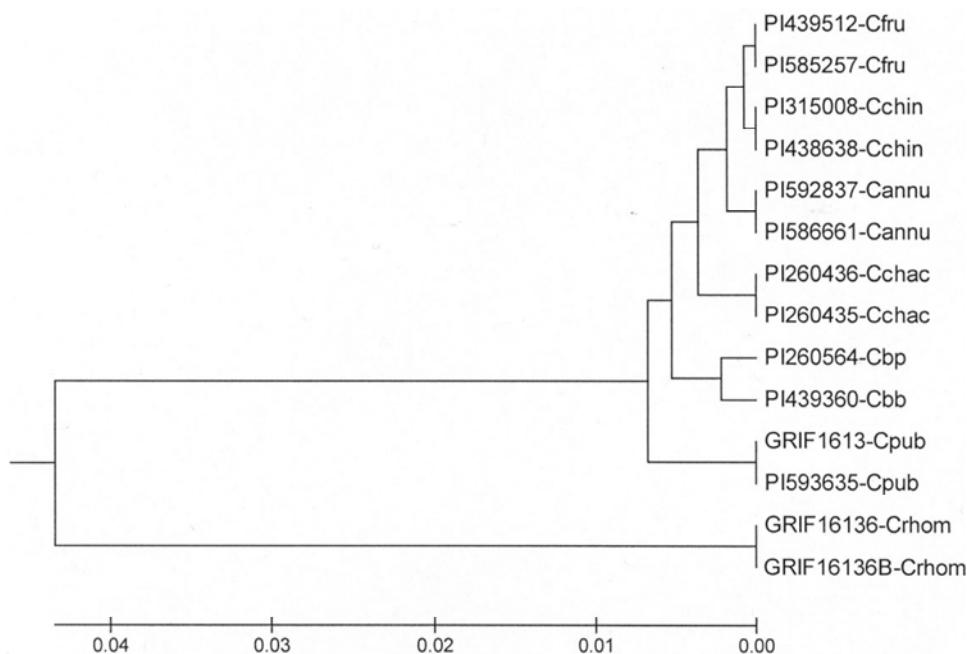
Fig. (3). UPGMA analysis of the *waxy* locus intron sequences from seven *Capsicum* spp.

species boundaries or unique morphotypes within existing taxa, become confused with their sister taxa or appear to be unique taxa, respectively. The possibility also exists that new taxa will go unrecognized.

Though not to be considered a replacement for traditional taxonomic approaches, the use of DNA barcodes for species identification could, in some instances, increase the efficiency of genebank operations. For example, seed samples of plant materials whose classification has not been determined (or confirmed) are sometimes received. In these instances, the sacrifice of a single seed/seedling for DNA extraction and barcoding is a desirable alternative to the utilization of expensive greenhouse space to grow out one or more plants to maturity in order to determine, or assess the validity of, its classification. For those materials already in germplasm collections whose classification is unknown or uncertain, a rapid means of species identification – even if only tentative - could aid in the planning of the regeneration process by providing an advance notice of specialized resources or environmental conditions that might be required for their culture (i.e. specific temperature ranges or supplemental lighting for photoperiod sensitive species).

According to Lahaye *et al.* [15] "A suitable barcode must exhibit high interspecific but low intraspecific divergence." An effective DNA barcode in a genebank setting is one that is accurate to the species level - within the plant materials of interest. That is, in order to be effective, it need not be universally adaptable to all plant families, all genebanks or all collections, but should to the extent possible provide an unambiguous identification. Given the importance of accuracy, the use of barcodes that performed well only within a specific family (or genus) could readily be justified if appropriate (more generalist) alternatives were not available, and if its use facilitated improvements in the identification of recalcitrant/novel phenotypes. Clearly, the use of a universal or near-universal barcode such as *trnH-psbA* provides additional benefits by facilitating direct

comparisons across diverse arrays of taxa and the development and adoption of near uniform assay conditions.

Other types of DNA markers have proven useful in clustering *Capsicum* spp. [24-27]. Generally, these detected higher levels of both interspecific and intraspecific diversity and were intended for use in DNA fingerprinting or linkage mapping studies. The cpDNA and *waxy* intron loci examined in the present study are more conserved than hypervariable loci such as microsatellites [28]. In the present study, length variants within the *trnH-psbA* poly-T microsatellite were not correlated with taxonomy and was greater within *C. annuum* (T_{10} to T_{12}) than among other taxa - that were uniformly T_{10}. In contrast, specific indels within *trnD-trnT* and other introns, were unique (within the materials examined) to one or more taxa including; *C. pubescens* (5 bp insertion), *C. chinense* and *C. frutescens* (5 bp deletion), *C. baccatum* (25 bp deletion), and *C. annuum* and *C. chacoense* (15 bp deletion). However, before any conclusions concerning the potential of these and other indels for species identification or for inferring phylogeny could be drawn, a much broader and larger array of plant materials would need to be examined.

Chase *et al.* [13] and Kress and Erickson [14] proposed the use of *trnH-psbA* for barcoding land plants. In the study of Lahaye *et al.* [15], *trnH-psbA* provided the highest interspecific divergence values of the barcodes tested. The next most variable barcode at the interspecific level was *matK*. Acknowledging that the evolution of *trnH-psbA* is complex [15], the preferential use of *matK* alone, or in combination with *trnH-psbA* - as opposed to the combination of *trnH-psbA* and *rbcL* proposed by Kress and Erickson [14], was recommended for the barcoding of Angiosperms. The combination of *matK* and *trnH-psbA* resulted in greater species identification - as evidenced by an improvement in recovering species monophyly [15]. However, we found *trnH-psbA* to be more effective than *matK* in differentiating the *C. annuum* complex from the other taxa examined,

although neither was satisfactory for purposes of barcoding individual species within the complex. The multi-base indels detected within *trnH-psbA* were excluded from the data analyses. However, one or more of these may yet prove useful in future examinations of the phylogeny of the *C. annuum* complex, in studying diversity within this genus at other levels [29], or for the characterization of the new species of *Capsicum* that have been recently described [30, 31].

It seems intuitive that markers that provide well resolved and strongly supported phylogenies, would be candidates for use as barcodes – even if their use might be limited to a restricted range of taxa. In their study of phylogenetic relationships of *Capsicum* using the *waxy* introns and the *atpB-rbcL* spacer region, Walsh and Hoot [23] identified a series of indels and substitutions within the *waxy* introns, some of which differentiated the members of the *C. a.* complex. The present study confirms the finding of Walsh and Hoot [23] and extends those findings to a larger sampling of plant materials. Within these introns, a single bp indel and a substitution separated *C. chinense* from *C. frutescens*. Several investigators have suggested that as a result of their morphological similarities, that *C. frutescens* and *C. chinense* be combined [3, 32]. Yet, plant materials of each of these species could be identified, based on the presence or absence of these anomalies. The *waxy* locus has not previously been utilized for purposes of barcoding, but has proven particularly useful for establishing phylogenetic relationships within the Solanaceae [33-35]. We suggest that the utilization of the *waxy* introns, in addition to *trnH-psbA* (or other cpDNA introns) for barcoding *Capsicum* spp., provides the benefits associated with the use of both uni-parentally and bi-parentally inherited markers [13] and shows potential as a means to quickly and (perhaps) inexpensively characterize these plant materials - until such time as superior markers are identified or key developed.

Sequence variation within four of the eight cpDNA introns examined permitted the separation of the *C. annuum* complex members from other taxa. Our data confirm an earlier report on the presence of indels and substitutions within the *waxy* introns whose presence or absence was linked with a morphology-based identification. While the evolutionary significance (if any) of these indels and substitutions is not known, this linkage suggests that certain of these indels and substitutions provide an additional means of characterizing these plant materials. The assay of large numbers of samples for the presence or absence of these and other single nucleotide polymorphisms lends itself to rapid analysis [36] and thus lends itself to the (typically) large numbers of accessions represented in crop genebanks. Analysis of *Capsicum waxy* and cpDNA introns may also prove useful in examining archeological specimens whose identifications are based on other alternative approaches [37].

REFERENCES

[1] Vaughan DA, Balazs E, Heslop-Harrison JS. From crop domestication to super-domestication. Ann Bot 2007; 100: 893-901.

[2] Jarret RL, Spinks M, Lovell G, Gillaspie AG. The S-9 plant germplasm collections at Griffin, GA. Diversity 1990; 6: 23-5.

[3] Pickersgill B. Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). Evolution 1971; 25: 683-91.

[4] McLeod MJ, Guttman SI, Eshbaugh WH. Early evolution of chile peppers (*Capsicum*). Econ Bot 1982; 36: 361-8.

[5] Smith PG, Heiser CB Jr. Taxonomy of *Capsicum sinense* Jacq. and the geographic distribution of the cultivated *Capsicum* species. Bull Torrey Bot Club 1957; 84: 413-20.

[6] D'Arcy WG, Eshbaugh WH. New World peppers (*Capsicum* – Solanaceae) North of Colombia: A resume. Baileya 1974; 19: 93-105.

[7] Pickersgill B, Heiser CB Jr, McNeill J. In: Hawkes JG, Lester RN, Skelding AD, Eds. The Biology and Taxonomy of the Solanaceae. NY, Academic Press 1979; 679-700.

[8] Hebert PDN, Cywinska A, Ball SL, de Waard JR. Biological identification through DNA barcodes. Proc R. Soc Lond, B Biol Sci 2003; 270: 313-21.

[9] Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. PLoS Biol 2004; 2: e312.

[10] Taberlet P, Coissac E, Pompanon F, *et al.* Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Res 2007; 35: e14;doi: 10. 1093/ nar/gk1938.

[11] Seifert KA, Samson RA, deWaard JR, *et al.* Prospects for fungus identification using *CO1* DNA barcodes, with *Penicillium* as a test case. Proc Natl Acad Sci USA 2007; 104: 3901-6.

[12] Robba L, Russell SJ, Baker GL, Broody J. Assessing the use of the mitochondrial *cox 1* marker for use in DNA barcoding of red algae. Am J Bot 2006; 93: 1101-8.

[13] Chase M, Salamin N, Wilkinson M, *et al.* Land plants and DNA barcodes: short-term and long-term goals. Phil Trans R Soc B 2005;doi.10.1098/rstb.2005.1720.

[14] Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: The coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. PLoS One 2007; 2: e508.doi:10.1371/journal.pone.0000508.

[15] Lahaye R, van der Bank M, Bogarin D, *et al.* DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA 2008; 105: 2923-8.

[16] Kress WJ, Wurdack KJ, Zimmer EA, Weight LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 2005; 102: 8369-74.

[17] Shaw J, Lickey EB, Beck JT, *et al.* The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 2005; 92: 142-66.

[18] Will K, Rubinoff D. Myth of the molecule: DNA barcodes for some species cannot replace morphology for identification and classification. Cladistics 2004; 20: 47-55.

[19] Rubinoff D, Cameron S, Will K. A genomic perspective on the shortcomings of mitochondrial DNA for 'barcoding' identification. J Hered 2006; 97: 581-94.

[20] Rubinoff D, Cameron S, Will K. Are plant DNA barcodes a search for the Holy Grail? Trends Ecol Evol 2006; 21: 1-2.

[21] Hunziker AT, Genera Solanacearum. Ruggell, A.R.G. Gantner Vela K.-G. 2001.

[22] Cuenoud PV, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K DNA sequences. Am J Bot 2002; 89: 132-44.

[23] Walsh BM, Hoot SB. Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: The chloroplast *atpB-rbcl* spacer region and nuclear *waxy* introns. Int J Plant Sci 2001; 162: 1409-18.

[24] Prince JP, Lackney VK, Angeles C, Blauth JR, Kyle MM. A survey of DNA polymorphisms within the genus *Capsicum* and the fingerprinting of pepper cultivars. Genome 1995; 38: 224-31.

[25] Rodriguez JM, Berke T, Engle L, Nienhuis J. Variation among and within *Capsicum* spp. revealed by RAPD markers. TAG 1999; 99: 147-56.

[26] Toquica SP, Rodriguez F, Martinez E, Duque MC, Tohme J. Molecular characterization by AFLPs of *Capsicum* germplasm from the Amazon Department in Colombia, characterization by AFLPs of *Capsicum*. Genet Res Crop Evol 2003; 50: 639-47.

[27] Portis E, Nagy I, Sasvan Z, Stagel A, Barchi L, Lanteri S. The design of *Capsicum* spp. SSR assays *via* analysis of *in silico* DNA sequence, and their potential utility for genetic mapping. Plant Sci 2006; 172: 640-8.

[28] Tautz D. Hypervariability of simple sequence repeats as a general source for polymorphic DNA markers. Nucleic Acids Res 1989; 17: 6444-71.

[29] Kawasaki M, Ohnishi O. Two distinct natural populations of *Fagopyrum urophyllum* (Bur. et Franch.) Gross revealed by the nucleotide sequence of a noncoding region in chloroplast DNA. Genes Genet Syst 2006; 81: 323-32.

[30] Barboza GE, L de Bianchetti. Three new species of *Capsicum* (Solanaceae) and a key to the wild species from Brazil. Syst Bot 2005; 30: 863-71.

[31] Nee M, Boyd, Knapp S. New species of *Solanum* and *Capsicum* (Solanaceae) from Bolivia, with clarification of nomenclature in some Bolivian *Solanum*. Brittonia 2006; 58: 322-56.

[32] Heiser CB Jr, Pickersgill B. Names for the cultivated *Capsicum* species (Solanaceae). Taxon 1969; 18: 277-83.

[33] Levin RA, Myers NR, Bohs L. Phylogenetic relationships among the "spiny solanums' (*Solanum* subgenus *Leptostemonum*, Solanaceae). Am J Bot 2006; 93: 157-69.

[34] Levin RA, Watson K, Bohs L. A four-gene study of evolutionary relationships in *Solanum* section *Acanthophora*. Am J Bot 2005; 92: 603-12.

[35] Weese TL, Bohs L. A three-gene phylogeny of the genus *Solanum* (Solanaceae). Syst Bot 2007; 32: 445-63.

[36] Montgomery J, Witttwer CT, Palais R, Zhou L. Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. Nat Protoc 2007; 2: 59-66.

[37] Perry L, Dickau R, Zarrillo S, *et al.* Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. Science 2007; 315: 986-8.