PHYLOGENY AND BIOGEOGRAPHY OF WATERMELON [*CITRULLUS LANATUS* (THUNB.) MATSUM. & NAKAI] BASED ON CHLOROPLAST, NUCLEAR SEQUENCE AND AFLP MOLECULAR MARKER DATA

Jiarong Liu

A thesis
submitted to
the Graduate Faculty of
Auburn University in Partial Fulfillment of the
Requirements for the
Degree of
Master in Science

Auburn, Alabama
August 8, 2005

PHYLOGENY AND BIOGEOGRAPHY OF WATERMELON [*CITRULLUS LANATUS* (THUNB.) MATSUM. & NAKAI] BASED ON CHLOROPLAST, NUCLEAR SEQUENCE AND AFLP MARKER DATA

Jiarong Liu

_____

Signature of Author

_____

Date of Graduation

VITA

Jiarong (Rona) Liu, daughter of Maosheng Liu and Hongmei Ye, was born on November 7, 1981 in Suzhou, Jiangsu Province, the People's Republic of China. She graduated in June 1999 from Suzhou No.3 High School. She attended Yangzhou University, Yangzhou, Jiangsu Province, P. R. China in 1999 and graduated in June, 2003 with a Bachelor of Science in Horticulture. In August 2003, she entered Graduate school at Auburn University, Auburn, Alabama to pursue a Master of science degree in Horticulture. She was employed as a Graduate Research Assistant and Teaching Assistant during her graduate studies.

THESIS ABSTRACT

PHYLOGENY AND BIOGEOGRAPHY OF WATERMELON [*CITRULLUS LANATUS* (THUNB.) MATSUM. & NAKAI] BASED ON CHLOROPLAST, NUCLEAR SEQUENCE AND AFLP MARKER DATA

Jiarong Liu

Master of Science, July 12, 2005
(B.S., Yangzhou University, 2003)

75 Typed pages

Directed by Fenny Dane

Watermelons [*Citrullus lanatus* (Thunb.) Matsum. & Nakai], together with cucumbers, melons of various sorts, summer squashes, winter squashes and pumpkins are the principal food plants of the gourd family (Cucurbitaceae). The phylogeny of *C. lanatus* was estimated from separate and combined analysis of noncoding regions of chloroplast (*trn*S-*trn*G and *trn*R-*atp*A) and nuclear *G3pdh* sequence data and amplified fragment length polymorphisms (AFLP) marker data. Sequences from 18 taxa included in the study provided an aligned length of 869 bp for *trn*S-*trn*G and 610 bp for *trn*R-*atp*A (with total length of the chloroplast sequences of 1479 bp and the outgroup *C. colocynthis* (34256) sequence of 1442 bp). The sequenced *G3pdh* region covers one intron and a short section of the transit

peptide region with the length of around 900 bp. Combined sequence analysis of cpDNA and *G3pdh* divides the *C. lanatus* accessions into 11different haplotypes. The cultivated watermelon, *C. lanatus* var. *lanatus* accessions, grouped into one major clade, the citron type, var. *citroides*, into another clade. Two distinct lineages within *C. lanatus* var. *citroides* were detected. Since accessions from southern Africa contain ancestral haplotypes and the highest frequency (44.4%) of different var. *citroides* haplotypes, it can be considered the area of origin for *C. lanatus* var. *citroides*, with colonization patterns to Zaire, USA, Europe, and India.

Combined analysis of cpDNA and G3pdh sequences showed the accumulation of unique nucleotide substitutions in *C. lanatus* var. *lanatus*, suggesting that the two varieties (*lanatus* and *citroides*) diverged from a common ancestor. AFLP marker data also indicate low levels of genetic diversity, possibly as a result of the domestication of watermelon.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Fenny Dane for her greate guidance and direction, constructive criticism and wonderful support throughout her Master's studies. She also would like to thank Dr. John Liu and Dr. Narendra K. Singh for their support and help serving as her committee members. Thanks are extended to all other graduate students, faculty members and staff in the Horticulture Department of Auburn University.

Thanks are also due to my lovely family members, Mom and Dad especially for their support and encouragement through during the courses of this investigation.

Style manual of journal used American Journal of Botany

Computer software used Microsoft word 2000

Vector NTI; PAUP; NTSYS 2.1; SAS 8.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

II

III

I. Literature Review

1.1 Domestication and the origins of plants

Agriculture is one of the major technological innovations of humankind. The initiation of food production by humans is based on the domestication of a relatively small number of local grain plants. The transition to agriculture not only had revolutionary ecological and economic consequences; it was also associated with the development of settled life and it led ultimately (in some parts of the world) to the emergence of urban civilization. Most importantly, since the large majority of hunter-gatherers switched to agriculture as their staple food source, crops and farm animals were domesticated. Domestication is the process of genetic selection that, by altering key traits, transforms wild forms into domesticated varieties of plants (Salamini et al. 2002). Domestication is a very complicated procedure which includes several components: A physical barrier of some sort which separates a species into distinct reproductive groups within its geographical range was first set up by human societies. Over successive generations the groups on both sides of the barrier begin to diverge as they respond to selection pressures determined by human societies. A new set of selection pressure comes into play when humans intervene in key aspects of the life cycle of the now “captive” population, creating new rules for survival and reproductive success. Only those individuals able to survive and produce offspring

under the new rules contribute genetic information to the next generation. Over generations, in response to the new rules for survival, the captive populations change in a number of ways, some deliberately caused by the domesticators, others incidental and automatic. All of the adaptations or adjustments made by a captive population can be described as that species – adaptive syndrome of domestication. Many of the changes that occur as part of the adaptive syndrome of domeatication are “phenotypic”, or observable, and it is such observable changes that often enable us to determine that the species has been domesticated (Smith, 1998). Associated with these observable changes, are changes at the molecular level, in the genes themselves.

The process of domestication started more than 10,000 years ago independently at several different locations: (1) Mesoamerica (the southern half of Mexico and the northern half of Central America); (2) the Andes of South America, including its foothills region on the eastern slope; (3) the “Near East” (Southwest Asia); (4) the Sahel region and the Ethiopian highlands in Africa; (5) China; and (6) Southeast Asia (Harris, 1996). It is very likely that global climate changes had an impact on the sequence of events. The area of domestication are generally located in tropical or subtropical regions, at middle elevation (approximately 1,000-1,200 m) in areas with varied topographies, including river valleys, hills and mountainsides, and plateaus. These regions often have a climate with distinct wet versus dry season, with either Mediterranean climates, savanna, or monsoon climates. People domesticated different sets of plants in each of these different regions (See Table 1). However, in most regions crops with similar uses were domesticated. For example, people domesticated cereals in most of the regions, as well as crops of the legume family, since cereals and legumes complement each other nutritionally (Gepts, 1998). Even though most plants were domesticated in a single region, some were domesticated in more than one place. For

example cotton was domesticated in Mesoamerica and also in south America and Africa (Rindos, 1984). Rice became domesticated in Africa, China and also southeast Asia (Smith, 1998). These are so-called vicarious domestications, meaning domestications of similar plants in widely different locations, whether they belong to the same or different species. Additionally, domestication may have occurred in other areas (various parts of Africa), although convincing archeological evidence for this has not yet been found.

1.2 Systematics and domestication of watermelon

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai], together with cucumber, melons of various sorts, summer squash, winter squash and pumpkin are the principal food plants of the gourd family (Cucurbitaceae). The term cucurbit denotes all species within the Cucurbitaceae (Maynard, 2001). The Cucurbitaceae, which is not closely related to any other plant family, consists of two well-defined subfamilies, eight tribes representing varying degrees of circumscriptive cohesiveness, and about 118 genera and 825 species (Jeffrey, 1990). Two subfamilies-Zanonioideae and Cucurbitoideae-are well characterized: the former by small, striate pollen grains and the latter by having the styles united into a single column. Waternmelon is assigned to the subfamily Cucurbitoideae, the tribe Benincaseae, subtribe Benincasinae and is native to Africa (Maynard, 2001). In 1941, L. H. Bailey proposed dividing cultivated watermelon *C. vulgaris*, into botanical variety *lanatus* and botanical variety *citrodes*. The variety *citrodes* includes the citron or preserving melon, which produces fruit with hard, inedible bitter flesh, and green or tan seeds. The species can be classified based on the cucurbitacin or bitter principle content. One group of closely related species (*C. lanatus*, *C. colocynthis*, and *C. ecirrhosus*) had cucurbitacin E as the bitter substance, while the other group (*C. naudinianus*) has cucurbitacin B and E (and their derivatives)

(Bailey et al, 1941). Morphological and cytogenetic studies have revealed that the four species are cross compatible with each other. The maintenance of identity of the different species was attributed to geographical isolation, differences in flowering habit, genetic differences, and structural changes in chromosomes. The genus *Citrullus* has now been revised to include *C. lanatus* (syn. *C. vulgaris*), *C. ecirrhosus*, *C. colocynthis*, and *C. rehmii* (De Winter, 1990). *Citrullus ecirrhosus* is more closely related to C. *lanatus* that either is to *C. colocynthis* (Maynard, 2001). There are other two closely related species: *Praecitrullus fistulosus* from India and Pakistan, and *Acanthosicyos naudinianus* from southern Africa. Cytogenetic investigation support the separation of *Praecitrullus* from *Citrullus* (Jeffrey, 1990).

Watermelon has 22 chromosomes (2n=22, x=11). Other members of the Cucurbitaceae with 22 chromosomes include *Gymnopetalum, Lagenaria, Momordica, Trichosanthes*, and *Melothria*. None appear to be closely related to watermelon. The basic chromosome number of *P. fistulosus* is (2n=24, x = 12). Watermelon is a monoecious, warm-season crop. Flowering and fruit development are promoted by high light intensity and high temperature. Watermelon is the only economically important cucurbit with pinnatifid (lobed) leaves; all of the other species have whole (nonlobed) leaves. Watermelon growth habit is a trailing vine. The stems are thin, hairy, angular, grooved, and have branched tendrils at each node. Watermelon has small flowers that are less showy than those of other cucurbits. The fruit of watermelon are round to cylindrical, up to 61 cm long and have a rind 1.0-3.7 cm thick. The edible part of the fruit is the endocarp (placenta). Seeds continue to mature as the fruit ripens and the rind lightens in color. There is no dormancy in watermelon seeds, so they can be harvested on one day and planted the next. Seeds germinate in 2 days to 2 weeks depending on temperature and moisture conditions.

The primary gene center for watermelon is not known, although tropical Africa and India have been suggested because it is found growing wild through out those areas. One theory proposes that watermelon was derived from a perennial relative, *Citrullus colocynthis*, which is endemic to Africa and can hybridize with watermelon. Colocynth seeds have been found in early archaeological sites preceding the finding of watermelon remnants (Maynard, 2001). Interestingly, *C. colocynthis* is reported to also grow wild in India. Another theory is that watermelon was domesticated in Africa from putative wild forms of *Citrullus lanatus*. Wild populations of *C. lanatus* var. *citroides*, which are common in central Africa and known as citron, are suggested to have given rise to the domesticated var. *lanatus* (Maynard, 2001). However, some botanist regarded citrons as varieties of watermelon and not progenitors.

However, even though there is not sufficient evidence for where and when domestication of watermelon began, by 2000 B.C., watermelons were cultivated in the Nile Valley. Watermelons were widely grown in prehistoric times by agriculture peoples of sub-Saharan Africa. They had developed many landraces, varying in fruit size, shape, flesh color, and rind color. Also a spectrum of seed color mutants had been selected. From Africa, watermelons were introduced to India about 800 AD and China about 1100 AD. From India and China, cultivation spread to Southeast Asia in the 15th century, and reached Japan in 16th century. The Moors introduced watermelon to Europe during their conquest of Spain. Cultivation spread into other parts of Europe, although slowly due to a less favorable growing climate (Zohary et al, 1988; Dane et al, 2003). Nevertheless, watermelons were mentioned in European herbal writings in the 1500s, and by 1625 were widely planted as a minor crop in European gardens. The first record of watermelon in England dates to 1597. The watermelon was transported and introduced to the Americans in post-Columbus times by the early European colonists.

Spanish colonists were growing watermelons in Florida by 1567, and before 1600 in Panama, Colombia, and Peru. By the middle 1600s, watermelons were commonly grown in Latin America, Brazil, and in British and Dutch colonies of the New World. Watermelons were reported to be grown in the Massachusetts colony as early as 1629. Watermelon was readily accepted and disseminated by native Americans, especially in the Mississippi Valley and the southwestern United States (Sauer, 1993).

1.3 Methods to study plant domestication

Progress in understanding plant domestication depends on multidisciplinary investigations - to which archaeologist, biologists, geographers and other scientists have all contributed. The first definite signs of plant cultivation in the Old World appear in a string of early Neolithic farming villages that developed in the Near East by 7500-7000 BC (Zohary, 1988). The study of the archaeobotanical properties of plant remains from archaeological sites reveals when and where domesticated crops first existed (Salamini, 2002). Archaeology can tell us when agriculture arose and where to locate the domestication origin which contributes to the distribution patterns of domesticated crops. Phylogeographers seek to interpret the extent and mode by which historical processes in population demographics, including but not confined to those related to natural selection over extended periods of time, may have left evolutionary footprints on the contemporary geographic distributions of gene-based organismal traits (Avise, 2001). Phylogeographical approaches focus on the present-day wild species and populations, their relationships to cultivated crops, their distribution, their ecology and the trends in those morphological characters that are associated with plant domestication.

During the past two decades, the tools of molecular biology have been applied to systematics with remarkable success, especially after the discovery of polymerase

chain reaction (PCR) in the mid – 1980s. New insights have been gained into such topics as phylogenetic reconstruction, introgression, genomic evolution, and levels of genetic variation in natural populations. Phylogeny reconstruction can play an important role in creating a logical framework for understanding the basis for plant organismal evolution. Molecular methods have provided greater resolution than was previously possible with other approaches and have dramatically reshaped our views of organismal relationships and evolution. Moreover, variation in DNA sequences is more readily subjected to statistical analysis than many previous types of data, and it can be less ambiguous, making interpretation of data more straightforward.

1.3.1 Phylogenies, genetic distances and cytological methods

The fraction of alleles that differ between two individuals can be scored and used to determine genetic distances among closely related taxa. Various algorithms exist for this purpose, some of which infer distance on the basis of the presence or absence of characters, whereas others infer an estimate of the number of nucleotide substitutions that might have occurred between individuals in the restriction-site sequences (Felsenstein, 1993). From a matrix of pairwise genetic distances, a phylogenetic tree can be constructed in molecular phylogenetic analyses, which has provided unparalleled insight into relationships at all levels of plant phylogeny (Soltis, 2000). A phylogeny is a graph that depicts the relatedness of individuals, populations and species (Li, 2000).

A different approach involves constructing trees for populations on the basis of overall similarity of their allele frequencies; for example, at Amplified-fragment length polymorphisms (AFLPs) loci. In studies of plant domestication genetics, phylogenies that are based on single genes are of very limited use, because the alleles at single nuclear genes are much older than the populations themselves. Instead, measures of

genome-wide similarity, as provided by AFLP or Single nucleotide polymorphism (SNP) alleles, are more useful for unraveling domestication history (Salamini, 2002). Through cytological methods, genetic variation among related taxa can be assessed by comparing the organization of their chromosomes. Various inversions, duplications, translocations and ploidy changes are known to distinguish crop plants from their wild progenitors.

1.3.2 Molecular markers

Various DNA-fingerprinting techniques have been used in recent years to reveal the existence of alternative alleles at DNA loci (encoded by the nucleus or chloroplast or mitochondria) (Martin, 2000). Among them are restriction-fragment length polymorphisms (RFLP), randomly amplified polymorphic DNA (RAPD), AFLP and single-nucleotide polymorphisms (SNPs). RFLP and RAPD alleles are often due to sequence variation at restriction-enzyme recognition sites and primer-binding sites, respectively, but can also be due to length polymorphisms in the restricted or amplified region. AFLPs can be detected through a PCR-based procedure. DNA is usually digested with two restriction enzymes (one with a tetrameric recognition site and one with a hexameric) to yield fragments with overhanging ends; these are ligated to adaptors with primer-binding sites, which allows selective amplification of the fragment. The use of a labeled primer, usually for the hexamer site, yields a pattern of bands in a sequencing gel that is dense enough to reveal differences between fragments but simple enough to be interpreted (Vos et al, 1995).

1.3.3 Genetics of domestication

Until recently researchers interested in domestication were limited to studying phenotypic changes or the genetics of simple Mendelian traits, when often the characters of most interest- fruit size, yield, height, flowering time, etc. – are quantitative in nature (Jeffrey, 2005). The last 15 years, however, have seen an outpouring of data on the genetic basis of quantitative traits (Jeffrey, 2005). Three general types of molecular tools feature prominently for quantitative trait loci (QTL) mapping. RFLP maps were used in most of the early QTL mapping efforts in major crops and many of these involved interspecific crosses between crops and their wild ancestors (Paterson, 2002). Simple sequence repeat (SSR) markers have been essential in the primary molecular mapping of taxa such as soybean and have been especially important in the detailed characterization of elite crop gene pools comprised of closely related individuals (Marino et al., 1995). However, SSRs are relatively difficult to develop, and have proven to detect only modest levels of DNA polymorphism in some recently formed polyploids such as groundnut (Hopkins et al., 1999). While several other methods have been described including arbitrary primer (AP) – PCR (Welsh and McClelland, 1990), and RAPD (William et al., 1990), the most widely used one is presently the AFLP method (Vos et al., 1995). They are suitable for rapid assembly of data with a minimum of *priori* sequence information. Perhaps the most widely used pattern to emerge from QTL mapping studies in study of domestication syndrome has been the clustering of QTL (Jeffrey, 2005). Linkage maps of watermelon (*C. lanatus*) are available to-date. The first one, covering 354 cM and based on isozymes and seed protein, revealed the loci for flesh color (Navot and Zamir, 1986). The second one constructed with RAPDs, isozymes and RFLPs spanning 524 cM revealed the loci for rind color and flesh color (Hashizume et al., 1996). The latest one constructed with RAPDs, RFLPs and ISSR markers covering 1,729 cM revealed the loci for hardness of

rind, Brix of flesh juice, fresh color (red or yellow) and rind color (Hashizume et al., 2003). Most mapping studies have found that QTL are not randomly or even uniformly distributed throughout the genome (Paterson, 2002). However, in the linkage map of watermelon, QTL clustering has not yet been detected (Hashizume, 2003).

1.4 Chloroplast Genome

Chloroplasts are highly polyploid organelles in plants containing circular DNA molecules of 85 to 200 kb as well as the entire machinery necessary for the process of photosynthesis. The isolation of a unique DNA species in chloroplasts (e.g. Sager and Ishidia, 1963) has led to intensive studies of both the structure and expression of chloroplast genomes. Information is available on the structure and characterization of genes on the chloroplast genome (e.g. Deno et al., 1982; Yamada et al., 1986; Woitsch et al., 2003;), expression and function of proteins coded by chloroplast genes (e.g. Wang et al, 2000; May et al, 2000.) and metabolitic pathways of chloroplast (e.g. Klaus et al., 2002). Chloroplasts are genetically autonomous and information specifying components of the organelle protein synthesizing system is divided between organelle and nucleus. Chloroplasts of higher plants synthesize more than 80 polypeptides. In a typical higher plant, the chloroplast genome contains two copies of an inverted repeat (IR), which is usually 20 to 30 kb in size and contains genes encoding the chloroplast rRNAs, certain tRNAs, and often one or more genes specifying proteins. The rRNA operon is usually oriented with the 23s rRNA gene closer to the small single-copy region and the 16s rRNA gene close to the large single-copy region. Nearly two-thirds of the variation in size among higher plants chloroplast genome is accounted for by expansion or contraction of the IR (Palmer, 1991). The pea (*Pisum sativum*), the broad bean (*Vicia faba*) and the liverwort are exceptions and only contain one of the repeats,

resulting in correspondingly smaller chloroplast genomes (Wallace, 1982). Chloroplast genes can be categorized functionally into three main classes: genes related directly to photosynthesis, such as rubisco subunit genes (*rbcL*), photosystem genes (*psa*), photosystem genes (*psb*), cytochrome b/f complex genes (*pet*), ATP synthase genes (*atp*) and genes that encode NADH dehydrogenase (*ndh*), those involved in transcription and translation, and those encoding enzymes involved in biosynthesis of small compounds. However, most photosynthesis system genes are encoded by nuclear genes, the components of which are synthesized in the cytoplasm and transported into the organelle (Sugiura, 1992). The chloroplast genome has a low rate of structural and sequence evolution {i.e. low intraspecific chloroplast DNA (cpDNA) variation}, recombination is rare (or absent) and the genome is inherited uniparentally (Harris & Ingram, 1991). In most angiosperms, chloroplast genome is inherited maternally, which allows a direct study of seed-meditated dispersal and gene flow. However, cpDNA can be biparentally inherited as in *Medicago* (Masoud et al., 1990) and *Pelargonium* (Metzlaff et al., 1981) or even paternally inherited as in several gymnosperms such as *Pinus* (Wagner et al., 1987). The gene expression in the chloroplast is regulated by the function of a core set of chloroplast gene products in photosynthesis and electron transport (Allen, 2003).

CpDNA is increasingly being used by plant systematists because cpDNA is highly conserved and evolves fairly slowly at the nucleotide sequence level so that it is very useful in determining phylogenetic relationships (Havey, 1990). CpDNA is still being used as a tool to study intrafamilial relationships, the evolutionary position of genera, the origin and evolution of species and the degree and partioning of cpDNA variation within species (Harris & Ingram, 1991). CpDNA is a relatively abundant component of plant total DNA, therefore facilitating extraction and analysis. After DNA sequencing

techniques were developed, chloroplast DNA molecules were selected as one of the first targets of ‘the genome projects’, as they are relatively small and simple compared to nuclear genome. Most genes in the chloroplast genomes are essentially single-copy. In contrast, most nuclear genes are members of multigene families, which can compromise the phylogenetic utility of these genes. The entire nucleotide sequences of 12 chloroplast genomes from higher plants have been determined, disclosing an enormous amount of functional and evolutionary information (Wakasugi, et al., 2001). The complete nucleotide sequence of cpDNA was established for dicot tobacco (*Nicotiana tabacum*) (Shinozaki et al, 1986), for bryophyte liverwort (*Merchantia polymorpha*)(Ohyama et al, 1986) and monocot rice (*Oryza sativa*) (Hiratsuka et al, 1989), for *Epifagus virginiana* (Wolfe et al., 1992), gymnosperm black pine (*Pinus thunbergii*) (Wakasugi et al, 1994), maize (*Zea mays*) (Maier et al, 1995), *Chlorella vulgaris* (Wakasugi et al, 1997) *Arabidopsis thaliana* (Sato et al, 1999), *Oenothera elata* ssp. *Hookeri strain Johansen* (Hupfer et al, 2000), *Lotus japonicus* Miyakojima MG-20 (Kato et al, 2000), wheat (*triticum aestivum* cv. Chinese spring) (Ogihara et al, 2000) and spinach (*Spinacia oleracea*) (Schmitz-Linneweber, 2001). The completion of the chloroplast genome sequence of the chlorophyte alga *Chlamydomonas reinhardtii*, which is the most genetically and biochemically tractable eukaryotic model system for photosynthesis and chloroplast gene expression has been announced by Simpson, et al in 2002. In 2003, the complete chloroplast DNA sequence (122,890 bp) of the moss *Physcomitrella patents* has been determined, which contains 83 protein, 31 tRNA and 4 rRNA genes and a pseudogene (Sugiura, 2003). Studies of the complete chloroplast DNA sequences from several other plants are still in progress.

Four main approaches employ the chloroplast genome to infer relationships: (1) restriction site analysis; (2) structural changes in the chloroplast genome, including

inversions, large deletions, and the loss of specific introns and genes; (3) comparative DNA sequencing and (4) PCR based approaches. PCR-RFLPs of cpDNA have been studied extensively in plants, and have proven to be valuable for molecular systematic studies above the species level (Clegg, 1993; Jansen et al, 1998) as well as for phylogeographic analyses within species (Newton et al, 1999; Schaal et al, 1998). A defined DNA sequence is amplified using a sequence-specific primer pair. This may already result in differently sized and hence informative PCR fragments. Then the PCR product is digested with a restriction enzyme, usually with a four-base recognition specificity. The digested amplification products may or may not reveal polymorphisms after seperation on agarose gel. Because only a subset of base substitutions is targeted, small insertion-deletion events may escape detection.

Comparative DNA sequences analysis is the area of molecular systematics in which the greatest advances have been made. PCR and methods for direct sequencing of PCR products have resulted in a mushrooming of sequence data. In theory, any degree of divergence is amenable to comparative sequencing analysis. In practice, plant systematists have focused on two slowly evolving sequences (*rbcL* and rRNA genes). The gene *rbcL* is located in the large single-copy region of the chloroplast genome and was one of the first plant genes to be sequenced. Sequence data derived from *rbcL* have been used to address phylogenetic relationships not only in angiosperms, but also in ferns (Hasebe et al, 1993), and various groups of algae (McCourt, 1995), The application of *rbcL* sequence data spans a very wide taxonomic range. More rapidly evolving DNA sequences, including rapidly changing chloroplast genes, chloroplast introns (e.g. *rpl*16, *rpo*C1, *ndh*A), and intergenic spacers (e.g. *acc*D-*psa*I, *trn*L-*trn*F, *trn*T-*trn*L, *atp*B-*rbc*L), and the noncoding portions of cpDNA,

also are being investigated for comparative purposes (Olmstead and Palmer, 1994). More recently, sequencing of cpDNA noncoding regions (introns and intergenic spacers) has become popular for analyses at various taxonomic levels (**Randall et al, 1998**). Noncoding regions have been presumed to be more useful at lower taxonomic ranks because they are less functionally constrained and are therefore freer to vary, thereby potentially providing more phylogenetically informative characters per unit of sequencing effort (Clegg et al., 1994).

1.5 Nuclear genome

Systematists have become increasingly aware that reliance on a single data set may result in insufficient resolution or an erroneous picture of phylogenetic relationships. As a result, it is now common practice to use multiple data sets for phylogenetic inference (Soltis et al., 2000). Moreover, inferring phylogenetic relationships among closely related plant species is often difficult due to the lack of molecular markers with enough nucleotide variability at this taxonomic level. A gene tree does not necessary represent the true species tree because of random sorting of polymorphic alleles in different lineages (Despres et al., 2003). Potential problems due to cpDNA introgression among closely related taxa and the lack of phylogenetic resolution stimulated the development of new approaches based on nuclear DNA. Nuclear DNA has been relatively unexplored compared with cpDNA, with the exception of the nuclear ribosomal DNA region, which has been sequenced on a large scale. The internal transcribed spacers (*ITS*1 and *ITS*2) are applied extensively to phylogeny reconstruction at low taxonomic levels (Grob et al., 2004). The 18S ribosomal RNA gene has been the most widely used nuclear sequence for phylogeny reconstruction at higher taxonomic levels in plants (e.g., Chaw et al., 1997; Hamby and

Zimmer, 1992; Soltis et al., 1997). Kuzoff et al. (1998) demonstrated the potential of entire 26S rDNA sequences for phylogeny reconstruction at taxonomic levels comparable to those investigated with 18S rDNA. The phylogenetic utility of 26S rDNA sequences in higher taxonomic levels plants has also been demonstrated by recent studies (e.g., Fan and Xiang, 2001; Fishbein et al., 2001; Zanis et al., 2002; 2003). 5.8S rDNA and 5S rDNA have also been used to infer phylogeny, however, these regions are too conserved and too small to be informative at even deep taxonomic level (Troitsky and Bobrova, 1986). Only the last few years nuclear DNA regions have been investigated for their phylogenetic utility. A few examples of these are the genes encoding small subunit of ribulose 1,5-biphosphate carboxylase (*rbc*S), chalcone synthase (*Chs*) (Clegg et al., 1997), alcohol dehydrogenase (*Adh*) in Paeoniaceae (Sang et al., 1997), Cycloidea in Gesneriaceae (Möller et al., 1999), granule-bound starch synthase (GNSSI, or waxy) in Poaceae (Mason-Gamer et al., 1998), malate synthase in Arecaceae (Lewis and Doyle, 2001), chloroplast-expressed glutamine synthetase (ncpGS) in Oxalidaceae (Emshwiller and Doyle, 1999), pistillata in Brassicaceae (Bailey and Doyle, 1999), vicilin in Sterculiaceae (Whitlock and Baum, 1999), and glyceraldehyde 3-phosphate dehydrogenase (*G3pdh*) in *Mitthyridium* (Musui: Calymperacear) (Wall, 2002). ITS sequences are sometimes unsuitable for phylogenetic studies due to high sequence divergence (Wilson, 2003), extensive length variations between copies (Liston et al., 1996), paralogy problems (Baker et al., 2000), or lack of resolving power (Whitcher and Wen, 2001). Genes encoding *rbc*S, *Adh*, and *Chs* exist as multi-gene families and thus also present problems of paralogous evolution similar to ITS sequences (Clegg et al., 1997). Moreover, these genes, especially *Adh* and *Chs*, may have undergone excessive recombination that could have clouded their actual history (Clegg et al., 1997). Most recently, the AFLP approach has

been used. This technique has the potential to solve such difficulties, particularly among closely related species, or at the intraspecific level (eg., Koopman et al., 2001; Talhinhas et al., 2003; Guo et al., 2005). AFLP techniques are highly reproducible and provide a large number of informative markers derived from loci dispersed throughout the nuclear genome (Ridout and Donini, 1999). The power of AFLP analysis is tremendously high for revealing genomic polymorphisms. All of the approximately 500,000 fragments generated by *EcoR - Mse* digestion of a 10^9-bp genome caused by deletions, insertions, and primer site base substitutions can be revealed by a full AFLP scan of the 4096 primer combinations (Liu and Cordes, 2004). In barley, for example, AFLP markers were found to be located on the long and short arms of all seven chromosomes, with a strong correlation between the number of markers per chromosome and the length of the chromosome (Waugh et al., 1997). In rice, the AFLP technique was used to map an F2 population from an indica japonica cross (each representing major breeding groups). More than 50 AFLP markers were spread across all the chromosomes except chromosome 12 (Mackhill et al., 1996). The fact that AFLP markers are generally distributed across the genome gives the technique some advantages over sequence analysis for closely related taxa. AFLP however are limited to the analysis of closely related species and infra-specific taxa (Vekemans et al, 2002). Above this level the multi-locus fingerprint becomes too variable, and markers are less likely to be homologous (Hodkinson et al, 2000).

1.6 Mitochondrial genome

Molecular study of the mitochondrial genome involving either restriction site analysis or sequencing has been a major focus of phylogenetic studies in animals (e.g. Avise, 1986, 1994; Mitton, 1994). In contrast, the mitochondria DNA has been little

used in studies of plant phylogeny. Mitochondria DNA is hard to isolate and less abundant in plant leaves compared to cpDNA. This lack of emphasis on the plant mitochondrial genome is due to the following (Palmer, 1992): (1) plant mtDNAs are very large and highly variable in size and stability as compared to animal mtDNAs; (2) many foreign DNA sequences, particularly chloroplast DNA sequences, are present in plant mitochondrial genomes; (3) large duplications (repeats) appear to be frequently created and lost; (4) recombination can occur among repeats with high frequency, creating a very complex genome structure; (5) small (1-11 kb) unstable, extra chromosomal plastids of unknown region and function are common in plant mitochondria and may be transmitted differently than the main mitochondrial genome; (6) plant mtDNA are characterized by many short (50-1,000 bp), dispersed repeats that appear scattered throughout the mitochondrial genome; such repeats are considered rare in other mtDNAs and in cpDNA; (7) plant mtDNA rearrange very quickly, the result being that even closely related species do not possess the same order of mtDNA genes; (8) the rate of nucleotide substitution is very low compared to the mitochondria genome of other organisms, as well as compared to the chloroplast and nuclear genomes of plants. Thus, most of the complete mtDNA sequences published are from animals, protests and ascomycete fungi, whereas the available sequences for plants are limited to the liverwort *Marchantia polymorpha* (Oda et al, 1992), the angiosperm *Arabidopsis thaliana* (Unseld et al, 1997), and the sugar beet (*Beta vulgaris* L.)(Kubo et al, 2000). There are 23 species for which both cpDNA and mtDNA data are available, including eight cases in which both genomes are maternally inherited (Petit et al., 2005). There are 29 species with data from both paternally inherited (cpDNA) and biparentally inherited markers (all conifers); for 13 of these, data was also available from mtDNA (i.e. data was available from three differentially inherited genomes)

(Petit et al, 2005).

1.6 Project objectives

This study was conducted to provide an intraspecific phylogenetic framework for *Citrullus lanatus* based on the comparative chloroplast and nuclear *G3pdh* sequence data and AFLP markers. The objectives of the present investigations are: (1) to detect the phylogenetic relationships within the species and the possible origin of the genus; (2) to shine light on the possible domestication pattern of the species; (3) to compare the substitution rate of different chloroplast regions with nuclear *G3pdh* region and evaluate the phylogenetic utility of the cpDNA regions; (4) to compare the cpDNA sequence data with nuclear *G3pdh* sequence data and AFLP data in order to determine which method is more useful for phylogenetic research.

Table 1. List of six centers of origin of major domesticated crops (Gepts, 1998)

Major Centers of Origin	Major Domesticated Crops
Mesoamerica	maize, beans, cotton, pepper, tomato and etc
South America	potato, cotton, pepper, beans
Near East	wheat, barley, carrot, apple
Africa	rice, cotton, sorghum, watermelon
China	rice, soybean, citrus, chestnut
Southeast Asia	rice, cucumber

II. PHYLOGENY AND BIOGEOGRAPHY OF WATERMELON [*CITRULLUS LANATUS* (THUNB.) MATSUM. & NAKAI] BASED ON CHLOROPLAST AND NUCLEAR SEQUENCE DATA

Introduction

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai], together with cucumber, melons of various sorts, summer squash, winter squash and pumpkin are the principal food plants of the gourd family (Cucurbitaceae). The genus *Citrullus* has now been revised to include C. *lanatus* (syn. C. *vulgaris*), C. *ecirrhosus*, C. *colocynthis*, and C. *rehmii* (De Winter, 1990). *Citrullus ecirrhosus* is more closely related to C. *lanatus* than either is to C. *colocynthis* (Maynard, 2001). Watermelon has 22 chromosomes (2n=22, x=11). Watermelon is a monoecious, warm-season crop. Flowering and fruit development are promoted by high light intensity and high temperature. Watermelon is the only economically important cucurbit with pinnatifid (lobed) leaves; all of the other species have whole (nonlobed) leaves. In the USA, watermelon is cultivated mainly in the southeastern (Alabama, Florida and Georgia), southwestern (California, Texas and Arizona) and central (Indiana) states. Despite the economic importance of watermelon, domestication events and phylogeographic relationships have only recently attracted scientific attention.

Chloroplasts are highly polyploid organelles in plants containing circular DNA

molecules of 85 to 200 kb as well as the entire machinery necessary for the process of photosynthesis. Information is available on structure and characterization of genes on the chloroplast genome (e.g. Deno et al., 1982; Yamada et al., 1986; Woitsch and RÖMER, 2003;), expression and function of proteins coded by chloroplast genes (e.g. Wang et al., 2000; May and Soll, 2000.) and metabolic pathways of chloroplast (e.g. Klaus et al., 2002). CpDNA is increasingly being used by plant systematists because cpDNA is highly conserved and evolves fairly slow at the nucleotide level so that it is very useful in determining phylogenetic relationships (Harvey, 1990) and is still being used as a tool to study intrafamilial relationships, the evolutionary position of genera, the origin and evolution of species and the degree and partitioning of cpDNA variation within species (Harris and Ingram, 1991).

Initial studies using PCR-RFLP analysis of cpDNA in *Citrullus* species were conducted using nine different chloroplast regions (Dane, 2002). CpDNA studies using PCR-RFLP analysis revealed 7 haplotypes with nucleotide diversity at the *ndh*F, *rpl*16, *trn*S-*trnf*M, *trn*C-*trn*D and *trn*T-*trn*F five main regions. Three haplotypes were detected within *C. lanatus*, one haplotype being associated with the cultivated watermelon and two with wild citron types (Dane, Lang, and Bakhtiyarova, 2004). Also, cpDNA region *ndh*F and the intergenic region of *atp*A sequence analysis showed two main clades among *Citrullus,* one of which contains *C. colocynthis* and in the other clade, *C. rehmii* is sister to a clade containing *C. ecirrhosus* and *C. lanatus* (Dane and Lang, 2004).

This study was conducted to provide an intraspecific phylogenetic framework for *Citrullus lanatus* based on comparative chloroplast and nuclear gene sequence data. The objectives of the present investigations are: (1) to detect the phylogenetic relationships within the species and its origin; (2) to shine light on the possible

domestication pattern of the species; (3) to compare the substitution of different chloroplast regions, evaluate the phylogenetic utility of the cpDNA regions and nuclear gene.

Materials and Methods

Plant materials and DNA extraction. Seeds from *C. lanatus* var. *citroides* and *C. lanatus var. lanatus* accessions were obtained from the Plant Introduction (PI) Station at Griffin, GA, from The Cucurbit Network (TCN) (http://www.cucurbit.org), and from National Plant Genetic Resources Center at Windhoek, Namibia (NAM). *C. lanatus* var. *lanatus*. 'Crimson Sweet' is a common US cultivar. *C. colocynthis* accession 34256 was obtained from Z. Yaniv at the Volvani Center, Israel and used as outgroup (Table 1). DNA was extracted from seeds or young leaf tissue using the mini Qiagen plant DNAeasy extraction kit (Qiagen, Valencia, California 91335, USA). DNA concentration was estimated visually using DNA low mass ladder [Invitrogen™ Life Technologies (Carlsbad, CA, USA)] and electrophoresis in a 2% agarose gel.

PCR-RFLP analysis. Noncoding cpDNA regions were initially amplified using the following primer pairs: primer set TrnS-F and TrnG(UCC)-R of *trn*S-*trn*G region (Doyle et al., 1992), ccSSR4F and ccSSR4R of *trn*R-*atp*A region (Chung, Decker-Walters, and Staub, 2003), psaA-f and trnS1-M of *psa*A-*trn*S region (Demesure, Sodzi, and Petit, 1995), ucp-a and ucp-b of *trn*T-*trn*L exon 1 region (Taberlet et al., 1991), trnE Doyle (Doyle et al., 1992) and trnT-M (Demesure, Sodzi, and Petit, 1995) of *trn*E-*trn*T region, trnF and trnV1 of region *trn*F-*trn*V (Dumolin-Lapegue, Pemong, and Petit, 1997) and ORF184-f and petA-r0 of orf184-*pet*A region (Grivet et al., 2001) (Table 2; Figure 1). Double stranded DNA amplification was performed in a 20-ul volume containing 1x PCR buffer of 20 mmol/L Tris HCL (pH 8.4) and 50 mmol/L KCL [5ul of 10x PCR buffer (Invitrogen)],

2 mmol/L $MgCl_2$, 200 umol/L of each dNTP, 0.2 umol of each primer, 2 U of T*aq* Polymerase (Invitrogen), and 2 ul template DNA (20 ng/ul). PCR cycling conditions for the *trn*S-*trn*G region were 94C for 4 min, followed by 35 cycles of 1 min 94°C, 1 min 52°C, 2 min 65°C, followed by a final extension of 10 min 65°C; for *trn*R-*atp*A region, *trn*T-*trn*L region, *trn*F-*trn*V region and orf184-*pet*A region the same procedure was used except for the different annealing temperature of 55°C; for *psa*A-*trn*S region the annealing temperature as at 57.5°C; for *trn*E-*trn*T region 50°C. PCR-products were digested with *Taq*I restriction enzyme at 65°C for at least 6 hours.

cpDNA and nuclear gene sequence analysis. Since PCR-RFLP analysis of the *trn*S-*trn*G and *atp*A-*trn*R region using *Taq*I revealed variability within *C. lanatus*, these regions were selected for sequence analysis. A total of 18 accessions with different restriction enzyme patterns and different areas of origins were chosen. PCR amplification was performed as described earlier except the volume of the reaction was increased to 50 ul. The same 18 *C. lanatus* accessions were used for sequence analysis of the intron 2 of the transit peptide region of the *G3pdh* nuclear gene. This region was amplified using primers [*G3pdh*F (CAG GCT AAT GGA AAG GGT TT) and *G3pdh*R (TTG TAT CCT CCG CTC CTT CC)] designed from the published expressed sequence tag (EST) of *C. lanatus* (GenBank accession PI563173). For *G3pdh* sequence analysis, the PCR procedure was as follows: 94°C for 4 min, followed by 35 cycle of 1 min at 94°C, 1 min at 52.5°C, 2 min at 72°C followed by a final extention of 10 min at 72°C. The amplified fragments were purified using the Qiaquick PCR purification kit (Qiagen). Sequencing analysis was performed using ABI 3100 sequencer at the Genomics and Sequencing lab of Auburn University. Sequences are deposited in GenBank and accession numbers are shown in Table 4.

Sequence alignment and Data analysis. ContigExpress program implemented in

the Vector NTI™ software was used to assemble many small fragments into longer contiguous sequences. Multiple alignments of the sequences were obtained using the AlignX program implemented in the VectorNTI software, followed by manual adjustments. Insertion / deletion events (indels) that were potentially parsimonious were scored and added to the end of the data sets as single binary (0 vs 1) characters (Graham et al., 2000). Areas of ambiguous alignment or with poly-n strings were excluded from all analyses. Phylogenetic analyses were conducted using maximum parsimony (MP), maximum likelihood (ML), as implemented in PAUP* software (Swofford, 2002). MP was performed by branch-and-bound search using stepwise addition. ML analysis was carried out in heuristic search using the tree-reconnection (TBR) branch swapping and random sequence addition with 100 replications. A (50%) majority-rule consensus tree was constructed and the robustness of nodes was inferred by bootstrap analysis (Felsenstein, 1985) of 1000 replicates. Nucleotides were treated as independent unordered characters of equal weight.

Results

CpDNA analysis of *trn*S-*trn*G intergenic spacer region. PCR-RFLP analysis of *trn*S-*trn*G/*Taq*I studies of 22 *C. lanatus* var. *lanatus* and 56 *C. lanatus* var. *citoides* accessions (Table 1) from widely different origin resulted in a total of 3 different patterns. Pattern II and pattern III were observed within *C. lanatus* var. *citroides*, while Pattern I was detected in *C. lanatus* var. *lanatus* and *C. colocynthis*. Only PI 532667 and 596656 have the unique pattern III (Figure 2; Table 3) at *trn*S-*trn*G. Sequences from 18 selected taxa included in the study provided an aligned length of 869 bp and A+T content of the sequences is 75.3% (Table 7). Two transversions, one 3 bp (T→A) and 1 bp (T→G) are unique to *C. lanatus* var. *citroides* accessions, one 30 bp deletion was detected in all *C. lanatus* var. *citroides* accessions with the exception of PI 532667

and PI 596656. The one base pair transversion unique to the *C. lanatus* var. *citroides* accessions resulted in an additional *Taq*I site. A total of five *Taq*I sites (TCGA) are present in *C. lanatus* var. *lanatus* and *C. colocynthis* accessions, resulting in fragments which can be detected on agarose electrophoresis gel of around 345, 200, 162, and 67 bp (Figure 3). The additional *Taq*I site in *C. lanatus* var. *citroides* accessions resulted in fragments of around 345, 162, 127 or 97, and 63 bp. The difference in PCR-RFLP pattern II and III is the result if a 30 bp deletion in all *C. lanatus* var. *citroides* accessions except for PI 532667 and 596656.The INDEL events and the segments of the aligned sequences that were defined are shown in Table 5. Both sites were flanked by A and/or T nucleotides.

CpDNA analysis of *trn*R-*atp*A intergenic spacer region. PCR-PFLP of the *trn*R-*atp*A region using *Taq*I resulted in two banding patterns. Part of the *C. lanatus* var. *citroides* accessions showed a 23 bp insertion (Table 5). Sequencing analysis of 18 selected taxa provided an aligned length of 610 bp with 85.1 % A+T content (Table 7) and showed nucleotide substitutions, 2 transversions (A→C, G→A) and 1 transition (T →A) between *C. lanatus* var. *lanatus* and all *C. lanatus* var. *citroides* accessions, while the different banding patterns observed in *C. lanatus* var. *citroides* were the result of a 23 bp insertion homologous to a flanking region. This insertion was detected in accessions from southern Africa, India and USA. The events of 30 bp deletion on *trn*S-*trn*G region and 23 bp insertion on *trn*R-*atp*A region must have happened relatively recently, because outgroup *C. colocynthis* 34256 and all other accessions of *C. colocynthis* do not have the deletion or duplication (Dane, unpublished results).

Phylogenetic Analysis. Combined sequence data analysis with indels scored as presence/absence (1/0) characters using maximum parsimony group the *C. lanatus* var. *lanatus* accessions into one clade (97% bootstrap support) and the *C. lanatus* var.

citroides accessions into three main clades [50% majority-rule consensus of 4483 trees with consistency index (CI) of 0.8387 and retention index (RI) of 0.8384]. Clade (99% bootstrap support) groups accessions PI596656 from S. Africa and PI532667 from Swaziland. Clade (62% bootstrap support) includes *C. lanatus* var. *citroides* accessions from Namibia, S. Africa, Cape Province, Transvaal, Zaire, India and USA. Clade can be divided into two subclades, with 69% bootstrap support for TCN1337 and PI485583 (Figure 4).

***G3pdh* sequence analysis.** The sequenced *G3pdh* region covers one intron (2), flanked by AG and CT motifs typical of nuclear introns (Peterson et al., 2003), and a short section of the transit peptide region. The region shows high similarity (78%) to the transit sequence of the mature subunit sequence (βA and αB) of *GpaB* from pea (*Pisum sativum*) (Dane, unpublished results). Sequence divergence is high only at intron 2 region of the transit peptide region. The length of sequences among 18 taxa varies from 596bp to 1018bp. A total of 10 parsimony informative sites are detected in this region with A+T content of 63.6% and a transition and transversion ratio of 4: 6. *C. lanatus* var. *lanatus* can be distinguished from all other *Citrullus* species by the presence of one unique transition, 2 unique transversions and one unique 4 bp deletion. Within *C. lanatus* var *lanatus* 2 haplotypes can be detected. PI 494529 and Crimson Sweet have a 3 bp deletion (Table 6). Within *C. lanatus* var. *citroides* accessions a total of 0, 1 (Tv), 2 (2Ts or 2 Tv), 3 (3Tv) or 4 (2Tv + 2 Ts) unique nucleotide substitutions were detected (Table 6).

Trees constructed using *G3pdh* sequences in PAUP similarly group the *C. lanatus* var. *lanatus* accessions into one clade (with 96% bootstrap support) and the *C. lanatus* var. *citroides* accessions into three sub clades [50% majority-rule consensus of 3012 trees with consistency index (CI) of 0.8519 and retention index (RI) of 0.8400]. Clade

 (with 64% bootstrap support) groups PI596656, PI296343 from South Africa, nam1569 from Namibia, PI271769 from Transvaal and PI189225 from Zaire together. Clade (with 67% bootstrap support) includes the rest of the *C. lanatus* var. *citroides* accessions from southern part of Africa, India and USA regions. Clade (with 64% bootstrap support) include nam958 and nam1612 from Namibia (Figure 5).

To test for incongruence between two data partitions, cpDNA sequences and *G3pdh* sequences, the partition homogeneity was implemented (Farris et al., 1995; Mason-Gamer and Kellogg, 1996). One hundred replicates were used for each partition to generate the null distribution using PAUP* 4.0. The result of partition homogeneity test demonstrated that cpDNA sequences and *G3pdh* sequences are congruent (P=0.24). Combined sequence analysis of cpDNA and G3pdh, divides all the *C. lanatus* accessions into 11 cytotypes (Table 6). By combining these 2 data partitions together, the 50% majority-rule consensus tree [consistency index (CI) of 0.7500 and retention index (RI) of 0.6897] changed slightly as followed: 100% support for the clade of *C. lanatus* var. *lanatus* accessions, clade (90% bootstrap support) includes PI271769 and PI189225. Clade I (with 61% bootstrap support) has TCN1337 and PI485583. Clade II (with 93 bootstrap support) groups nam958 and nam1612 together. Clade III (with 100% bootstrap support) includes PI596656, 296343, 270563, 288316, 532667, nam1569, 1884 and TCN1126 (Figure 6).

Discussion

The nature of intraspecific cpDNA polymorphism detectable using PCR-RFLP is typically limited to restriction site changes and INDEL mutations (Tsumura et al., 2000). One 30bp deletion was found on *trn*S-*trn*G region and one 23bp homologous duplication was found on *trn*R-*atp*A region. The high A+T content of *trn*R-*atp*A region (85.1%) may explain the 23bp duplication. No substitutions were detected in

*trn*S-*trn*G and *trn*R-*atp*A regions within *C. lanatus* var. *citroides* while a total of 12 substitutions (1.8%) in *trn*S-*trn*G region were detected in *Vaccinium uliginosum* L. *sensu lato* (Alsos et al, 2005), indicating that the evolution of *trn*S-*trn*G region on chloroplast genome of *C. lanatus* is slow and conservative. Also, a Cucurbitaceae study showed that the *trn*R-*atp*A region has more useful sequence information than any other region studied (Chung et al., 2003). The 5 substitutions detected were found between *C. lanatus* var. *citroides* and *C. lanatus* var. *lanatus* only, which indicates low genetic variability on both *trn*S-*trn*G and *trn*R-*atp*A cpDNA regions. Also studies of other cpDNA regions such as *psa*A-*trn*S, *trn*T-*trn*L, *trn*E-*trn*T, *trn*F-*trn*V and orf184-*pet*A regions (Dane & Lang, 2004; Liu, results) failed to detect variation using PCR-RFLP and *Taq*I, and were consequently not studied further.

Higher sequence variability was detected at the nuclear *G3pdh* region. Sequencing analysis did identify two different *C. lanatus* var. *lanatus* types and based on cpDNA and *G3pdh* sequencing data, 9 cytotypes were identified among *C. lanatus* var. *citroides* accessions which can be grouped into two distinct lineages (Figure 6) with different nuclear profiles. Since chloroplast genomes are maternally inherited in cucurbits (Harvey, 1990), it can be hypothesized that PI 532667 and 596656 are probably the oldest *C. lanatus* var. *citroides* accessions. The accessions with the 30 bp deletion, TCN1337, nam958 and 1612 might have evolved from PI 532667. Similarly PI 189225 and 271769 might have evolved from PI 596656. Accessions with the 23 bp insertion can be considered more recent. Because of possible recombination during seed increases, nuclear lineages are more problematic and speculative. The early European colonists in the middle of 16th century introduced watermelon into the USA in post-Columbus times. Since the cpDNA haplotype of TCN1337 resembles that of PI

189225 and 271769, we can hypothesize that the accession probably originated in southern Africa. The 50% majority consensus tree based on combined cpDNA and *G3pdh* sequences groups PI 485583 and TCN 1337 into a clade with 61% bootstrap support, which supports the hypothesis that this US accession migrated from the southern part of the African continent. The accession from India PI 288316 can be grouped with accessions from southern Africa and might similarly have originated from southern Africa. It can be deduced that southern Africa (Transvaal, Cape Province and Namibia) might be the area of origin of *C. lanatus* var. *citroides*, with colonization routes from this area all over the world (Figure 7). However, more accessions with wider geographic distribution should be studied to support this conclusion.

It is clear from the cpDNA and nuclear *G3pdh* sequence analysis that there are 2 distinct variable lineages within the species *C. lanatus*. Intermediate types between *C. lanatus* var. *citroides* and var. *lanatus* were not detected and almost no diversity within the *C. lanatus* var. *lanatus* types. PI 454929 is an Egusi type watermelon, which seeds are used in Nigeria as a source of grain. This accession is completely homologous with cultivated US accession and only differs from PI 179881 from India by a 3 bp deletion in a nuclear transit peptide region. This indicates that var. *lanatus* and *citroides* evolved from the same ancestor millions of years ago. Since no diversity was detected within var. *lanatus*, little can be postulated about its origin.

In conclusion, this study demonstrates the use of both noncoding cpDNA and single copy nuclear DNA to examine the phylogeography of a plant species. Few nuclear genes are currently available for phylogenetic studies, especially ones that allow reconstruction of historical relationships. Three haplotypes were detected using

a total of 1.5 kb region of cpDNA, however with the genetic variation at a 0.6 kb region of the transit peptide of the nuclear *G3pdh* a total of eleven unique *C. lanatus* haplotypes were detected. Information on both data partition was congruent and phylogenetically informative. The haplotypes are structured into two distinct nuclear lineages. The substitution rate at cpDNA is very low, which suggest that the *trn*S-*trn*G and *trn*R-*atp*A cpDNA regions evolved very slowly and conservatively. The occurrence of two unique indels at cpDNA region suggests South Africa as the area of origin and colonization patterns from South Africa to middle and northern part of African continent, India and America. *C. lanatus* var. *lanatus* and *C. lanatus* var. *citroides* appear to have evolved independently from a same common ancestor.

Table 1. List of investigated *C. lanatus* and *C. colocynthis* accessions and their geographical origin.

Species or variety	PI or TCN number	Origin
C. lanatus var. *citroides*	288316	India
	189225, 532738	Zaire
	244018, 255137, 271769, 296343, 596665, 244019, 271767, 596666	Transvaal
	270563, 596656, 596671	S. Africa
	248774, 296335, 485579, nam814, nam838, nam958, nam960, nam1310, nam1444, nam1525, nam1569, nam1607, nam1612, nam1621, nam1626, nam1628, nam1634, nam1678, nam1688, nam1884, nam1885, nam1901	Namibia
	379243	Yugoslavia
	482246, 482259, 482279, 482303, 482311, 482319, 482324, 482361, 482298, 482299, 482315, 532624, 482252	Zimbabwe
	532664, 532666, 532667	Swaziland
	TCN1126, TCN1360, TCN1337	U.S.
	512385, 512854	Spain
	296339, 296343, 296341, 596667	Cape Province
	500308, 500335	Zambia
	296335, 432717	Natal
C. lanatus var. *lanatus*.	179881	India
	255136, 271779, 295850, 295842	Transvaal
	254742, 254744	Senegal
	482334, 482336, 482273, 482293	Zimbabwe
	176492	Turkey
	211011	Afghanistan
	271778	S. Africa
	385964	Kenya
	494527, 494529	Nigeria
	500324, 500353	Zambia
	507858	Hungary
	‘Crimson Sweet’	USA

	536453	Maldives
C. colocynthis	386018	Iran
	220778	Afghanistan
	TCN955	Morocco
	14202	India
	525081	Egypt

Table 2. Primers used for PCR reaction and sequencing (from 5’ to 3’).

Primers	Cp regions	Length	Primer sequence	Reference
*Trn*S-F	TrnS	19	TACAACGGATTAGCAATCC	Doyle et al., 1992
*Trn*G(UCC)-R	*trn*G	20	ATACCACTAAACTATACCC	Doyle et al., 1992
ccSSR4F	*trn*R	23	AGG TTC AAA TCC TAT TGG ACG CA	Chung et al., 2003
CcSSR4R	*atp*A	24	TTT TGA AAG AAG CTA TTC ARG AAC	Chung et al., 2003
*psa*A-f	*psa*A	22	ACT TCT GGT TCC GGC GAA CGA A	Demesure, Sodzi, and Petit, 1995
*trn*S1-M	*trn*S	22	AAC CAC TCG GCC ATC TCT CCT A	Demesure, Sodzi, and Petit, 1995
ucp-a	*Trn*T	20	CAT TAC AAA TGC GAT GCT CT	Taberlet et al., 1991
ucp-b	*Trn*L exon1	20	TCT ACC GAT TTC GCC ATA TC	Taberlet et al., 1991
*trn*E Doyle	*Trn*-E	20	GCCTCC TTG TTG TTG AAA GAGAGA TG	Doyle et al., 1992
*trn*T-M(P*)	*Trn*-T	20	CTA CCA CTG AGT TAA AAG GG	Demesure, Sodzi, and Petit, 1995
ORF184-f	ORF184	19	TGG CGA TCA GAA CAY ATA TGG ATA G	Grivet et al., 2001
*pet*A-r0	*pet*A	25	CCCATTTTTGCACAGCAGGGTTATG	Grivet et al., 2001

Table 3. List of investigated *C. lanatus* and *C. colocynthis* accessions and their restriction enzyme (*Taq*I) patterns at cpDNA regions *trn*S-*trn*G and *trn*R-*atp*A.

Species or variety	PI or TCN number	*Trn*S-*trn*G	*Trn*R-*atp*A
C. lanatus var. *citroides*	244018, 255137, 288316, 596666, 270563, 296335, 296339, 296343, 596665, 596666, 596671, nam838, nam1310, nam1444, nam1569, nam1688, nam1884, nam1885, nam1901, 432717, 482361, 482298, 482299, 482361, 482315, 485583 532624, 532666, 596667, TCN1126	Pattern I	Pattern I
	189225, 296334, 271769 379243, 482246, 482259, 482279, 482303, 482311, 482324, 482361, 482252, 512854, 512385, 532664, 500308, 500335, TCN1337, TCN1360, nam814, nam958, nam960, nam1525, nam1607, nam1612, nam1621, nam1634, nam1626, nam1628, nam1678	Pattern II	Pattern I
	532667, 596656	Pattern III	Pattern I
C. lanatus var. *lanatus*.	179881, 244019, 271767, 295842, 482273 255136, 254742, 254744, 271778, 271779, 295850, 482319, 482334, 482336, 482293, 494527, 494529, 500324, 532738	Pattern II	Pattern II
C. colocynthis	14202, 386018, 220778, TCN955, 525081	Pattern II	Pattern II

Table 4. Genebank numbers of *trn*S-*trn*G, *trn*R-*atp*A, and *G3pdh* sequences from *C.lanatus* and *C.colocynthis* accessions.

C. lanatus var. *citroides*	Origin	*Trn*S-*Trn*G	*Trn*R-*atp*A	*G3pdh*
P.I. 596656	S. Africa			
P.I. 296343	Cape Province			
P.I. 271769	Transvaal			
P.I. 270563	S. Africa			
P.I.189225	Zaire			
P.I. 288316	India			
P.I. 532667	Swaziland			
P.I. 485583	Botswana			
NAM 958	Namibia			
NAM 1569	Namibia			
NAM 1884	Namibia			
NAM 1612	Namibia			
TCN 1337	U.S.A.			
TCN 1126	U.S.A			
C. lanatus var. *lanatus* P.I. 179881	India			
P.I. 494529	Nigeria			
Crimson Sweet	U.S.A.			
C. colocynthis 34256	Israel			

Table 5. Indels observed on cpDNA regions *trn*S-*trn*G and *trn*R-*atp*A in different *C. lanatus* accessions and *C. colocynthis*

Species or variety	Accessions	*trn*S-*trn*G	*trn*R-*atp*A
		Aligned sequence position 690-719	Aligned sequence position 190-212
C. lanatus var. *citroides*	PI596656	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTA----------
	PI296343	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	PI271769	----------	ACTAATAATTCTATTCTGTTTTA----------
	PI270563	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	PI189225	----------	ACTAATAATTCTATTCTGTTTTA----------
	PI288316	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	PI532667	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTA----------
	PI485583	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	Nam958	----------	ACTAATAATTCTATTCTGTTTTA----------
	Nam1569	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	Nam1884	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
	Nam1612	----------	ACTAATAATTCTATTCTGTTTTA----------
	TCN1337	----------	ACTAATAATTCTATTCTGTTTTA----------
	TCN1126	----------	ACTAATAATTCTATTCTGTTTTAACTAATAATTCTATTCTGTTTTA
C. lanatus var. *lanatus*	PI179881	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTG----------
	PI494529	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTG----------
	Crimson Sweet	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTG----------
C. colocynthis	34256	ATTATATATGTCTATAATTATATATCTATA	ACTAATAATTCTATTCTGTTTTG----------

Table 6. Characterization of *C. lanatus* haplotypes observed at the *G3pdh* intron 2 using *C.colocynthis* as outgroup.

			Deletion at *trn*S-*trn*G	Insertion at *trn*R-*atp*A	Variable substitution sites at intron 2 position within *C. lanatus* var. *citroides*					
Haplotypes	Accessions	Location	690-719	189-212	220-222	203	234	296	436	687
C. lanatus var. *citroides* 1	PI596656	S. Africa	+	-	+	G	A	T	A	T
2	PI532667	Swaziland	+	-	+	G	A	C	G	A
3	TCN1337	U.S.A.	-	-	+	G	A	C	G	T
4	Nam958,Nam1612	Namibia	-	-	+	C	C	C	G	A
5	PI271769 PI189225	Transvaal Zaire	-	-	+	G G	A A	T T	A A	T T
6	PI270563 PI288316 PI485583 Nam1884	S. Africa India Botswana Namibia	- - - -	+ + + +	+	G G G G	A A A A	C C C C	G G G G	T T T T
7	Nam1569	Namibia	-	+	+	G	A	T	A	T
8	PI296343	Cape Province	-	+	+	C	C	T	A	T
9	TCN1126	U.S.A.	-	+	+	G	A	C	G	A
C.lanatus var. *lanatus* *10*	PI494529 Crimson Sweet	Nigeria U.S.A.	+ +	- -	-	G G	A A	C C	G G	T T
11	PI179881	India	+	-	+	G	A	C	G	T

C.colocynthis	34256	Israel	+	-	+	G	A	C	G	T

Table 7. Characterization of variable cpDNA regions and *G3pdh* transit peptide intro 2 section with in *C. lanatus* var. *citroides* and var. *lanatus*.

	*atp*A-*trn*R	*trn*S-*trn*G	*G3pdh* intron 2
Raw Length (bp)	464	672	584
A + T content	85.1%	75.3%	63.1%
Variable characters	6	2	13
Parsimony informative sites	4	4	10
T_s:T_v:ID[2]	1:2:1	0:2:2	4:4:1
Unique indels	23 bp insertion	30 bp deletion	3 bp deletion, 4 bp deletion

T_s=transition, T_v=transversion, ID=indel

Figure 1. Different regions on cpDNA genome

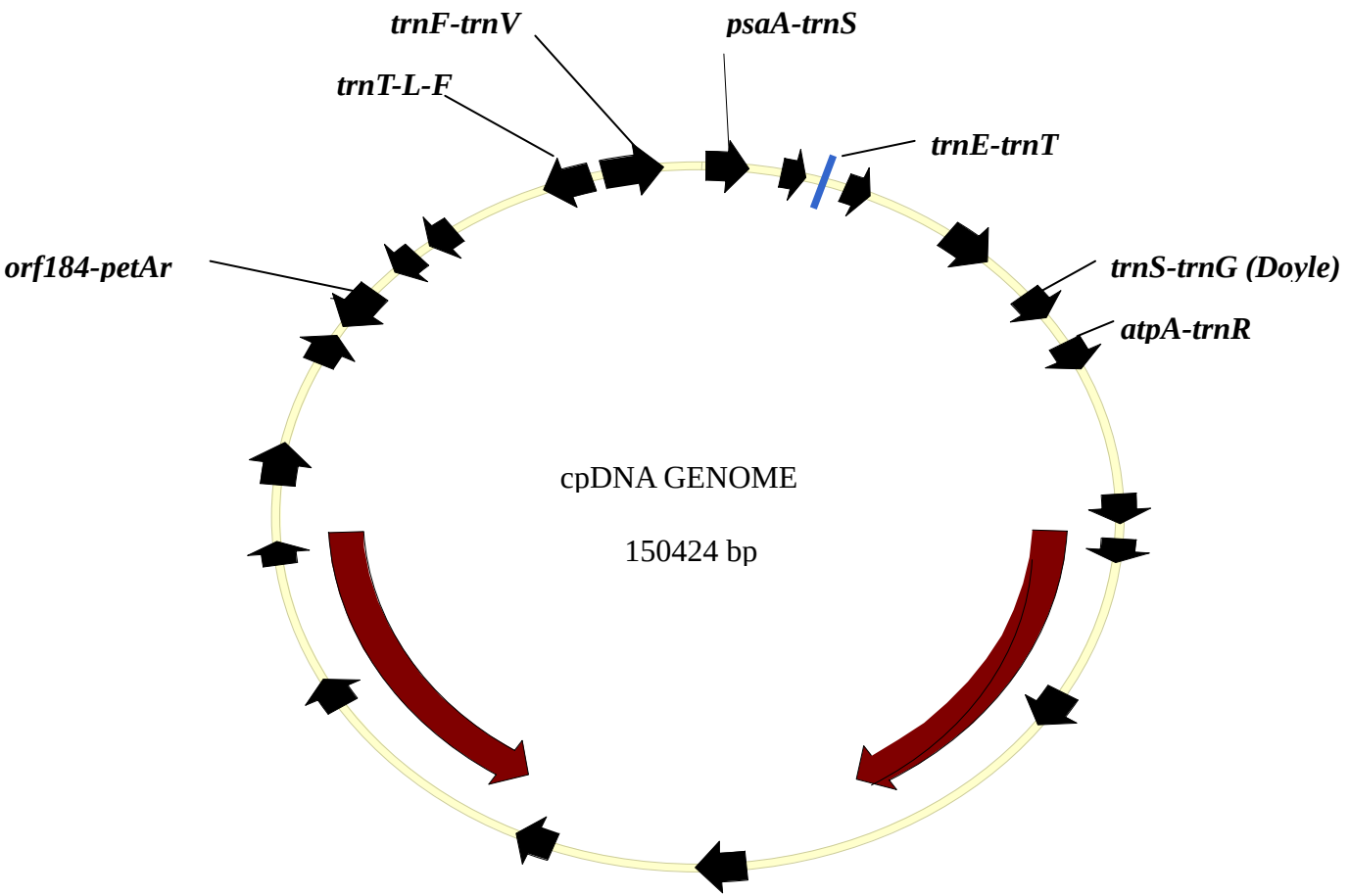
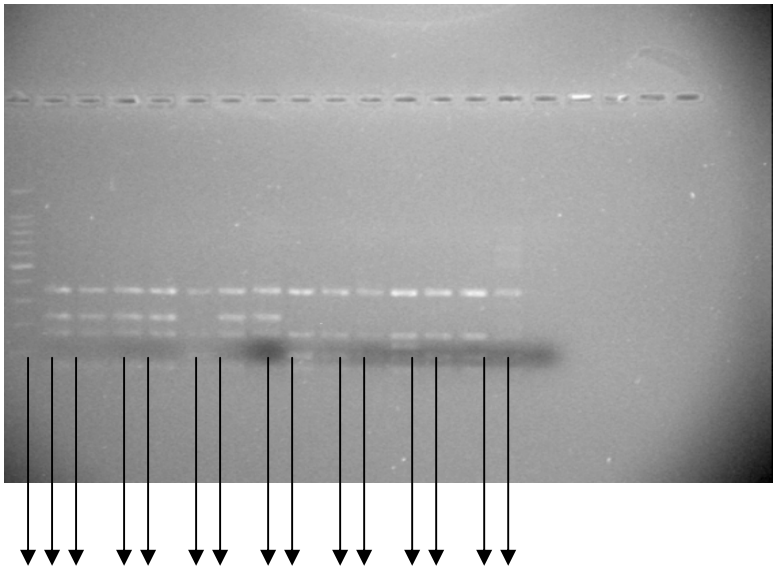
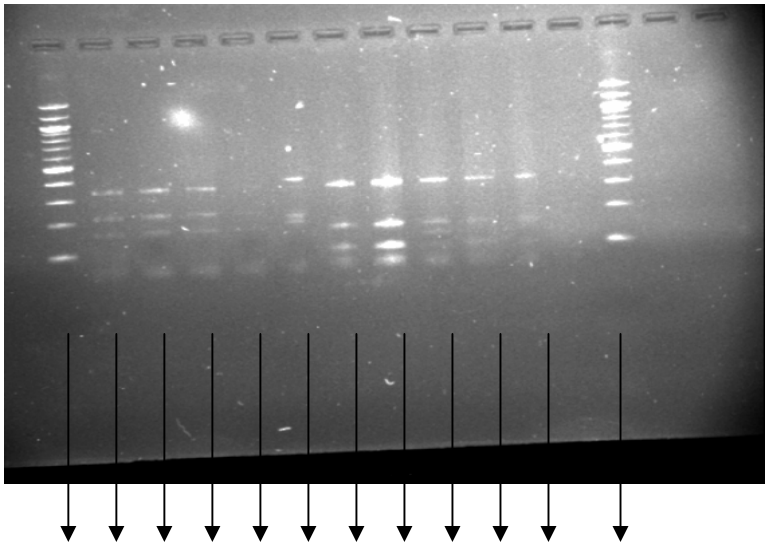


Figure 2. Different banding patterns of some of the surveyed accessions on cpDNA *trn*S-*trn*G region using *Taq*I restriction enzyme.



L 1 2 3 4 5 6 7 8 9 10 1112 13 L

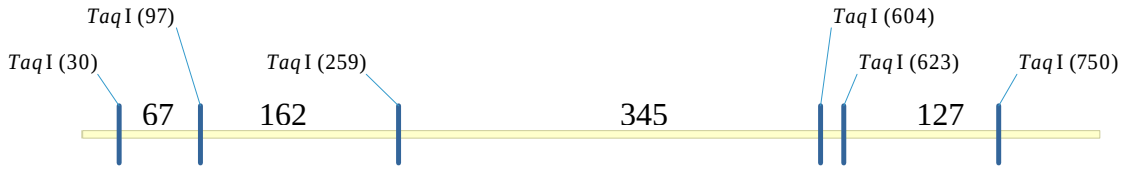
L-Ladder 1-PI295842(I) 2-PI211011(I) 3-PI254742(I)
4-PI494527(I) 5-PI296343(II) 6-PI255136(I)
7-PI482336(I) 8-PI296334(II) 9-TCN1126(II)
10-PI482279(II) 11-PI532667(III) 12-TCN1360(II)
13-PI596671(II)



L 1 2 3 4 5 6 7 8 9 10 L

1-34250(I) 2-34262(I) 3-34256(I)
4-34267(I) 5-TCN1317(I) 6-PI542114(II)
7-PI482331(II) 8-PI596656(III) 9-PI271769(II)
10-PI482315(II)

Figure 3. The *Taq*I site positions on *trn*S-*trn*G noncoding cpDNA region of *C. lanatus* var. *citroides* PI 596656, 532667, and 485583, and *C. lanatus* var. *lanatus* PI 494529.

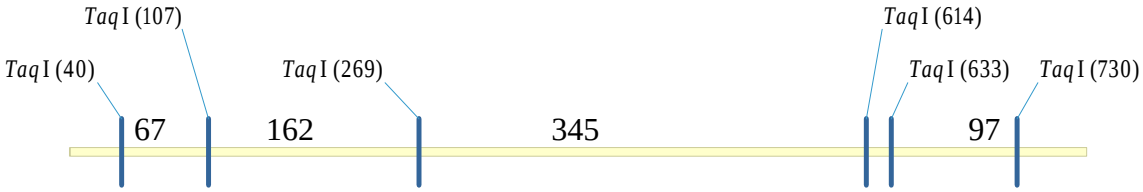


Doyle 596656 & 532667

832 bp

Doyle 494529 RC new

813 bp



Doyle 485583 F RC

783 bp

Figure 4. 50% majority-rule consensus tree of cpDNA sequences data of 16 *C. lanatus* accessions and *C. colocynthis* 34256.

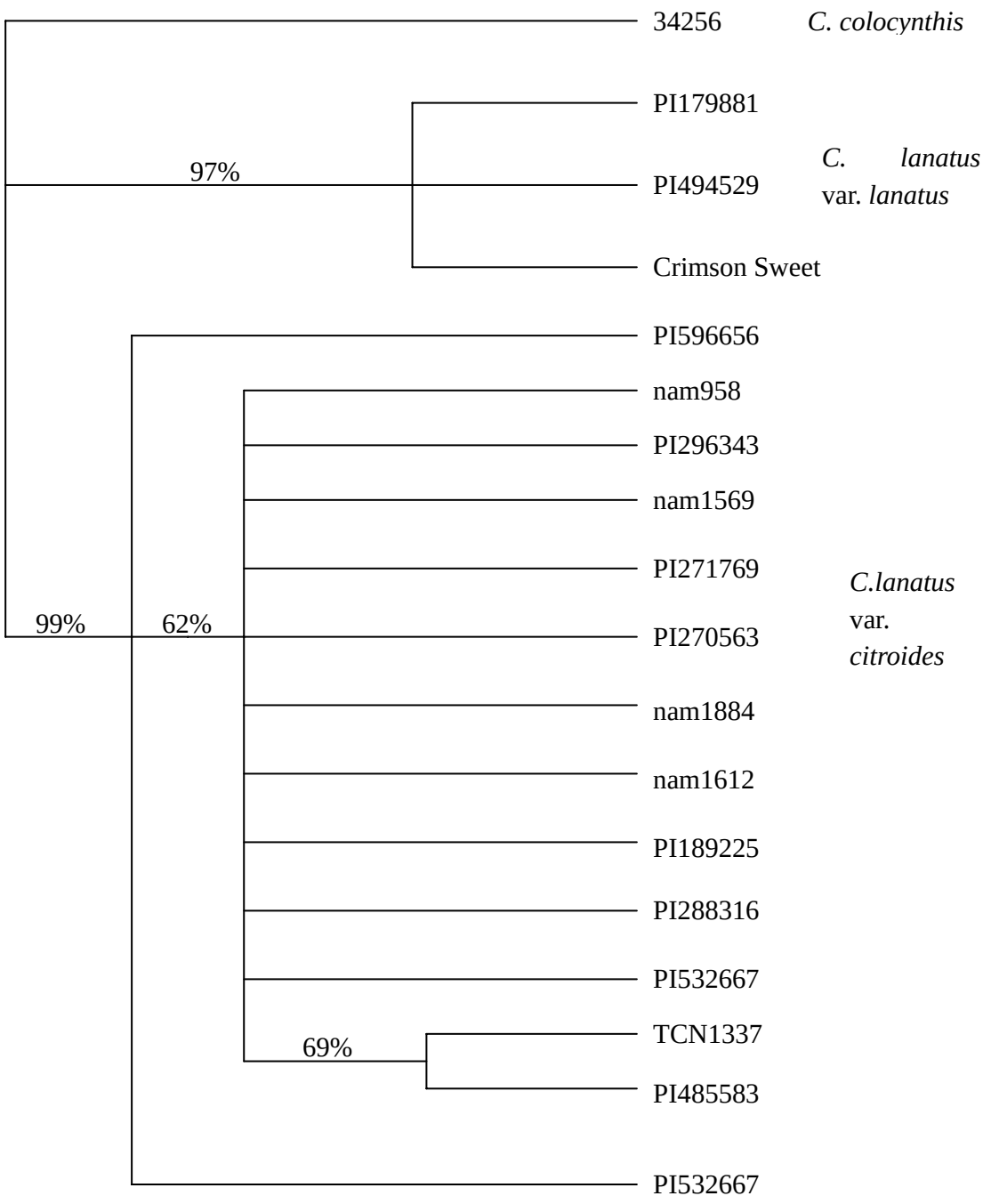


Figure 5. 50% majority-rule consensus tree based on *G3pdh* sequences data of 16 *C. lanatus* accessions and *C. colocynthis* 34256.

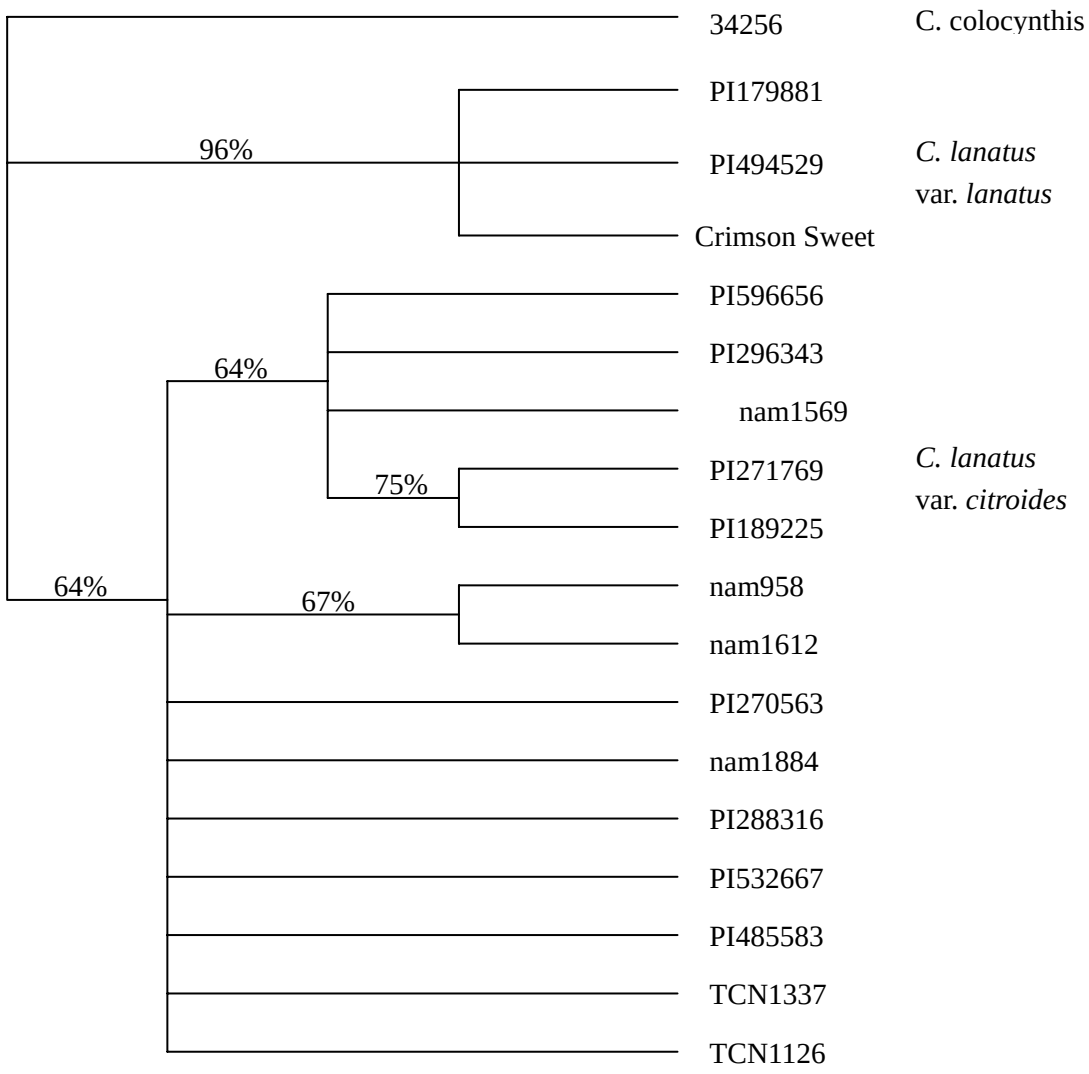


Figure 6. 50% majority-rule consensus tree of cpDNA sequences and *G3pdh* sequences of 16 *C. lanatus* accessions and *C. colocynthis* 34256.

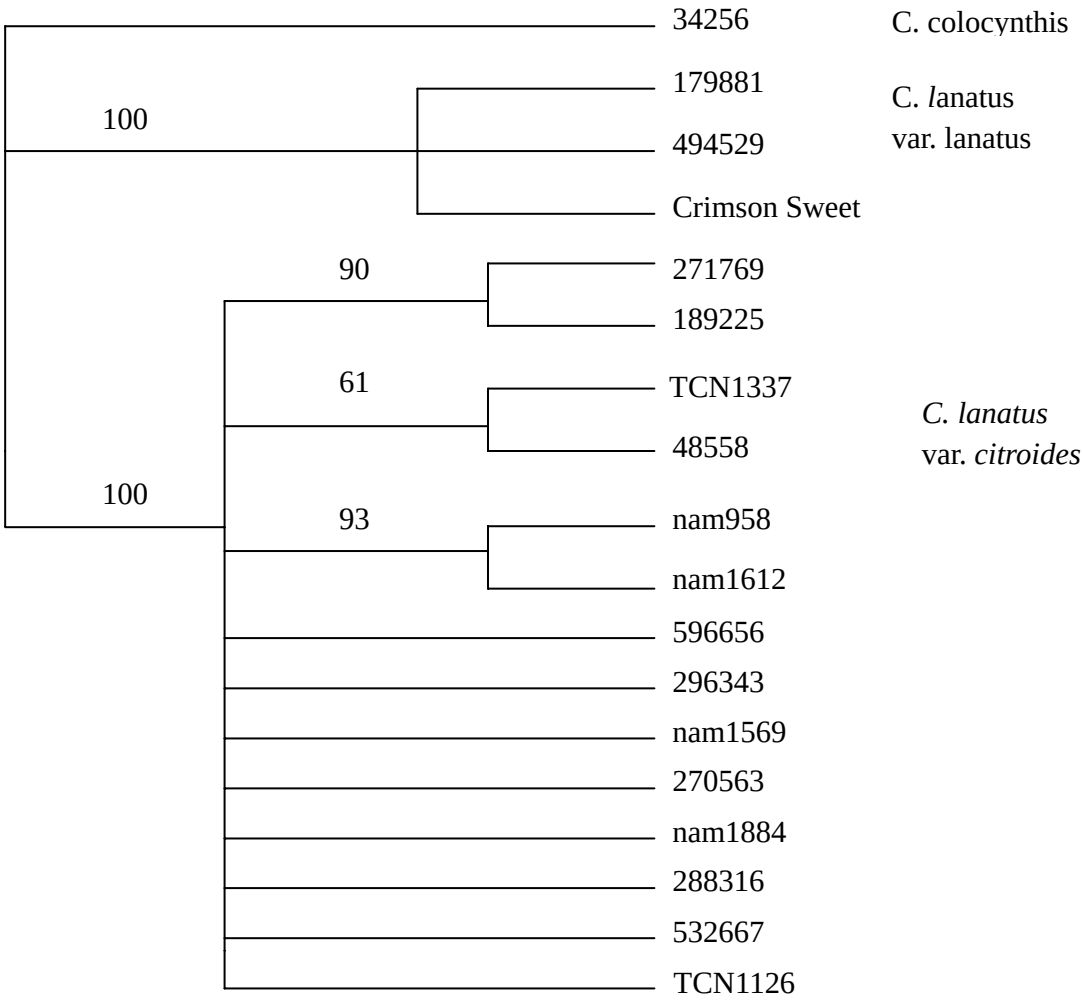


Figure 7. Two distinct lineages with different nuclear profiles among 14 *C. lanatus* var. *citroides* accessions.

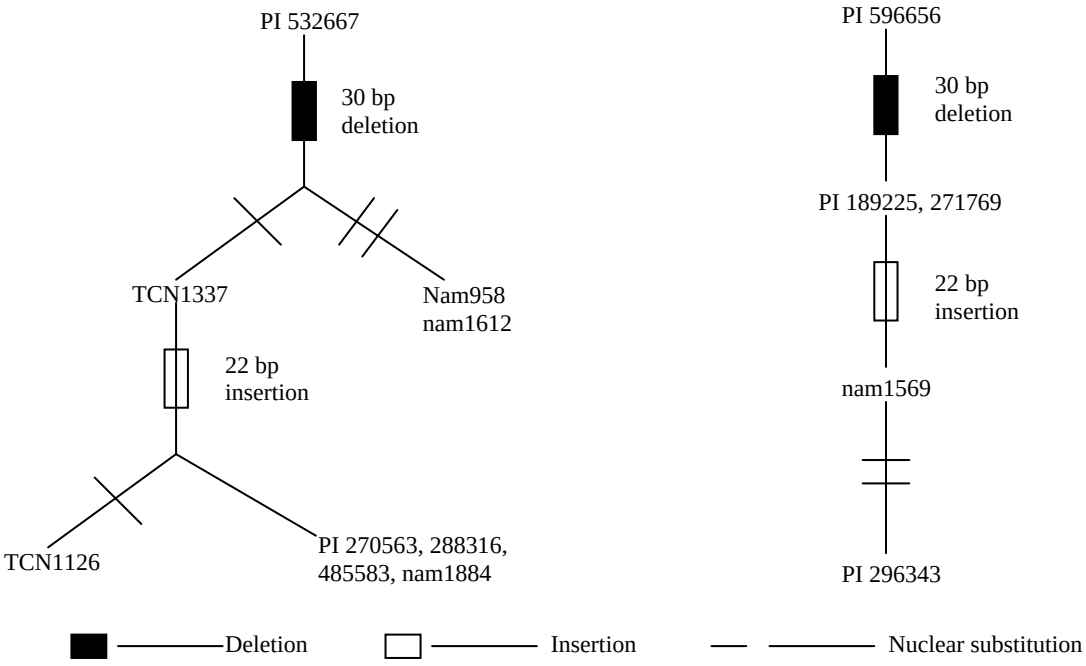
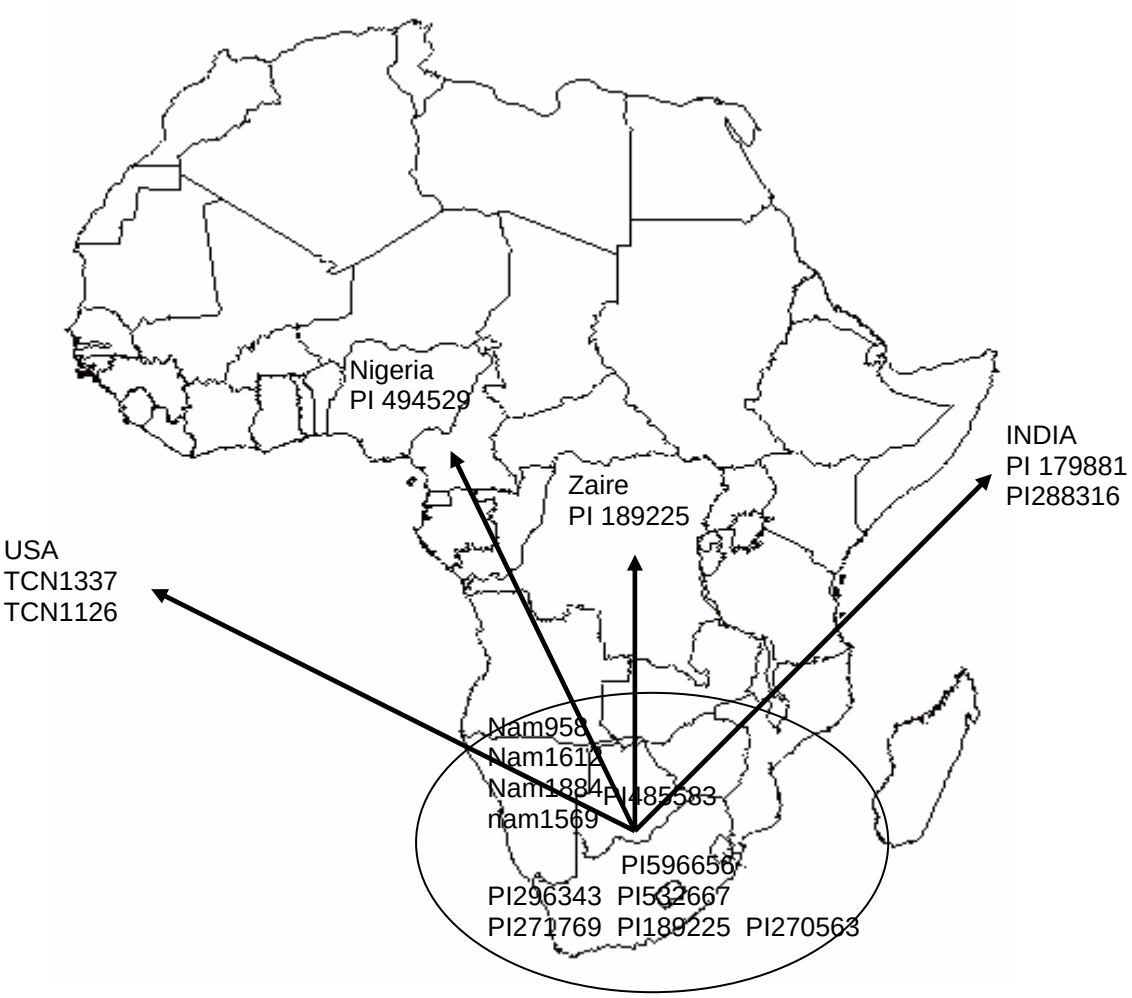


Figure 8. Colonization pattern of 17 *C. lanatus* accessions.



III. PHYLOGENY AND BIOGEOGRAPHY OF WATERMELON [*CITRULLUS LANATUS* (THUNB.) MATSUM. & NAKAI] BASED ON AFLP MARKERS

Introduction

In previous studies, isoenzyme and DNA polymorphisms have been used to conduct genetic identification of watermelon accessions [*Citrullus lanatus* (Thunb.) Matsum&Nakai]. However, the detection efficiency was limited because of the narrow genetic background among *C. lanatus* accessions and cultivars (Zamir et al., 1984; Navot and Zamir, 1987; Biles et al., 1989; Lee at al., 1996; Jarret et al., 1997; Levi et al., 2001). Zamir et al. (1984) found that 12 commercial cultivars of *C. lanatus* were monomorphic for all 19 enzymatic loci examined. Likewise, Biles et al. (1989) detected no polymorphic isoenzyme loci among eight cultivars. Using 14 random RAPD primers to differentiate 39 *C. lanatus var. lanatus* and *C. lanatus var. citroides* accessions, Lee et al. (1996) found that several genotypes were indistinguishable. Levi et al. (2001) found similar results using RAPDs, while, Jarret et al. (1997) reported in their analysis of *C. lanatus* 32 watermelon genotypes using greater SSR markers that several accessions could not be differentiated from each other.

Most recently, the AFLP approach has been used. This technique has the potential to solve such difficulties, particularly among closely related species, or at the intraspecific level (eg., Koopman et al, 2001; Talhinhas et al, 2003; Guo et al., 2005).

AFLP techniques are highly reproducible and provide a large number of informative markers derived from loci dispersed throughout the nuclear genome (Ridout and Donini, 1999). The power of AFLP analysis is tremendously high for revealing genomic polymorphisms. In 2003, Che et al. used eight AFLP primer combinations to assess genetic diversity among 30 genotypes of watermelon [*Citrullus lanatus* (Thunb.) Mansf.] and found each genotype could be successfully distinguished based on AFLP scoring. Likewise, Levi et al (2004) found that AFLP markers identified 97.8% of genetic similarity among heirloom cultivars. AFLPs were highly effective in differentiating watermelon cultivars or elite lines with limited genetic diversity.

The objectives of this study were to a (1) construct a dendrogram based on AFLP markers that delineate phylogenetic relationships among *C. lanatus* accessions and cultivars with low genetic diversity, (2) to compare the cpDNA sequence data with nuclear *G3pdh* sequence data and AFLP data in order to determine which method is more useful for phylogenetic research.

Material and Methods

Plant material and DNA extraction

Accessions used for AFLP analysis were the same as those used for the chloroplast and nuclear genome study (Table 4 in chap II) except for nam1884 because of poor seed quality. DNA was extracted from seeds or young leaf tissue using the Dneasy Plant Maxi kit (Qiagen, Valencia, California 91335, USA) with minor modification. At the final elution step 2x25ul water was added instead of 2x750ul to obtain samples with high concentrations of DNA. DNA concentration was estimated visually using a DNA low mass ladder (Invitrogen) and electrophoresis on a 2% agarose gel.

AFLP analysis

Digestion of DNA

The IRDye^{TM} Fluorescent AFLP Kit for Large Plant Genome Analysis (Li-Cor Biosciences, Lincoln, NE68504, USA) was used. DNA (0.1mg) was digested for 2 hours at 37°C using 1.0 ul EcoR1/Mse1 enzyme mix [1.25 units/ul each in 10 mM Tris-HCL (pH 7.4), 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 200 ug/ml BSA, 50% (v/v) glycerol, 0.15% Triton X-100], 2.5 ul 5X reaction buffer [50mM Tris-HCL (pH 7.5), 50 mM Mg-acetate, 250 mM K-acetate] and deionized water in a final volume of 12.5 ul.

Adaptor ligation

Adaptor ligation was achieved by adding 12 ul adaptor mix [EcoR1/Mse1 adapters, 0.4 mM ATP, 10 mM Tris-HCL (pH 7.5), 10 mM Mg-acetate, 50 mM K-acetate] and 0.5 ul T4 DNA ligase [1 unit/ul in 10 mM Tris-HCL (pH 7.4), 0.1 mM EDTA, 1 mM DTT, 50 mM KCl, 200 ug/ml BSA, 50% (v/v) glycerol] in a final volume of 25 ul. The mixture was incubated at 20°C for 2 hours and diluted (1: 10) with TE buffer.

Pre-amplification

2.5 ul of diluted (1: 10) ligation mixture was mixed with 20 ul AFLP pre-amp primer mix, 2.5 ul 10x PCR reaction buffer [200 mM Tris-HCL (pH 8.4), 15 mM $MgCl_2$, 500 mM KCl] and 0.5 ul Taq DNA polymerase (5 units/ul) in a final volume of 25.5 ul. The PCR program consisted of 20 cycles at 94°C for 30 seconds, 56°C for 1 minute, and 72°C for 1 minute, and finally stay at 4°C. Five ul of the pre-amplification DNA mixture was diluted 1: 20 with deionized water.

Selective PCR amplification

Selective restriction fragment amplification was performed with an unlabeled MseI primer and a IRDye700/800 labeled EcoRIprimer. Five different primer combinations were used (Table 1). Each 11 ul PCR reaction consisted of 2 ul diluted (1:20) pre-amplified DNA, 2 ul Mseⅰprimer containing dNTPs, 0.5 ul IRDye700/800 labeled EcoRⅠprimer and 6 ul Taq DNA polymerase working mix [148 ul deionized water, 40 ul 10x amplification buffer, 2 ul Taq DNA polymerase (5 units/ul) (for 33 reactions)]. The following cycle profile ensured optimal primer selectivity: 1 cycle of 94°C for 30 seconds, 65°C for 30 seconds, and 72°C for 1 minute followed by 12 cycles of subsequently lowering the annealing temperature (65°C) by 0.7°C per cycle while keeping the denature temperature at 94°C for 30 seconds and extension step at 72°C for 1 minute, followed by 23 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute.

Gel electrophoresis

The PCR product was mixed with 3 ul of Blue Stop Solution and denatured for 3 minutes at 94°C and placed on ice immediately. A 25-cm LI-COR glass plate, 0.2-mm thick spacers and 64-tooth rectangular comb were used. The long ranger polyacrylamide (8%) gel was prepared by mixing 9.4-9.5 g urea, 2.7 ml 10xTBE (80mM Tris base, 40mM Boric acid, 2mM EDTA), 3.6 ml Long Ranger solution

(Cambrex Bio Science, Rockland, ME, USA) and 7.5 ml deionized water and adding 150 ul 10% APS and 15 ul TEMED. The gel was poured at least 3 hours before use and pre-run at 40W for 30 minutes. Approximately 0.8 to 1.0 ul of each denatured sample was loaded per lane using a 8-channel Hamilton syringe and gels were run for about 5 hours. Real-time IRDye labeled AFLP data (TIF images) were automatically collected and recorded during electrophoresis.

Data analysis

AFLP fragments were visually scored as present (1) or absent (0) and weak bands were removed from the matrix before analysis. Fragment scoring was imported in NTSYSpc version 2.1 (Exeter Software, Setauket, New York, USA) as a (1, 0) binary data matrix. Data from five primer combinations were combined in one data set. Clustering methods and similarity coefficients were tested using the procedures Similarity (Qualitative Distance), Clustering [unweighted pair group method of the arithmetic average (UPGMA)], and Graphics (TREE Plot) from the program NTSYSpc version 2.1. Dice's (1979) genetic distance was used for cluster analysis. Species relationship among the *C. lanatus* accessions were studied using principal coordinate analysis (PCA) in SAS Version 8.0e (SAS Institute Inc, Cary, NC, USA).

Results

AFLP fingerprinting of 17 taxa of *C. lanatus* with five different primer combinations (Table 1) revealed a total number of 1094 clearly identifiable bands (with an average of 218.8 band per primer), of which 533 (48.72%) were polymorphic. Concomitantly, high values of genetic similarity were obtained, ranging from 0.6825 between *C. lanatus* var. *lanatus* 'Crimson Sweet' and *C. lanatus* var. *citroides* TCN1126 to 0.9635 between *C. lanatus* var. *lanatus* 'Crimson Sweet' and PI179881. Within *C. lanatus* var. *citroides*, the lowest similarity value of 0.7719 between

PI485583 and TCN1126 and the highest value of 0.9629 between PI296343 and TCN1337/PI596656 were obtained. Overall similarity values are presented in Table 2.

The UPGMA clustering of 17 taxa based on Dice's coefficient (1979) gave more information on grouping of different accessions (Figure 1). UPGMA clustering yields the highest cophenetic correlation (Koopman et al, 2001) and is therefore considered suitable for determining phenetic relationships in *C. lanatus*. Two major clusters were obtained at a distance of 0.29. The first cluster included all 3 *C. lanatus* var. *lanatus* accessions and PI532667, which were separated at a distance of 0.14. In the other main cluster, all except TCN1126, PI 189225 and 485583 were closely clustered together.

Principal coordinate analysis

The five first principal coordinates accounted for 73.24% of the total variance in the data matrix (Table 3). Only the first three principal coordinates (explaining 57.27% of the total explained variance) were retained, to establish spatial representations of the individuals. The first principal coordinate (describing 33.50% of the total variation) and the second (describing 13.87% of the total variation) separated two groups. Group 1 contains all 3 *C. lanatus* var. *lanatus* accessions. Group 2 includes PI189225, PI270563, PI596656, PI271769, PI 288316, PI296343, nam958, nam1612, nam1569, TCN1126 and TCN1337 (Figure 2). These two groups were identical to the two major cluster in the UPMGA clustering analysis, except for PI532667 which was not in group 1 and PI485583 and outgroup 34256 which were not in group 2. Principal coordinate 3 (describing 9.9% of the variation) separated PI189225 from Group 2 to cluster it with PI485583 and outgroup *C. colocynthis* 34256 (Figure 3).

Discussion

The results of the AFLP study demonstrate that the intra-genetic diversity between *C. lanatus* var. *lanatus* and *C. lanatus* var. *citroides* is consistent with that observed using cpDNA and nuclear region *G3pdh* sequences. AFLP analysis supports cpDNA and *G3pdh* sequence analysis in that two main clusters were detected, one of *C. lanatus* var. *lanatus* and the other of *C. lanatus* var. *citroides*. PI 532667 has very similar AFLP banding pattern as *C. lanatus* var. *lanatus* accessions (Figure 4), as a result, in the UPGMA clustering analysis, PI532667 was grouped in the same cluster as the 3 *C. lanatus* var. *lanatus* accessions, which supports the conclusion that PI532667 and *C. lanatus* var. *lanatus* may have evolved from the same ancestor. While 'Crimson Sweet' and PI494529 have exactly the same cpDNA and *G3pdh* haplotype, AFLP analysis indicated a closer relationship between PI179881 and 'Crimson Sweet'. AFLP analysis grouped PI596656 and PI296343 together, consistent with result from sequence analysis. TCN1337 and PI270563 have the highest genetic similarity, which is also supported by their identical *G3pdh* sequences. In the AFLP UPGMA dendrogram, PI485583, PI189225 and TCN1126 have relatively low values of genetic similarity. In the PCA, PI485583 and PI189225 were separated from other *C. lanatus* var. *citroides* accessions and grouped together with *C. colocynthis* 34256; TCN 1126 was also separated from all the other 16 taxa. The relatively greater genetic diversity among these three accessions may be due to recombination during reproduction. TCN1337 and TCN1126, both from USA, were possibly obtained from sources in southern Africa.

Although AFLP is a very powerful molecular marker technique to distinguish intra-/inter-species relationships (Amsellem et al, 2000; Anderberg et al, 2002; Cervera et al, 2002; Papa and Gepts, 2003), since it did provide more and accurate differentiation among *C. lanatus* compared to sequences data analysis, the genetic

diversity is limited among *C. lanatus* var. *citroides* and *C. lanatus* var. *lanatus*. Levi et al. (2004), also found the genetic diversity to be limited among watermelon genotypes. Long term domestication of watermelon and human selection and inbreeding may be the major reason to cause this genetic bottleneck and loss of genetic diversity.

In conclusion, *G3pdh* nuclear sequences and AFLP nuclear marker analysis were probably the most informative at detecting differences among *C. lanatus* accessions in this study. As far as lineage and phytogeography are concerned, however, the application of *G3pdh* nuclear sequences and AFLP nuclear marker might have been obscured by recombination and cpDNA sequences analysis was most reliable and more informative.

Table 1. AFLP primer combinations, primer sequences, total number of bands generated by each primer set, number of polymorphic bands detected, and percentages of polymorphic bands used in the study of *C. lanatus*

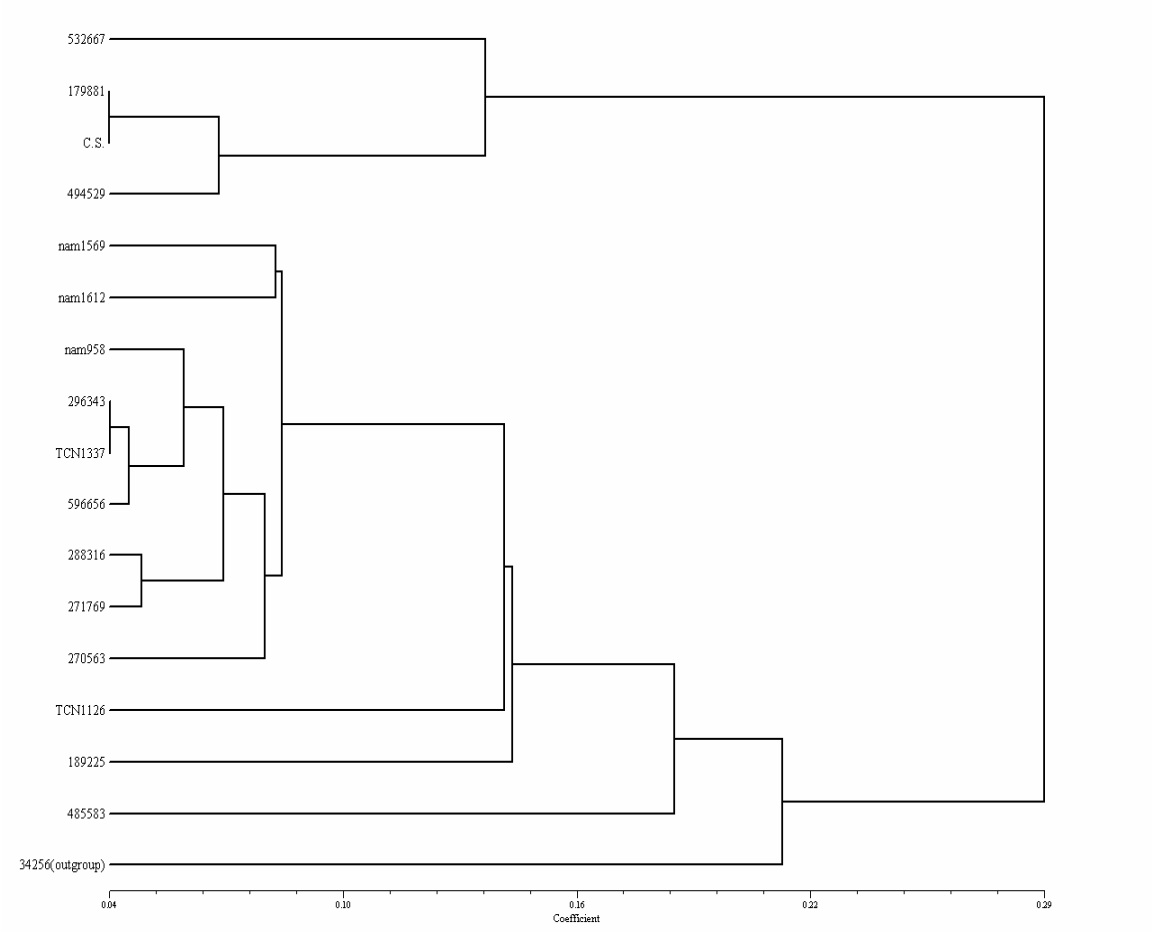
Primer	Total no. of bands	No. of polymorphic bands	% of polymorphic bands
IRD700E-AAC + M-CAC	232	79	34.05%
IRD700E-ACC + M-CAA	268	115	42.91%
IRD800E-ACT + M-CTA	202	83	41.09%
IRD800E-AGG + M-CAG	203	152	74.88%
IRD800E-AGC + M-CAC	189	104	55.03%
Total	1094	533	48.72%
Average	218.8	106.6	48.72%

Table 2 Similarity matrix calculated with Dice coefficient for the 17 *C. lanatus* taxa from banding patterns with AFLP

Rows/ Cols	PI532667	Nam1569	Nam958	TCN1126	PI189225	PI288316	PI271769	PI485583	Nam1612	PI296343	TCN1337	PI270563	PI596656	34256	PI179881	PI494529	Crimson Sweet
PI532667	1																
Nam1569	0.8154	1															
Nam958	0.8529	0.9118	1														
TCN1126	0.7742	0.8387	0.8923	1													
PI189225	0.7879	0.8485	0.8551	0.8571	1												
PI288316	0.8209	0.9104	0.9429	0.8594	0.8529	1											
PI271769	0.7969	0.9531	0.9104	0.8689	0.8769	0.9545	1										
PI485583	0.8000	0.8500	0.8095	0.7719	0.8361	0.8065	0.8475	1									
Nam1612	0.7813	0.9219	0.8955	0.8689	0.8769	0.9091	0.9365	0.8644	1								
PI296343	0.8296	0.9185	0.9503	0.8837	0.8613	0.9352	0.9173	0.8160	0.9323	1							
TCN1337	0.8154	0.9231	0.9412	0.8710	0.8788	0.9403	0.9375	0.8333	0.9219	0.9630	1						
PI270563	0.7874	0.8976	0.9023	0.8430	0.8837	0.9160	0.9120	0.8205	0.9120	0.9242	0.9606	1					
PI596656	0.8308	0.9231	0.9412	0.8710	0.8636	0.9552	0.9375	0.8333	0.9375	0.9630	0.9538	0.9291	1				
34256	0.7500	0.8000	0.7937	0.7544	0.7869	0.8065	0.7966	0.7818	0.7966	0.8000	0.8333	0.8547	0.8333	1			
PI179881	0.9037	0.7556	0.7801	0.6977	0.7591	0.7626	0.7519	0.7680	0.7368	0.7714	0.7556	0.7273	0.7704	0.7200	1		
PI494529	0.8480	0.7360	0.7328	0.6891	0.7244	0.7132	0.7317	0.7826	0.7480	0.7231	0.7200	0.6885	0.7200	0.6957	0.9231	1	
Crimson Sweet	0.8636	0.7424	0.7391	0.6825	0.7463	0.7206	0.7385	0.7705	0.7384	0.7299	0.7273	0.6977	0.7273	0.7213	0.9635	0.9449	1

Table 3. Five principal coordinates of the principal coordinate analysis (PCA) and their respective contributions to the total variance.

Principal coordinates	Contributions to the original variance
1	33.50%
2	13.87%
3	9.99%
4	8.71%
5	7.26%
Total	73.24%



Dice's distance

Figure 1. UPGMA dendrogram of AFLP marker based on Dice's distance (1979) matrix of 17 accessions.

Figure 2. Principal coordinate analysis of AFLP marker data from 17 *C. lanatus* accessions and *C. colocynthis* 34256. Principal coordinate 1 and 2.

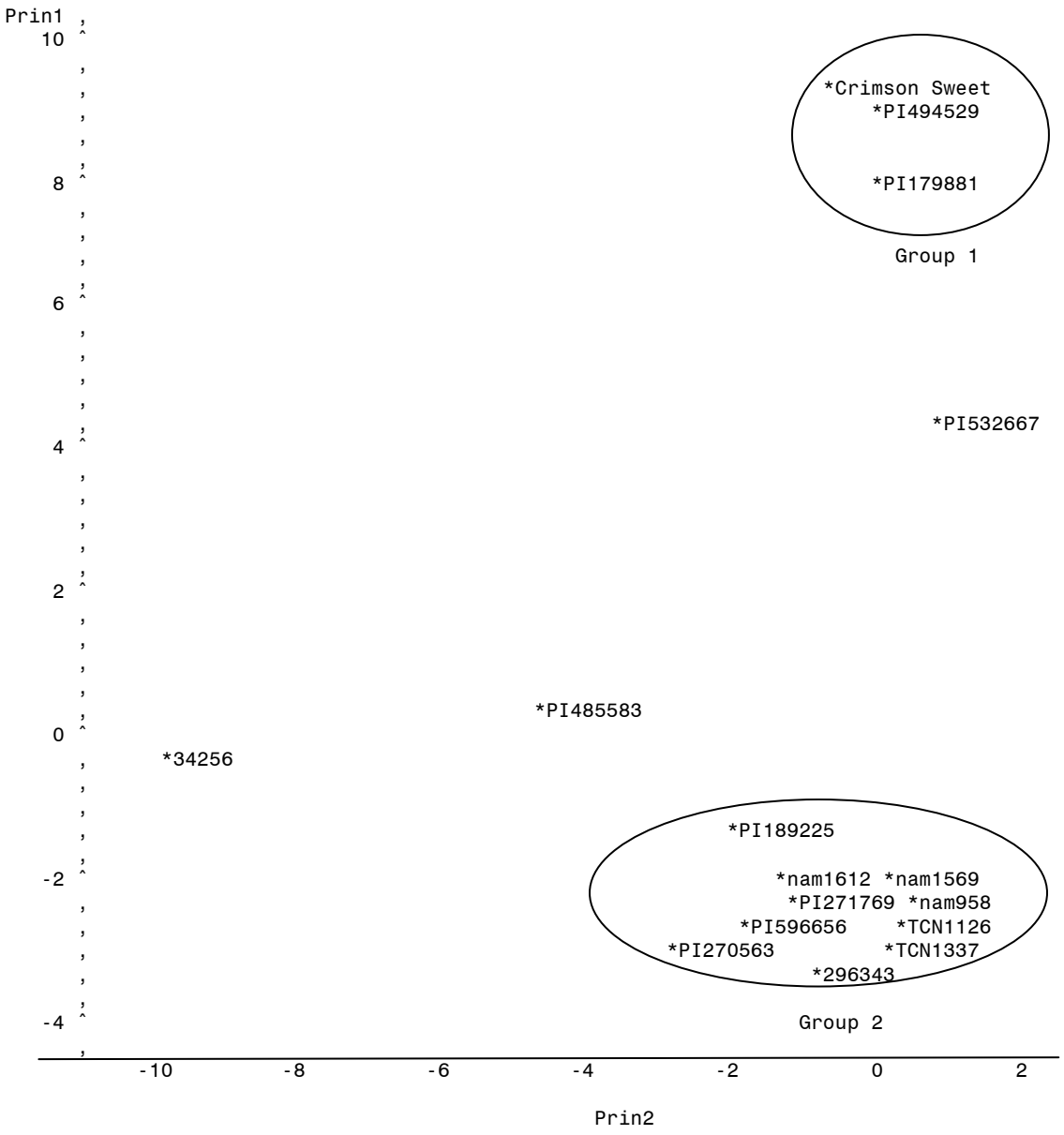
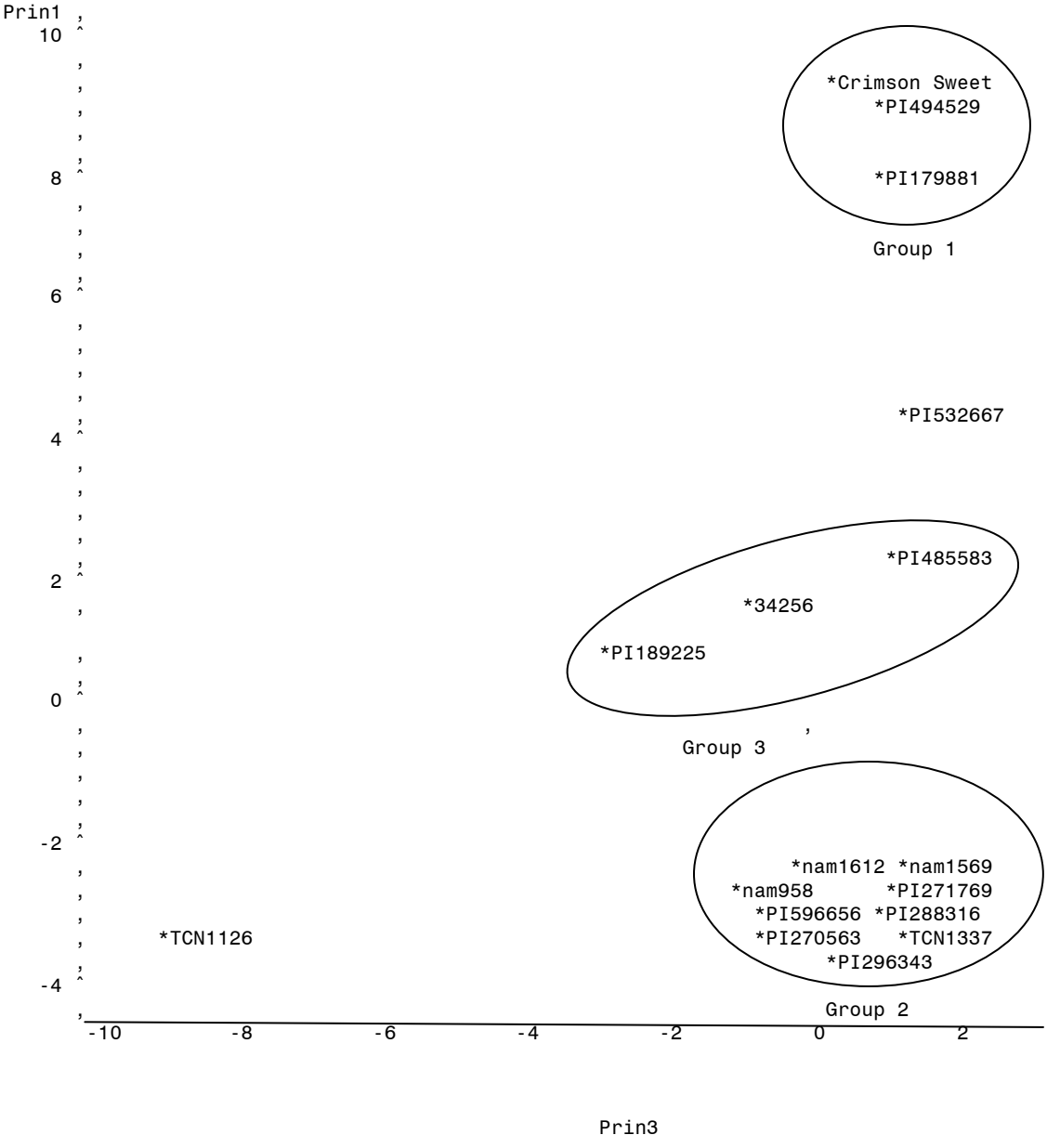


Figure 3. Principal coordinate analysis of AFLP marker data from17 *C. lanatus* accessions. Principal coordinate 1 and 3.



IV LITERATURE CITED

ALSOS, I. G., T. ENGELSKJON, L. GIELLY, P. TABERLET, AND C. BROCHMANN. 2005. Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology* In press.

Allen, J. F. 2003. State transitions – a question of balance. *Science* 299: 1530-1532.

ANDERBERG, A. A., C. RYDIN, AND M. KALLERSJO. 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany* 89: 677-687.

AVISE, J. C. 1994. Molecular markers, Natural history, and evolution. Chapman & Hall, New York, NY, USA.

AVISE, J. C. 2001. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.

BAILEY, C. D., AND J. J. DOYLE. 1999. Potential phylogenetic utility of the low-copy nuclear gene pistillata in dicotyledonous plants: comparison to nrDNA ITS and *trn*L intron in *Sphaerocardamum* and other Brassicaceae. *Molecular and Phylogenetic Evolution* 13: 20-30.

BAILEY, L. H., AND E. Z. BAILEY. 1941. A concise dictionary of gardening, general horticulture and cultivated plants in North America (Hortus second). The Macmillian Company, New York, NY, USA.

BAKER, W. J., T. A. HEDDERSON, AND J. DRANSFIELD. 2000. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rps*16 intron sequence data. *Molecular phylogenetics and evolution* 14: 195-217.

BILES, C. L., R. D. MARTYN, AND H. D. WILSON. 1989. Isoenzymes and general proteins from various watermelon cultivars and tissue types. *HortScience* 24: 810-812.

CERVERA, M. T., L. RUIZ-GARCIA, AND J. M. MARTINEZ-ZAPATER. 2002. Analysis of DNA metylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP Markers. *Molecular & Genomics* 268: 543-552.

CHAW, S.-M., A. ZHARKIKH, H.-M. SUNG, T.-C. LAU, AND W.-H. LI. 1997. Molecular phylogeny of gymnosperms and seed plant evolution: analysis of 18S rRNA sequences. *Journal of molecular and evolution* 14: 56-58.

CHE, K.-P., C.-Y. LIANG, D.-M. J. WANG, AND B. WANG. 2003. Genetic assessment of watermelon germplasm using the AFLP technique. *HortScience* 38: 81-84.

CHUNG, S.-M., D. S. DECKER-WALTERS, AND J. E. STAUB. 2003. Genetic relationships within the Cucurbitaceae as assessed by consensus chloroplast simple sequence repeats (ccSSR) marker and sequence analyses. *Canadian Journal of Botany* 81: 814-832.

CLEGG, M. T. 1993. Chloroplast gene sequenced and the study of plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* 90: 363-367.

CLEGG, M. T., M. P. CUMMINGS, AND M. L. DURBIN. 1997. The evolution of plant nuclear genes. *Proceedings of the National Academy of Sciences of the United States of America* 94: 7791-7798.

DANE, F. 2002. Chloroplast DNA in investigation in *Citrullus* using PCR-RFLP analysis. Cucurbitaceae. ASHS Press, Naples, Fla, USA.

DANE, F., AND P. LANG. 2004. Sequence variation at cpDNA regions of watermelon and related species: implications for the evolution of *Citrullus* haplotypes. *American Journal of Botany* 91: 1922-1929.

DANE, F., P. LANG, AND R. BAKHTIYAROVA. 2004. Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species. *Theoretical and Applied Genetics* 108: 958-966.

DEMESURE, B., N. SODZI, AND R. J. PETIT. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129-131.

DENO, H., A. KATO, K. SHINOZAKI, AND M. SUGIURA. 1982. Nucleotide sequences of tobacco chloroplast genes for elongator tRNAMet and tRNAVal (UAC): the tRNAVal (UAC) gene contains a long intron. *Nucleic Acids Research* 10: 7511-7520.

DESPRES, L., L. GIELLY, B. REDOUTET, B. REDOUTET, AND P. TABERLET. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution* 27: 185-196.

DE WINTER, B. 1990. A new species of *Citrullus* (Benincaseae) from the Namib Desert, Namibia. *Bothalia* 20: 209-213.

DOYLE, J. J., J. I. DAVIS, R. J. SORENG, D. GARVIN, AND M. J. ANDERSON. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences of the United States of America* 89: 7722-7726.

DUMOLIN-LAPEGUE, S., M.-H. PEMONGE, AND R. J. PETIT. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* 6: 393-398.

EMSHWILLER, E., AND J. J. DOYLE. 1999. Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic studies with an example from oxalis (Oxalidaceae). *Molecular phylogenetics and evolution* 12: 310-319.

FAN, C., AND Q.-Y. XIANG. 2001. Phylogenetic relationships within *Cornus* (Cornaceae) based on 26S rDNA sequences. *American Journal of Botany* 88: 1131-1138.

FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, AND C. BULT. 1995. Constructing a significance test for incongruence. *Systematic Biology* 44: 570-572.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

FELSENSTEIN, J. 1993. http://evolution.genetics.washington.edu/phylip.html.

FISHBEIN, M., C. HIBSCH-JETTER, D. E. SOLTIS, AND L. HUFFORD. 2001. Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. *Systematic Biology.* 50: 817-847.

GEPTS, P. 1998. Ten thousand years of crop evolution. Plants, genes, and crop biotechnology. Jones and Barlett, Sudbury, Massachusetts, USA.

GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS, AND R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Science* 161 (6 Suppl.): s83-s96.

GRIVET, D., B. HEINZE, G. G. VENDRAMIN, AND R. J. PETIT. 2001. Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. *Molecular Ecology Notes* 1: 345-349.

GROB., G. B. J., B. GRAVENDEEL, AND M. C. M. EURLINGS. 2004. Potential phylogentic utility of the nuclear FLORICAULA/LEAFY second intron: comparison with three chloroplast DNA regions in Amorphophallus (Araceae). *Molecular phylogenetics and evolution* 30: 13-23.

GUO, Y.-P., J. SAUKEL, R. MITTERMAYR, AND F. EHRENDORFER. 2005. AFLP analyses

demonstrate genetic divergence, hybridization, and multiple polyploidization in the evolution of *Achillea* (Asteraceae-Anthemiddeae). *New Phytologist* 166: 273-290.

HAMBY, K. R., AND E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Chapman & Hall, New York, NY, USA.

HARRIS, A. S., AND R. INGRAM. 1991. Chloroplas DNA and biosystematics: The effects of intraspecific diversity and plastid transmission. *Taxon* 40: 393-412.

HARRIS, D. R. 1996. The origins and spread of agriculture and pastoralism in Eurasia. Smithsonian Institution Press, Washington, D.C., USA.

HARVEY, M. J. 1990. Phylogenetic relationships among cultivated *Allium* species from restriction enzyme analysis of the chloroplast genome. *Theoretical & Applied Genetics* 81: 752-757.

HARVEY, M. J., J. D. MCCREIGHT, B. RHODES, AND G. TAURICK. 1998. Differential expression of the Cucumis organellar genomes. *Theoretical & Applied Genetics* 97: 122-128.

HASEBE, M., P. G. WOLF, K. M. PRYER, M. UEDA, R. ITO, G. J. SANO, J. GASTONY, R. J. YOKOYAMA, N. MANHART, E. H. MURAKAMI, E. H. CRANE, C. H. HAUFLER, AND W. D. HAUK. 1995. Fern phylogeny based on *rbc*L nucleotide sequences. *American Fern Journal* 85: 134-181.

HASHIZUME, T., I. SHIMAMOTO, AND M. HIRAI. 2003. Construction of a linkage map and QTL analysis of horticultural traits for watermelon [*Citrullus lanatus* (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. *Theoretical & Applied Genetics* 106: 779-785.

HASHIZUME, T., I. SHIMAMOTO, M. YUI, T. SATO, T. LMAI, AND M. HIRAI. 1996. Constriction of a linkage map for watermelon (*Citrullus lanatus*) using random amplified polymorphic DNA (RAPD). *Euphytica* 90: 265-273.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C. SUN, B. MENG, Y. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, AND M. SUGIURA. 1989. The complete sequence of rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular & General Genetics.* 217: 185-194.

HODKINSON, T. R., S. A. RENVOIZE, G. N. CHONGHAILE, C. M. A. STAPLETON, AND M. W. CHASE. 2000. A comparison of ITS nuclear rDNA sequence data and AFLP marker for phlogenetic studies in phyllostachys (Bambusoideae, Poaceae). *Journal of Plant Research* 113: 259-269.

HOPKINS, M. S., A. M. CASA, T. WANG, S. E. MITCHELL, R. E. DEAN, G. D. KOCHERT, AND S. KRESOVICH. 1999. Discovery and characterization of polymorphic simple

sequence repeats (SSRs) in peanut. *Crop science* 39: 1237-1242.

HUPFER, H., M. SWIATEK, S. HORNUNG, H. R.G., M. R.M., W. L. CHIU, AND B. SEARS. 2000. Complete nucleotide sequence of the Oenothera elata plastid chromosome, representing plastome I of the five distinguishable Euoenothera plastomes. *Molecular & General Genetics.* 263: 581-585.

JARRET, R. L., L. C. MERRICK, T. HOLMS, J. EVANS, AND M. K. ARADHYA. 1997. Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome* 40: 433-441.

JANSEN, R. C., J. L. WEE, AND D. MILLIE. 1998. Comparative utility of chloroplast DNA restriction site and DNA sequence data for phylogenetic studies in plants. Kluwer Academic, Dordrecht, Netherlands.

JEFFREY, C. 1990. Systematics of the Cucurbitaceae: an overview. Cornell University Press, Ithaca, New York, USA.

JEFFREY, R.-I. 2005. Quantitative trait loci and the study of plant dometication. *Genetica* 123: 197-204.

KATO, T., T. KANEKO, S. SATO, Y. NAKAMURA, AND S. TABATA. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Research* 7: 323-330.

Klaus, D., H. Härtel, L. Fitzpatrick, J. F. Froehlich, J. Hubert, C. Benning, P. Dörmann. 2002. Digalactosyldiacylglycerol synthesis in chloroplasts of the *Arabidopsis thaliana dgd*1 mutant. *Plant Physiology* 128: 885–895.

KOOPMAN, W. J. M., M. J. ZEVENBERGEN, AND R. G. VAN DEN BERG. 2001. Species relationships in *Lactuca* S.L. (Lactuceae, Asteraceae)inferred from AFLP fingerprints. *American Journal of Botany* 88: 1881-1887.

KUBO, T., S. NISHIZAWA, A. SUGAWARA, N. ITCHODA, A. ESTIATI, AND T. MIKAMI. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys(GCA). *Nucleic Acids Research* 28: 2571-2576.

KUZOFF, R. K., J. A. SWEERE, D. E. SOLTIS, P. S. SOLTIS, AND E. A. ZIMMER. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Molecular Biology and. Evolution.* 15: 251-263.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 11-20.

KUMAR, S., K. TAMURA, M. NEI, AND I. B. JACOBSEN. 1993. MEGA2: Molecular

evolutionary genetic analysis software. Arizona State University, Tempe, Arizona, USA.

LEE, S. J., J. S. SHIN, K. W. PARK, AND Y. P. HONG. 1996. Detection of genetic diversity using RAPD-PCR and sugar analysis in watermelon (C*itrullus lanatus* (Thunb.) Mansf.) germplasm. *Theoretical & Applied Genetics* 92: 719-725.

LEVI, A., C. E. THOMAS, T. C. WEHNER, AND X. P. ZHANG. 2001. Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon. *HortScience* 36: 1096-1101.

LEVI, A., C. THOMAS, M. NEWMAN, O. U. K. REDDY, AND X. ZHANG. 2004. ISSR and AFLP markers sufficiently differentiated among American watermelon cultivars with Limited Genetic Diversity. *Journal American Society Hortscience* 129: 553-558.

LEWIS, C. E., AND J. J. DOYLE. 2001. Phylogenetic utility of the nuclear gene malate synthase in the palm family (Arecaceae). *Molecular phylogenetics and evolution* 19: 409-420.

LI, W.-H., AND D. GRAUR. 2000. Fundamentals of molecular evolution. Sinauer, Sunderland, Massachusetts, USA.

LISTON, A., W. A. ROBINSON, J. M. OLIPHANT, AND E. R. ALVAREZ-BUYLLA. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. *Systematic Botany* 21: 109-120.

LIU, Z. J., AND J. F. CORDES. 2004. DNA marker technologies and their applications in aquaculture genetics (review). *Aquaculture* 238: 1-37.

MACKILL, D. J., Z. ZHANG, E. D. REDONA, AND P. M. COLOWIT. 1996. Level of polymorphism and genetic mapping of AFLP markers in rice. G*enome* 39: 969-977.

MAIER, R. M., K. NECKERMANN, G. L. IGLOI, AND H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *Journal of Molecular Biology* 251: 614-628.

MARINO, M. A., L. A. TURNI, S. A. DEL RIO, P. E. WILLIAMS, AND P. B. CREGAN. 1995. The analysis of simple sequence repeat DNA in soybean by capillary gel electrophoresis. *Applied and Theoretical Electrophoresis* 5: 1-5.

MARTIN, W., AND F. SALAMINI. 2000. A meeting at the gene. Biodiversity and natural history. *EMBO Journal* 1: 208-210.

MASON-GAMER, R. J., AND E. A. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology*

45: 524-545.

MASON-GAMER, R. J., C. F. WEIL, AND E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molacular Biology and Evolution* 15: 1658-1673.

MASOUD, S. A., L. B. JONSHON, AND E. L. SORENSEN. 1990. High transmission of paternal plastid DNA in alfalfa plants demonstrated by restriction fragment polymorphic analysis. *Theoretical & Applied Genetics* 79: 49-55.

MAY, T., AND J. SOLL. 2000. 14-3-3 Proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell* 12: 53-64.

MAYNARD, D. N. 2001. An introduction to the watermelon. ASHS Press, Alexandria, VA, USA.

MCCOURT, S., M. LAVIN, AND R. A. SHARROCK. 1995. Using *rbc*L sequences to test hypotheses of chloroplast and thallus evolution in conjugating green algae. *Journal of Phycology* 31: 989-995.

Metzlaff, M., T. Borner, AND R. Hagemann. 1981. Variations of chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. *Theoretical & Applied Genetics* 60: 37-41.

MÖLLER, M., M. CLOKIE, P. CUBAS, AND Q. C. B. CRONK. 1999. Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study. Taylor & Francis, London, England.

MITTON, J. B. 1994. Molecular approaches to population biology. *Annual review of Ecology and Systematics* 25: 45-69.

MSELLEM, L., J. T. NOYER, T. LE BOURGOIS, AND M. HOSSAERT-MCKEY. 2000. Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology* 9: 443-455.

NEWTON, A. C., T. R. ALLNUTT, A. C. M. GILLIES, A. J. LOWE, AND R. A. ENNOS. 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution* 14: 140-145.

NISHIKAWA, K. 1992. A guide to the wheat aneuploids. *Wheat information service* 74: 1-3.

ODA, K., K. WAMATO, E. OHTA, Y. NAKAMURA, M. TAKEMURA, N. NOZATO, K. AKASHI, T. KANEGAE, Y. OGURA, T. KOCHI, AND K. OHYAMA. 1992. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA : A primitive form of plant mitochondrial genome. *Journal of Molecular Biology* 233: 1-7.

O'DONNELL, K. 1992. Ribosomal DNA internal transcribed spacers are highly divergent

in the phytopathogenic ascomycete *Fusarium sambucinum (Gibberella pulicaris)*. *Current Genetics* 22: 213-220.

OGIHARA, Y., K. ISONO, T. KOJIMA, A. ENDO, M. HANAOKA, T. SHIINA, T. TERACHI, S. UTSUGI, M. MURATA, N. MORI, S. TAKUMI, K. IKEO, T. GOJOBORI, R. MURAI, K. MURAI, Y. MATSUOKA, Y. OHNISHI, H. TAJIRAI, AND K. TSUNEWAKI. 2000. Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: Complete sequence and contig clones. *Plant Molecular Biology Reporter* 18: 243-253.

OHYAMA, K., H. FUKUZAWA, T. KOCHI, H. SHIRAI, T. SANAO, S. SANO, K. UMESONE, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S. AOTA, H. INOKUCHI, AND H. OZEKI. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572-574.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205-1224.

PALMER, J. D. 1991. Plastid chromosomes: structure and evolution. *Cell Culture Somatic Cell Genetics of Plants* 7: 55-92.

PALMER, J. D. 1992. Mitochondrial DNA in plant systematics: application and limitations. Chapman & Hall, New York, NY, USA.

PAPA, R., AND P. GEPTS. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical & Applied Genetics* 106: 239-250.

PETERSON, J., H. BRINKMANN, AND R. CERFF. 2003. Origin, evolution, and metabolic role of a novel glycolytic GAPDH enzyme recruited byland plant plastids. *Journal of Molecular Evolution* 57: 16-26.

PETIT, J. R., J. DUMINIL, S. FINESCHI , A. HAMPE , SALVINI , D., AND G. G. VENDRAMIN. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14: 689-701.

RANDALL, L. S., A. R. JULIE, C. C. RICHARD, S. TOSAK, AND F. W. JONATHAN. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301-1315.

RIDOUT, C. J., AND P. DONINI. 1999. Use of AFLP in cereals research. *Trends of Plant Science* 4: 76-79.

RINDOS, D. 1984. The origins of agriculture- an evolutionary perspective. Academic Press, Orlando, Florida, USA.

ROBINSON, R. W., AND D. S. DECKER-WALTERS. 1997. Cucurbits. CAB International, New York.

SAGER, R., AND M. R. ISHIDIA. 1963. Chloroplast DNA in *Chlamydomonas*. *Proceedings of the National Academy of Sciences of the United States of America* 50: 725-730.

SALAMINI, F., H. OZKAN, A. BRANDOLINI, AND W. MARTIN. 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics.* 3: 429.

SANG, T., D. J. CRAWFORD, AND T. F. STUESSY. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). A*merican Journal of Botany* 84: 1120-1136.

SATO, S., Y. NAKAMURA, T. KANEKO, E. ASAMIZU, AND S. TABATA. 1999. Complete stucture of the chloroplast genome of *Arabidopsis thaliana*. D*NA Research* 6: 238-290.

SAUER, J. D. 1993. Historical geography of crop plants. CRC press, BocaRaton, FL, USA.

SCHAAL, B. A., D. A. HAYWORTH, K. M. OLSEN, J. T. RAUSCHER, AND W. A. SMITH. 1998. Phylogeorgraphic studies in plants: problems and prospects. *Molecualr Ecology* 7: 465-474.

SCHMITZ-LINNEWEBER, C., R. M. MAIER, J. P. ALCARAZ, A. COTTER, R. G. HERMANN, AND R. MACHE. 2001. The plastid chromosome of spinach (Spinacia oleracea): Complete nucleotide sequence and gene organization. *Plant Molecular Biology Reporter* 45: 307-315.

SHINOZAKI, K., M. OHEM, M. TANAKA, T. WAKASUGI, N. HAYSHIDA, T. MATSUBAYASHI, N. ZAITA, J. CHUNGWONGSE, J. OBOKATA, K. YAMAGUCHI-SHINOZAKI, H. DENO, T. KAMOGASHIRA, K. YAMADA, J. KUSUDA, F. TAKAIWA, A. KATO, N. TOHDON, H. SHIMADA, AND M. SUGURIA. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal* 5: 2043-2049.

SIMPSON, C. L., AND D. B. STERN. 2002. The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiology* 129: 957-966.

SMITH, B. D. 1998. The emergency of agriculture. W.H.Freeman Company, New York, NY, USA.

SOLTIS, D. E., P. S. SOLTIS, D. L. NICKRENT, L. A. JOHNSON, W. J. HAHN, S. B. HOOT, J. A. SWEERE, R. K. KUZOFF, K. A. KRON, M. W. CHASE, S. M. SWENSEN, E. A. ZIMMER, S.-M. CHAW, L. J. GILLESPIE, W. J. KRESS, AND K. J. SYTSMA. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1-49.

SOLTIS, E. D., AND P. S. SOLTIS. 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant Molecular Biotechnology* 42: 45-75.

SUGIURA, M. 1992. The chloroplast genome. *Plant Molecular Biolog* 19: 149-168.

SUGIURA, S., Y. KOBAYASHI, S. AOKI, C. SUGITA, AND M. SUGITA. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpo*A from the chloroplast to the nucleus. N*ucleic Acids Research.* 31: 5324-5331.

SWOFFORD, D. L. 2002. Phylogenetic analysis using parsimony (* and other methods). Sinauer Associates, Sunderland, MA, USA.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

TALHINHAS, P., J. NEVES-MARTINS, AND J. LEITAO. 2003. AFLP, ISSR and RAPD markers reveal of genetic diversity among *Lupinus* spp. *Plant Breeding* 122: 507-510.

TROITSKY, A. V., AND V. K. BOBROVA. 1986. 23S rRNA-derived small ribosomal RNAs: their structure and evolution with references to plant phylogeny. I*n* S. K. Dutta [ed.], In DNA systematics, vol 2: Plants, 137-170. CRC press, Boca Raton, Florida.

UNSELD, M., J. R. MARIENFELD, P. BRANDT, AND A. BRENNICKE. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15: 57-61.

VEKEMANS, X., T. BEANUWEMS, M. LEMAIRE, AND I. ROLDAN-RUIZ. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular ecology* 11: 139-151.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VAN DE LEE, M. HORNES, A. FRIJTERS, J. POT, J. PELEMAN, AND M. KUIPER. 1995. AFLP: a new concept for DNA fingerprinting. *Nucleic Acids Research.* 23: 4407-4414.

WAKASUGI, T., J. TSUDZUKI, S. ITO, M. SHIBATA, AND M. SUGURIA. 1994. A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology Reporter* 12: 227-241.

WAKASUGI, T., T. TSUDZUKI, AND M. SUGIURA. 2001. The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. *Photosynthesis Research.* 70: 107-118.

WALL, D. P. 2002. Use of the nuclear glyceraldehydes 3-phophate dehygrogenase for phylogeny reconstruction of recently diverged lineages in *Mitthyridum* (Musci: Calymperaceae). *Molecular Phylogenetics and Evolution* 25: 10-26.

WALLACE, C. D. 1982. Structure and evolution of organelle genomes. *Microbiological Reviews* 46: 208-240.

WANG, H., G. DUBY, B. PURNELLE, AND M. BOUTRY. 2000. Tobacco VDL Gene encodes a plastid DEAD box RNA helicase and is involved in chloroplast differentiation and plant morphogenesis. *Plant Cell* 12: 2129-2142.

WANGER, D. B., D. R. GOVINDARAJU, AND B. P. DANCIK. 1987. Chloroplast DNA variation in the sympatric zone of lodgepole pine *(Pinus contorta* Dougl.) and jack pine (*P. banksiana* Lamb.). *Proceedings of the Southern Forest Tree Improvement Conference.* 41: 119-127.

WAUGH, R., N. BONAR, E. BAIRD, B. THOMAS, A. GRANER, P. HAYES, AND W. POWELL. 1997. Homology of AFLP products in three mapping populations of barley. *Molecular and General Genetics* 255: 311-321.

WELSH, J., AND M. MCCLELLAND. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 236-239.

WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Kluwer Academic publisher, Boston, MA, USA.

WHITCHER, I. N., AND J. WEN. 2001. Phylogeny and biogeography of *Corylus* (Betulaceae): inferences from ITS sequences. *Systematic Botany* 26: 283-298.

WHITLOCK, B. A., AND D. A. BAUM. 1999. Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene Vicilin. *Systematic Botany* 24: 128-138.

WILLIAMS, J., A. KUBELIK, K. LIVAK, J. RAFALSKI, AND S. TINGEY. 1990. Oligonucleotide primers of arbitrary sequence amplify DNA polymorphisms which are useful as genetic markers. N*ucleic acids research* 18: 6531-6535.

WILSON, C. A. 2003. Phylogenetic relationships in Iris series Californicae based on ITS sequences of nuclear ribosomal DNA. *Systematic Botany* 28: 39-46.

WOITSCH, S., AND S. RÖMER. 2003. Expression of Xanthophyll Biosynthetic genes during light-dependent chloroplast differentiation. *Plant Physiology* 132: 1508-1517.

Wolfe, K. H., C. W. Morden, AND J. D. Palmer. 1992. Function and Evolution of a Minimal Plastid Genome from a Nonphotosynthetic Parasitic Plant. *Proceedings of the National Academy of the Sciences of the United States of America* 89: 10648-10652.

YAMADA, T., AND M. SHIMAJI. 1986. Peculiar feature of the organization of rRNA genes of the *Chlorella* chloroplast DNA. *Nucleic Acids Research* 14: 3827-3839.

ZAMIR, D., N. NAVOT, AND J. RUDICH. 1984. Enzyme polymorphism in C*itrullus lanatus*

and C. *colocynthis* in Israel and Sinai. *Plant Systematics and Evolution* 146: 163-170.

ZANIS, M. J., D. E. SOLTIS, P. S. SOLTIS, S. MATHEWS, AND M. J. DONOGHUE. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of the Sciences of the United States of America* 99: 6848-6853.

ZANIS, M. J., D. E. SOLTIS, P. S. SOLTIS, Y.-L. QIU, AND E. ZIMMER. 2003. Phylogenetic analyses and perianth evolution in basal angiosperms. *Annals of the Missouri Botanical Garden* 90: 129-150.

ZOHARY, D., AND M. HOPF. 1988. Domestication of plants in the old world. Oxford University Press, Oxford, UK.