BIOCHEMICAL, MOLECULAR AND PHYLOGENETIC STUDIES OF SELECTED WILD *SOLANUM* SPECIES OF MEGHALAYA

By
INSAN ARA RAHMAN

THESIS SUBMITTED
IN FULFILMENT OF THE DEGREE OF
DOCTOR OF PHILOSOPHY IN BOTANY

NORTH EASTERN HILL UNIVERSITY
SHILLONG 793022, INDIA
2014

DEDICATED TO .....

BABA & MAA

The North Eastern Hill University

January, 2014

DECLARATION

I, Insan Ara Rahman, hereby declare that the subject matter of this thesis entitled "Biochemical, molecular and phylogenetic studies of selected wild *Solanum* species of Meghalaya" is the record of work done by me, that the contents of this thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/Institute.

This is being submitted to the North-Eastern Hill University for the award of the degree of Doctor of philosophy in Botany.

Insan Ara Rahman

Insan Ara Rahman

Prof. Arvind K. Misra

(Supervisor)

(Head of the Department)

Head
Department of Botany
Schools of Life Sciences
N.E.H.U. Shillong-22

Prof. D. Syiem

(Joint-supervisor)

Prof. D. Syiem
Department of Biochemistry
North-East Hill University
Shillong - 793022

Dr. Y. Kumar

(Joint-supervisor)

ACKNOWLEDGEMENTS

The journey of Ph.D. as everyone experiences is never through a fine carpet, but still I have enjoyed each and every moment of this journey only because of the surrounding atmosphere created by my near and dear ones with their support and encouragement. Now here I am standing in the last point of my journey where I have ultimately got the chance to acknowledge each and everyone related to me directly or indirectly during this journey.

First and foremost I want to pay my homage to that person without whose guidance, support and encouragement I could have never fulfill y my dream of having Ph.D. degree and he is none other than my Supervisor Prof. Arvind K. Misra. I feel, the words "Thank you Sir..." is not enough to express my gratitude towards him. He is not only an owner of attractive personality but also a good human being. For any kind of problem he has always the solution. I was never disheartened if wanted to consult him even for smallest matters, in spite of his busy schedule. He has enlightened me not only with knowledge but also teaches me how to serve people and also how to be a disciplined as well as loyal person.

At this time of accomplishment I want to thank my Joint-supervisors Prof. D. Siyem and Dr. Y. Kumar for their support and many insightful suggestions. Prof. D. Siyem has always welcomed me in his lab for any kind of help that I wanted during my biochemical work. Dr. Y. Kumar has also helped me in understanding the taxonomical terms.

I would always be grateful to the present Head of the Botany department, Prof. N. Venugopal and former Head of the department Prof. M. S.

Dkhar for providing necessary infrastructure and resource to accomplish my research work.

I want to specially thank Prof. V. K. Kapoor, Punjab University, presently the Principal of Khalsa College of Phermaceutical Science, Punjab, whom I will never forget during my life time. He is a person whom I have never met personally, I talked to him over phone and he sent me the pure Solasodine which has a very high cost value in the market and moreover it is almost out of market now a days. I will always remain grateful to him for his immense help.

I was lucky to be a part of Molecular Genetics Lab where I have got a good support system. My senior lab mates Mhathung Yanthan, David L. Biate, Meena Elangbam, Juabanry Nongkynrih and Asem Mipeshwaree Devi and junior Chipem Vashi were always encouraging and supportive. They all were just like members of the same family.

I am indebted to many of my student colleagues who provided a funfilled environment. Here I should mention the name of Nilofer Sheikh who is always a wonderful sister cum friend. She was always supportive and she had 'ready to help' kind of attitude at any situation. I shall always be grateful to her. I should also acknowledge Mr. Bodhisattwa Banerjee for his immense help during DNA sequencing. I should thank Neeta Pathaw, Ganga Ayam, Rajkumari Jashmi, Lucy Nongbri, Debashree Nath, Wilson Monsang, Val Saio, Kiran Gogoi and Sushmita Barua for their constant moral support. There are so many names that come to my mind whom I should thank but the list would be so long to mention here. I want to thank all of them (you know who you are...!).

I expand my thanks to all the non-teaching staff of Botany department. They are always very helpful and approachable. I want to pay my sincere gratitude towards the members of Bioinformatics Centre of NEHU, especially Mr. Santosh Vishwakarma for his helping hand during handling the DNA sequences.

I specially thank my Baba, Maa, my brother, sister and aunt. They were always supportive throughout my life. I really fill lucky whenever I think about the faith that they have on me. Having a Ph.D. degree is not only my dream, it's my family's dream. I want to fulfill their dream too. Here I want to thank my in-laws who all are very supportive and caring.

I also want to thank both UPE-Biosciences programme and Maulana Azad National Fellowship Scheme for Minority students of UGC for providing financial support.

I want to thank my best friend, soul mate and my husband Md. Saiful Islam. He has been a true and great supporter. Whenever I am disappointed he consoles me in such a way that all my problems seem to become very small. According to him there is nothing like 'impossible' kind of thing in the world, he believes in hard work and wanted me to do the same and that's how he encourages me. Here I should specially mention the name of the new member of our family, my daughter Saarah, who is only of few months now but still she is very supportive to her mom while her mom is busy writing her thesis. It would have been very difficult to finish writing the thesis without both of their cooperation. So I specially thank my husband and my daughter.

At last I want to thank God for making this journey possible.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

CHAPTER 1

INTRODUCTION

The evolution and diversification of animals and plants go hand in hand and they depend on each other for successful completion of their life cycles. The need for food, medicine, timber, etc. resulted in the exploitation of wild plants. The origin of human civilization and simultaneous domestication and cultivation of some useful plants were the main turning points which paved the way for modern world. But still the majority of the plants on this earth are in the wild state and their usefulness for humans has yet to be discovered. The systematists and evolutionary biologists are trying to establish the relationship between the wild and cultivated plant species. Phylogenetic analyses have revealed the closest relatives of many crop and 'model organisms' for genetic and molecular studies and indicate the possible relatives for use in comparative studies and plant breeding. According to most systematists and evolutionary biologists phylogeny should be the central underpinning of research in much of biology. Dobzhansky (1973) stated that nothing in biology makes sense except in the light of evolution. A close corollary is that everything makes a lot of sense in the light of phylogeny. For example, the phylogenetic inferences that the model organism *Arabidopsis thaliana* is actually part of a broadly circumscribed Brassicaceae that also includes Capparaceae, and that Brassicaceae are, in turn, part of well-supported 'glucosinolate clade', are important to those hoping to test the applicability of the ABC model of floral development. If

this model holds throughout a broadly defined Brassicaceae, then the closest relatives of the family become the focal point for future studies.

The classical method of studying phylogeny was based on morphological characteristics of plants which used to create lot of taxonomic confusion. This problem was overcome in 1950s when two events took place simultaneously creating renaissance in phylogenetics. The first was the discovery of molecular structure of DNA – the double helix by Watson and Francis Crick in 1953 and second event took place in 1955 when Fred Sanger and colleagues, published the first comparison of amino acid (the product of DNA) sequences from three different species- cattle, pig and sheep. This was the first study at the molecular level which reveals that the insulin of cattle, pig and sheep differed in three amino acids indicating that their insulins had evolved along with their anatomical features. One of the first example of molecular systematic- the use of gene sequences to reconstruct phylogenetic relationship was the work of Sarich and Wilson (1969) which changed our perspective on human origins and opened the 'molecules versus morphology' debate about which data set are the best markers of evolutionary relationship. The field of molecular phylogenetics blossomed with the emergence of new techniques such as Polymerase Chain Reaction (PCR) technique in 1983, invented by Kary Mullis. The PCR technique enables us to amplify specific regions of DNA to million folds. Now a day's various PCR based techniques have been utilized to develop different types of molecular markers to assess the diversity within plants and animals in different taxonomic levels e.g. Sequence Tagged Sites (STS), Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), Amplicon Restriction Patterns (ARPs), Variable Number Tandem Repeats (VNTRs),

Single Nucleotide Polymorphisms (SNPs), Short Sequence Repeats (SSRs) etc. In case of STS marker the nucleotide sequences for the targeted segments of DNA are already known. This is a very reliable process for developing molecular marker where pair of PCR primers are designed for amplification of specific DNA regions/loci which may or may not be coding regions. There are two approaches utilizing this technique - Sequence Characterized Amplified Region (SCAR) and Sequence Tagged Amplified Region (STAR). Expressed Sequence Tags (EST) is special type of STS which target coding regions and are useful for identification of superior genotypes. AFLP is an ideal technique to detect polymorphism within the genome of species as well as at varietal level. PCR primers are designed to anneal to the oligonucleotide adaptor + restriction site + few bases of original DNA fragment. RAPD is also used for diversity study where a single PCR primer of 10 bases is used which generates large number of bands after PCR amplification. This is a random amplification process because the primer does not target a specific region in the DNA. That is why this technique is not reliable and non-reproducible.

The use of molecular data sets in phylogenetic studies revolutionized the field of plant systematics and phylogeny. One most important example is *Solanum lycopersicum* which was formerly treated and classified as *Lycopersicon esculentum* representing a distinct genus *Lycopersicon* in the family Solanaceae. But molecular phylogenetic analysis revealed that tomato, formerly classified as *Lycopersicon esculentum*, is clearly embedded within the genus *Solanum* and therefore does not represent a distinct genus. *Lycopersicon esculentum* has been formerly transferred to *Solanum* as *Solanum lycopersicon* (Spooner and Sytsma, 1992). Another interesting dispute lying within the genus *Myrica* was also clarified

using molecular data by Yanthan *et al.* (2011). Morphological diversity within the genus *Myrica* led to the confusion among various workers. Some workers suggested that different morphotypes (morphotype-1, morphotype-2 and morphotype-3) of the plant belong to different species. According to one school of thought there are two separate species (*Myrica nagi* Thunb. and *Myrica esculenta* Buch.-Ham. ex D. Don) whereas others suggested that *M. nagi* and *M. esculenta* are synonyms. Based on rDNA sequence analysis, Yanthan *et* al. (2011) proposed that *M. nagi* and *M. esculenta* should be treated as two separate species. At the upper end of the taxonomic hierarchy, molecular phylogenies have not only provided insights into the origin of land plants (Kranz *et al.*, 1995; Huss and Kranz, 1997; Bhattacharya and Medlin, 1998), but have also clarified relationship among: all green plants (Källersjö *et al.*, 1998), all major groups of extant land plants (Manhart, 1994; Kranz *et* al., 1995; Kranz and Huss, 1996; Qiu and Palmer, 1998; Soltis *et al.*, 1999) and major groups of angiosperms (Chase *et al.*, 1993; Soltis *et al.*, 1998; Savolainen *et al.*, 2000).

Within Angiosperm families the Solanaceae ranks as one of the most important plant family in regard to its use for human beings. Species of the family are used for food (e.g. *Solanum tuberosum* L., the potato; *S. lycopersicum* L., the tomato; *S. melongena* L., the eggplant), drugs (e.g. *Nicotiana tabacum* L. and *N. rustica* L., tobacco; *Atropa belladonna* L., deadly nightshade; *Mandragora officinarum* L., mandrake; *Duboisia* spp., sources of commercial alkaloids) and as ornamentals (e.g. *Petunia hybrida* hort., petunia; *Salpiglossis sinuata* Ruiz & Pav'on, velvet tongue; *Schizanthus pinnatus* Ruiz & Pav'on, butterfly flower). About 90 genera and 2000 species of Solanaceae are found in tropical and temperate regions of

the world. In India, 15 genera and 88 species are reported out of which 10 genera and 28 species are found in Meghalaya (Haridasan and Rao, 1987). Owing to its economic and medicinal importance several studies including taxonomy and comparative genomics have been undertaken. Members of the Solanaceae are extremely diverse; in terms of habit, ranging from trees to small annual herbs; in habitat, from deserts to the wettest tropical rain forests; and in morphology, with astounding variation in many characters of both flowers and fruits. Due to its large size and tropical centre of diversity, resolving evolutionary relationships across Solanaceae as a whole has been challenging. The family has also served as a model for the study of incompatibility and for nuclear genome organization and evolution (Tanksley *et al.*, 1988). There are several genetic maps for Solanaceae: one for *Capsicum* (Livingstone *et al.*, 1999) and one for *Solanum* section *Etuberosum* (Perez *et al.*, 1999), the immediate sister to the clade of tomato and potato. From last few decades, the Solanaceae have remained in the forefront of molecular plant systematics e.g., chloroplast RFLP analysis (Palmer and Zamir, 1982; Hosaka et al., 1984); chloroplast restriction site mapping (Olmstead and Palmer, 1992; Spooner *et al.*, 1993); chloroplast *ndhF* and *rbcL* sequences (Olmstead and Sweere, 1994; Bohs and Olmstead 1997); nuclear ITS and "waxy" sequences (Walsh and Hoot, 2001; Levin and Miller, 2005; Whitson and Manos, 2005; Levin *et al.*, 2005; Martine *et al.*, 2006); nuclear SAMT sequences (Martins and Barkman, 2005); AFLP analysis (Mace *et al.*, 1999a, b; Spooner *et al.*, 2005a, b) and nuclear retroposon markers (Yuan *et al.*, 2006). A summary of progress in molecular phylogenetic studies of Solanaceae indicates that with over 50 published studies, more than 90% of genera and 37% of species have been sampled. Several clades in Solanaceae have been subject to detailed study, including *Nicotiana* (Aoki and Ito, 2000; Clarkson *et al.*,

2004), *Capsicum* (Walsh and Hoot, 2001; E. Dean and L. Bohs, pers. commun.), *Lycium* (Miller 2002; Levin and Miller, 2005); Goetzeoideae (Santiago-Valentin and Olmstead, 2003), Anthocercideae (Garcia and Olmstead, 2003), Physalinae (Whitson and Manos, 2005), *Petunia* (Ando *et al.*, 2005; Kulcheski *et al.*, 2006); Iochrominae (Smith and Baum, 2006), *Nolana* (M. Dillon and J. Wen, pers. commun.), and, of course, *Solanum* (e.g., Spooner *et al.*, 1993; Bohs, 2005; Levin *et al.*, 2006; Weese and Bohs, 2007). According to International Solanaceae Genomics Network (SOL) - "For several thousand years, Solanaceae crops have been subjected to intensive human selection. This has led to an enormous phenotypic diversity within species and the adaptation of individual varieties to widely different habitats. Showing this great interspecific and intraspeicific diversity and being the most important plant family of vegetable crops, Solanaceae have recently become a model of comparative and evolutionary genomic research."

Solanum L. is placed within the subfamily Solanoideae, characterized by flattened seeds and with curved embryos (Hunziker, 1979) with centre of diversity in South America. With an estimated 1700 species, *Solanum* is the largest genus in the family Solanaceae and one of the largest genera of flowering plants with high diversity. In India 40 species are found of which 13 are found in Meghalaya (Haridasan and Rao, 1987). Many of the economically important crop plants belong to this genus e.g. tomato (*Solanum lycopersicon*), potato (*Solanum tuberosum*), and egg plant (*Solanum melongena*). Various species of *Solanum* are also used in various parts of India in the treatment of diabetes, bronchitis, asthma, dysentery, etc. The hyperdiversity in this genus makes *Solanum* interesting from an evolutionary standpoint as well as for its usefulness to humans.

The genus *Solanum* has been extensively studied to resolve its systematic position and to assign the subgeneric groups within the genus. Some of the infrageneric groups in which molecular phylogenetic studies have been thoroughly conducted are– sect. *Petota* by Ames and Spooner (2010); subg. *Leptostemonum* by Levin *et al.*, (2006); sect. *Gonatotrichum* by Stern and Bohs (2012) and sect *Acanthophora* by Levin *et al.* (2005), etc. Ortiz (2001), Bradshaw *et al.* (2006) and Jansky (2009) have stated that wild potatoes have been proven to be a source of important traits which can be utilized in plant breeding of crop plants. The wild *Solanum* like *Solanum nigrum* have been identified as a valued genetic resource for plant breeding program as it is resistant to a disease called Late blight which is one of the economically most important Solanaceous plants disease caused by *Phytophthora infestans* (Mont.) de Bary (Kamoun et.al, 1999; Campos et.al, 2002; Flier et.al, 2003; Lebecka 2008).

The genus *Solanum* contains a number of species which possess active physiological properties and some of which have been used to a considerable extent in medicine. Various secondary metabolites like phenolics, alkaloids and terpenoids are present in *Solanum.* Most prominent medicinal properties are due to presence of alkaloids, solamargin and solasonin, which yield solasodine as glycone has great demand in pharmaceutical industry. Solasodine, the nitrogen analogue of diosgenin, has been reported as a valuable steroidal precursor for the supplementary source of the commercial synthesis of several steroidal drugs (Rodriguez *et al.*, 1979; Sree *et al.*, 1982).

Not much work has been done till date in the field of molecular phylogeny and biochemical analysis of wild relatives of *Solanum* found in

Meghalaya. So this study was conducted toassess the phylogenetic relationship among the wild species of *Solanum* found in Meghalaya by using sequence information from two nuclear regions, *18S rRNA* gene and *rrn* ITS and two chloroplast genes *rbcL* and *matK*, protein profiling and phytochemical screening of fruits.

REVIEW OF LITERATURE

CHAPTER2

REVIEW OF LITERATURE

Genus *Solanum* has been extensively studied morphologically for systematic positioning. *Solanum* is distinguished from most other genera in the tribe Solaneae by its poricidal anther dehiscence, a character present in nearly all *Solanum* species. Although poricidal anther dehiscence is the striking synapomorphy which recognises *Solanum* as a genus, its infrageneric subdivision is less clear. Linnaeus (1753) divided *Solanum* into two groups, *Spinosa* and *Inermia*, based on presence or absence of spines. Dunal (1852) established two major infrageneric divisions ("sections") based on anther shape and presence or absence of spines – '*Pachystemonum*' which includes species that lack spines and have relatively short, broad anthers with large terminal pores; and '*Leptostemonum*' that include spiny species with relatively narrow, distally tapered anthers with small terminal pores. But Seithe (1962), in contrast gave more emphasis on hair types rather than spininess or anther morphology. Though these classifications were not popular worldwide, but these provided the elements of D'Arcy's (1972) classification scheme which is the most widely used system today. D'Arcy's scheme recognized seven subgenera in *Solanum* (D'Arcy, 1972, 1991). These seven subgenera are - 1. *Solanum* subg. *Archaesolanum* Marzell, 2. *Solanum* subg. *Bassovia* (Aubl.) Bitter, 3. *Solanum* subg. *Leptostemonum* (Dunal) Bitter, 4. *Solanum* subg. *Lyciosolanum* Bitter, 5. *Solanum* subg. *Minon* Raf. [subg. *Brevantherum* (Seithe) D'Arcy, in D'Arcy (1972)], 6.

Solanum subg. *Potatoe* (G. Don) D'Arcy, 7. *Solanum* subg. *Solanum.* These subgenera ranged in size from monotypic subgenus *Lyciosolanum* to genera *Leptostemonum*, *Solanum* and *Potatoe* which includes hundreds of species. Whalen (1984) provided a detailed study of the subgenus *Leptostemonum.* An infrageneric extensive work was carried out by both Nee (1999) and Child and Lester (2001). Nee (1999) listed only New World taxa in his studies while Child and Lester (2001) listed only the type species for each of their infrageneric groups. Hunziker (2001) modified D'Arcy's (1972) system and provided descriptions for each section.

According to Daunay and Lester (1988) great degree of taxonomic confusion exists as regard to genus *Solanum.* The advent of molecular data has revolutionized the field of plant systematics and has led to new insights into phylogenetic relationships at all taxonomic levels. In the Solanaceae, Olmstead and colleagues have used restriction site and sequence data to examine phylogenetic relationships across the entire family (Olmstead & Palmer, 1992; Olmstead *et al.*, 1999). Molecular studies above the sectional level in *Solanum* include the works of Spooner *et al.* (1993), Olmstead and Palmer (1997), and Bohs and Olmstead (1997, 1999, 2001). These studies provide information on major clades within *Solanum*, but none have sampled from all the subgenera recognized by morphological systematists such as Bitter, Seithe, Danert, and D'Arcy.

Molecular data from chloroplast *ndhF* sequences identify about 13 major clades within *Solanum* (Bohs, 2005). Bohs (2005) presented results of a molecular phylogenetic study designed to identify major clades within *Solanum* using sampling from a broad spectrum of *Solanum* subgroups. Results are presented from an analysis of sequence data from the chloroplast gene *ndh*F. Sampling includes members of all seven of D'Arcy's subgenera and over 40 of the 62 sections listed in

D'Arcy (1991). All the sections listed in D'Arcy's (1972) conspectus as well as many sections described after 1972 are discussed in context of the major *ndh*F clades. Major lineages are described with informal clade names and their component sectional groups are listed. Non-molecular synapomorphies are proposed for many of the clades. Continued use of informal taxonomic designations is advocated for new infrageneric groups within *Solanum*. *Solanum laciniatum* is a member of the Archaeosolanum clade, a strongly supported monophyletic group in molecular analyses using plastid DNA sequences, and is of no strong affinity with other groups of *Solanum* (Bohs, 1995). Symon (1981) placed *S. laciniatum* in series *Laciniata* Geras., with *S. vescum* and *S. linearifolium*. In Symon's (1994) preliminary phylogeny based on morphology, *Solanum laciniatum* is sister to *S. aviculare* + *S. multivenosum*; that small group is sister to *S. vescum*.

With new molecular data from the chloroplast *trn*T-F region and the nuclear granule-bound starch synthase gene (GBSSI or waxy), it is confirmed that *Solanum vespertilio* and *Solanum lidii* are phylogenetically associated with *Solanum* lineages from Africa, rather than with previously suggested Maxican species (Anderson *et al.*, 2006). Weese and Bohs (2007) suggest that most traditionally recognized *Solanum* subgenera are not monophyletic based on chloroplast *ndh*F and *trn*T-F, and nuclear waxy locus. The *Thelopodium* clade is sister to *Solanum*, which is split into two large clades. These two large clades are further divided into at least 10 subclades, for which informal names are provided and morphological synapomorphies are proposed.

Nuclear DNA restriction enzyme site and chloroplast DNA analysis showed *Solanum bulbocastanum* and *Solanum cardiophylum* to be part of two main

clades (Rooriguez and Spooner, 2002). Leptostemonum were investigated to clarify the species circumscription and taxonomic issues. Eighty samples were analysed based on sequence variability of the internal transcribed spacers (ITS) of the nuclear ribosomal DNA, as well as the chloroplast intron *trn*L, *trn*L-*trn*F and *trn*S-*trn*G spacers. Five different matrixes were analyzed as the basis for a phylogenetic approach. The combined data from the chloroplast analyses formed well-supported trees. *Solanum* section *Torva* was proposed as a monophyletic group and it is close to *S. melongena*, *S. jamaicense* and *S. sisymbriifolium* (Miz *et al.*, 2007).

Regional treatments of *Solanum* have been provided by Symon (1981), Jaeger (1985), and Nee (1999). However, the most comprehensive treatment of the ''spiny *Solanums*'' is that of Whalen (1984). Based on morphology and biogeography, Whalen (1984) recognized 33 informal species groups within subgenus *Leptostemonum* as well as 36 unplaced *Solanum* species. Whalen (1984) summarized the diagnostic characters, geographical distributions, and component species of each of the 33 species groups and placed them in a hypothetical phylogenetic scheme based on morphological characters. Although .Whalen's treatment is arguably the most useful, but Whalen (1984) was not the first to recognize taxonomic groups of species within subgenus *Leptostemonum*. Molecular phylogenetic studies of the genus *Solanum* based on chloroplast DNA restriction sites (Olmstead and Palmer, 1997) support the monophyly of subgenus *Leptostemonum* sensu (Whalen, 1984). However, more recent studies using chloroplast and nuclear DNA sequence data and greater taxon sampling (Bohs and Olmstead, 1997, 1999, 2001; Bohs, 2005) suggest that the *S. wendlandii* group and perhaps also the *S. nemorense* group, both of which have prickles but lack stellate

hairs, may not belong within the subgenus. Phylogenetic hypotheses are presented for the Australian species in *Solanum* subgenus *Leptostemonum* (the'spiny *solanums*') section *Melongena* which contains10 of the 14 currently described dioecious species in the genus. Phylogenetic analysis of the ITS and *trn*K – *mat*K gene regions supports a single origin of dioecy from andromonoecy in Australian *Solanum* (Martine *et al.*, 2009).

2.1 Role of 18S rDNA in Molecular phylogeny

The 18S rDNA region, which encodes the RNA of the small ribosomal subunit, is the most widely used gene in molecular phylogenetics studies. This is because of its ubiquitous presences across the tree of life. It also exhibits similar phylogenetic signals at equivalent taxonomic ranks (Gerbi, 1985; Hillis and Dixon, 1991; Embley *et al.*, 1994; D. Soltis and Soltis, 1995). The first major effort to apply 18S data to angiosperm phylogeny was undertaken in the late 1980s and based on direct RNA sequencing of portions of both the 18S and 26S regions (Hamby and Zimmer, 1988, 1992; Zimmer *et al.*, 1989). Subsequent studies (Nickrent and Franchina, 1990; Boulter and Gilory, 1992; Bharathan and Zimmer, 1995) contributed to a growing plant small subunit database, with the trend toward complete 18S DNA sequences.

2.2 The ITS sequences of the Nuclear Ribosomal DNA

The nuclear ribosomal DNA has proved to be a powerful phylogenetic tool because of its ubiquitous presence in all organisms and the presence of repeat units in high copy number (Hamby and Zimmer, 1992). The internal transcribed spacer (ITS) region of 18S- 26S nuclear rDNA have become a major focus of

comaparative sequencing at the specific and generic levels in angiosperms (Baldwin *et al.*,1995). The internal transcribed spacers (ITS1 and ITS2) are part of the nuclear rDNA transcript but are not incorporated into ribosome. They appear to play a role in the maturation of nuclear *rRNA*s , bringing the large and small subunits in close proximity within a processing domain (Baldwin *et al.*,1995). ITS sequences have been utilized for reconstruction of phylogenetic relationship in angiosperms (Baldwin *et al.*, 1995), algae (Bakker *et al.*, 1995; Coleman *et al.*, 1994). ITS sequences have also been used to infer phylogeny at lower taxonomic levels in a diverse array of organisms.

2.3 The *rbcL* gene and its role in phylogenetic study

Organelle DNA sequences have been used extensively in phylogenetic studies on plants. Chloroplast DNA (cpDNA) variation has proven to be immensely valuable in reconstructing phylogenies at the species and higher taxonomic level. The gene *rbcL* is located in the large single copy region of the chloroplast genome and encodes the large subunit of ribulose 1,5- bisphosphate carboxylase (RUBISCO). At family level and above, *rbcL* has, by far, been the preferred gene for inferring phylogeny. One of the factors that had been suggested to have potentially distorting effects on the *rbcL* topology was the occurrence of lineage-specific rate variation. The largest molecular phylogenetic analysis published was based on *rbcL* sequences (Chase, Soltis, Olmstead,*et al.*, 1993). Sequence data derived from *rbcL* have been used to address phylogenetic relationships not only in the angiosperms but also in major groups of green algae and all major lineages of ferns (Hasebe *et al.*, 1994; McCourt *et al.*, 1995). The application of *rbcL* sequence data spans a very wide taxonomic range. The lower limit of applicability of *rbcL* sequences typically

extends to the generic level, but in some groups reaches the specific level. Analysis of *rbcL* sequences have been used to resolve generic relationships within several families of flowering plants (Morgan *et al.*,1994; Xiang *et al.*, 1993; Conti *et al.*, 1993; Soltis *et al.*, 1993).

2.4 The *matK* gene as a tool for phylogenetic study

matK is located in large single-copy region of the chloroplast genome. It is approximately 1,550 bp in length and encodes a maturase involved in splicing type II introns from RNA transcripts. Among protein coding regions in the chloroplast genome, *matK* (formerly ORFK) is one of the most rapidly evolving (Wolfe, 1991). In all photosynthetic land plants so far examined, *matK* is located within an intron of approximately 2,600 bp positioned between the 5' and 3' exons of the transfer RNA gene for lysine, *trnK*. The rate of evolution of *matK* makes this gene appropriate for resolving at intergeneric or interspecific levels in seed plants. In Saxifragaceae *matK* sequences provide a level of resolution compararable to that achieved with cpDNA restriction sites (Johnson and Soltis, 1995). Well resolved generic and species-level phylogenies have been obtained using *matK* sequences in Saxifragaceae (Johnson and Soltis, 1995; Soltis *et al.*, 1996), Polemoniaceae (Johnson and Soltis, 1995; Johnson *et al.*, 1996), and Cornaceae (Xiang *et al.*, 1997), Ericaceae (Kron, 1997) . In several angiosperm families, *matK* and *rbcL* data have been combined, providing enhanced resolution and internal support compared to either gene alone (Soltis *et al.*, 1996; Xiang *et al.*, 1998; Plunket *et al.*, 1997).

2.5 Secondary metabolites and medicinal properties of *Solanum*

The genus *Solanum* contains a number of species which possess active physiological properties and some of which have been used to a considerable extent in medicine. Various secondary metabolites like phenolics, alkaloids and terpenoids are present in *Solanum.* Several studies to identify and isolate the chemical components found in *Solanum* species of nightshades (*Solanum*) in the tomato family (Solanaceae) which reportedly contain a complex of toxic alkaloidal glycosides. Amongst the components identified is solanine, which is now known to consist of six different glycosides, each composed of a particular sugar and the steroidal alkaloid solanidine (Gelder, 2006). Phytochemical investigations have shown the presence of glycoalkaloids like solasodine, solakhasanin, solamargine, and khasinin in *S. khasianum* (Anonymous, 2004; Putalun *et al.*, 2000). *Solanum nigrum* L. (Black night shade) a member of the Solanaceae, has a wide range of medicinal values. The herb is antiseptic, antidysenteric and antidiuretic used in the treatment of cardiac and skin diseases, psoriasis, herpsvirus and inflammation of kidney. The root bark is laxative, useful in the treatment of ulcers on the neck, burning of throat, inflammation of liver and chronic fever. Berries are bitter and pungent and are useful in treatement of heart disease, piles and dysentery (Kritikar and Basu, 1935). Most prominent medicinal properties are due to presence of alkaloids, solamargin and solasonin, which yield solasodine as glycone ha ving a great demand in pharmaceutical industry. Solasodine, the nitrogen analogue of diosgenin, has been reported as a valuable steroidal precursor for the supplementary source of the commercial synthesis of several steroidal drugs (Rodriguez *et al.*, 1979; Sree *et al.*, 1982). It could be primarily obtained from various plants of genus *Solanum* (Crabbe

and Fryer, 1982). Solasodine exists in many forms of glycoalkaloids, which occur mainly as triosides such as, solasonine and solamargine (Crabbe and Fryer, 1982). In recent years, medicinal uses of glycoalkaloids have been a focus of scientific and pharmacological attention. For example, solamargine and solasodine exhibit potent cytotoxicity to human hepatoma cells (Hep3B) by apoptosis which is the major process responsible for cell death in various physiological events (Cheng *et al.*, 1998). Solasodine, solamargine, and solasonine from *Solanum incanum* L. showed liver protective effects against CCl4-induced liver damage (Lin *et al.*1990). Plants that belong to genus *Solanum* have various medicinal properties. *S. nigrum, S. sysimbrifolium, S. khasianum, S. torvum, S. kurzii,* etc. are found in wild state in Meghalaya. People of Meghalya use *S. khasianum* and *S. torvum* for toothaches (Kharkongor and Josheph, 1981). Root decoction of *S. khasianum* eases body pains and indigestion (Roeklein and Leung, 1987). S. *nigrum* has a long history of medicinal usage. The plant has emollient, diuretic, antiseptic and laxative properties. It shows protective effect on the liver and hepatoprotective activity in cases of toxicity induced by drugs and chemicals. It is also effective in the treatment of cirrhosis of the liver.

MATERIALS AND METHODS

CHAPTER3

MATERIALS AND METHODS

3.1 MATERIALS

Young leaves were collected from *Solanum sisymbriifolium* Lamk., *Solanum khasianum* Clark. (presently known as *Solanum aculeatissimum*), *Solanum clavatum* Rusby., *Solanum nigrum* L., *Solanum kurzii* Prain., *Solanum gilo* Dunal. (presently known as *Solanum aethiopicum*) and *Solanum torvum* Sw (Fig.3.2-3.4), some wild species of genus *Solanum* found in Meghalaya. Samples were collected from NEHU campus, Shillong town, Smit, Umiam, Upper Shillong, Shora and Sohryngkham (Fig. 3.1). Five replicates for each species were taken. Collected leaves were washed with water and preserved at -80° C for further use. Voucher specimens have been lodged in the Herbarium of Botany Department, North Eastern Hill University (Table 3.1).

3.2 METHODS

3.2.1 DNA EXTRACTION FROM LEAVES

Total genomic DNA was extracted following the method of Doyle and Doyle (1987) with minor modifications. The protocol is given below:

- The leaf samples were finely crushed in Liquid Nitrogen.
- The pulverized plane material was transferred to a 1.5 mL microcentrifuge tube and to 0.2 mg of the sample 800μL of preheated extraction buffer

Fig. 3.1 Physical map of India showing the location of Meghalaya and also the locations of sample collection sites.

(a)

(b)

(c)

Fig. 3.2 Study samples in natural habitat. a- *S. sisymbriifolium*, b- *S. khasianum*, c- *S. clavatum*.

(d)

(e)

(f)

Fig. 3.3 Study samples in natural habitat. d- *S. nigrum*, e- *S. kurzii*, f- *S. gilo*.

(g)

Fig. 3.4 Study samples in natural habitat. g- *S. torvum*.

Species	Specimen voucher[a]	Gene Bank Accession No.			
		18S rRNA	ITS	*rbcL*	*matK*
Solanum khasianum Clarke	NEHU-11928	KC535785	KC535792	KC535806	KC535798
Solanum sisymbriifolium Lamk.	NEHU-11929	KC535786	KC535789	KC535805	KC535799
Solanum gilo Req. ex Dunal	NEHU-11931	KC535787	KC535795	KC535808	KC535801
Solanum kurzii Brace ex Prain	NEHU-11934	KC535784	KC535794	KC535807	KC535800
Solanum clavatum Rusby	NEHU-11933	KC535783	KC535790	KC535804	KC535796
Solanum nigrum L.	NEHU-11930	KC535782	KC535791	KC535803	KC535797
Solanum torvum Sw.	NEHU-11932	KC535788	KC535793	KC535809	KC535802
Solanum tuberosum L.	-	X67238	-	DQ231562*	DQ231562*
Solanum lycopersicum L.	-	X51576	AJ300201	AM087200*	AM087200*
Nicotiana tabacum L.	-	AJ236016	AJ02364	Z00044*	Z00044*

Sequences have been obtained from *Complete chloroplast genome: *S. tuberosum, S. lycopersicum, Nicotiana tabacum.* [a]Specimen vouchers deposited at the Herbarium, Department of Botany, North Eastern Hill University.

Table 3.1: Specimens voucher and accession numbers of deposited sequences of all the seven *Solanum* spp. Sequences retrieved from GenBank data base are also listed.

(2%CTAB, 1%PVP, 100mM Tris-HCl, pH8.0, 1.4M NaCl, 20mM EDTA pH8.0) was added.

- To this solution 2 μL of β-meracaptoethanol was added.
- The solution was vortexes and incubated at 65℃ for an hour in a water bath.
- An equal volume of chloroform-isoamyl alcohol (24:1) was added and mixed by inversion to form an emulsion.
- The extract was centrifuged for 10 minutes at 13,000 rpm (16060g) and the aqueous supernatant transferred to a new microtube using a wide-bore pipette tip.
- The chloroform-isoamyl alcohol (24:1) extraction was repeated once more.
- To the final aqueous solution recovered, 1/3 volume of ice cold Isopropanol were added. A DNA pellet became visible upon gentle swirling. The solution was left overnight to precipitate at -20℃.
- Next morning, DNA was pelleted by centrifugation at 13,000 rpm for 15 minutes at 4^0C and pellet was washed twice with 70% ethanol (v/v) twice for 30 minutes each.
- The DNA was dried using vacuum, and dissolved in 50μL of 10X TE buffer (10mM Tris-HCl, 1mM EDTA, pH8.4) for an hour. The DNA from ten microtubes was pooled into one tube. 3μL of RNase (10mg/mL) was added and the DNA was incubated at 37^0C for 1 hour. After the RNase treatment, 3μL of proteinaseK (1mg/mL) was added and the DNA was incubated at 37^0C in the water bath for approximately 30 minutes.
- An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the sample and spinned at 13,000 rpm for 15 minutes.

- The supernatant was taken and chloroform-isoamyl alcohol (24:1) was added and spinned at 13000 rpm for 15 mins.
- Precipitation and washing was done in similar manner.
- Pellets were lyophilized using vacuum drier. To the dried pellet 100µL 10X TE buffer (pH 8.4) was added and stored at -20^{0}C for further use.

Quality of extracted DNA was assessed both by the agarose gel electrophoresis and spectrophotometrically using absorbance ratio of A260 and A280.

3.2.2 TARGET DNA SEGMENTS AND PCR PRIMER DESIGNING

Primer pairs were designed targeting nuclear *18SrRNA* gene, *rrn* ITS region and chloroplast *rbcL* and *matK* genes using DS GENE version 1.1 software (Table 3.2). The primers were designed from the consensus sequences of respective genes/ segments of related species available in GenBank database. Internal primers were designed for *18SrRNA, rbcL* and *matK* genes for getting complete sequences of the respective genes.

3.2.3 DNA AMPLIFICATION REACTIONS

DNA amplification reactions for the target regions were carried out using the thermal cycler (ABI GeneAmp 9700 Gold).

The annealing temperatures for the primers were roughly calculated using the following formula:

Annealing temperature= [4 (G+C) + 2(A+T)] – 5

Primer name	Sequence
18SF	5'ACTACTCTGGATAACCGTAGTC3′
18SR	5'ATAAGGTTCAATGGACTTCTCG3'
18SF790	5'TGAAAGACGAACAACTGCG3'
18SR832	5'GAAAACATCCTTGGCAAATGC3'
SITSF	5'AAACCTGCACAGCAGAACGAC3'
SITSR	5'GGTCGCGGTCGGAGCGCG 3'
SrbcLF	5'ATGTCACCACAAACAGAG3'
SrbcLR	5'TACGATCTCTTTCCATACC3'
SrbcF679	5'AAAGCACAGGCTGAAACAG3'
SrbcR791	5'ATCAAAACGCCCAATTCTC3'
SmtKF	5'CACAAACTAGACGAAGCTC3'
SmtKR	5'TATGCACTTGCTCAGGATC3'
Smtk1	5' AATATATTTCTATGGAAAAAG3'
Smtk2	5' ATCAAAGGATCCTTGAATAAC3'

Table 3.2: The primers designed and used in this study and their sequences.

However the annealing temperatures were modified (increased or decreased) depending upon the results. At the initial stage, the amplifications were carried out at lower temperature and then the annealing temperatures were gradually increased to avoid the nonspecific amplification. Only those amplification temperatures were considered as optimized for a particular pair of primers (Table 3.2) when a single amplification band of expected size was formed. The following protocols were used for amplification of respective target regions.

3.2.3.1 Amplification of *18S rRNA*

Amplification of *18S rRNA* gene was carried out using the primer pairs 18SF and 18SR. Each reaction mix contained 2.5 µL of 10X PCR assay buffer, 2.5 µL of $MgCl_2$ (25mM), 3 µL each of the individual dNTP (1.25 mM), 1.5 µL of each primer pair (5 pM), 0.3 µL of *Taq* polymerase (3 Units/ µL), and the final volume was made up to 25 µL by adding ultra pure water. Polymerase chain reaction was carried out targeting the genes/ segment using the respective primer pairs. The thermalcycler was programmed with the following parameters: premelt at 94^0C for 5 min followed by 35 cycles consisting of denaturation at 94^0C for 30 sec, annealing at 55^0C for 2 mins, and extension at 72^0C for 1 min. The main programme was followed by a final extension step of 72^0C for 10 mins. Internal primers 18SF790 and 18SR832 were used for sequencing purpose.

3.2.3.2 Amplification of ribosomal *ITS*

Amplification of *ITS* region was carried out by using the primer pairs SITSF and SITSR. Reaction mixture was prepared as mentioned in 3.2.3.1.

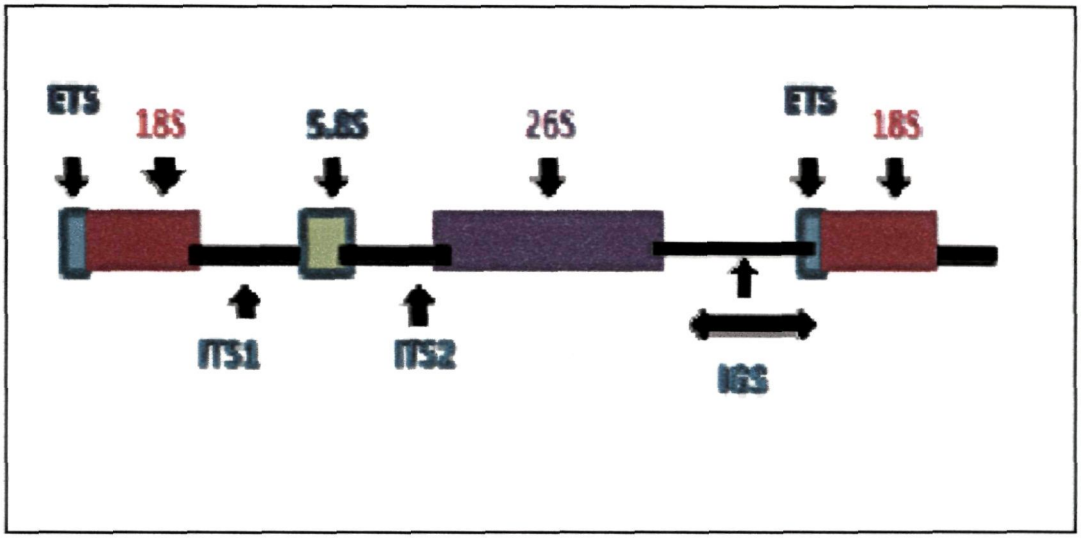


Fig. 3.5 Schematic diagram of rDNA repeat of plants showing 18S rDNA and ITS region.

Polymerase chain reaction was carried out targeting the genes using the respective primer pair. The thermalcycler was programmed with the following parameters: premelt at 94 ^{0}C for 5 mins, 35 cycles each consisting of a denaturation step at 94 ^{0}C for 1min, annealing step at 65 ^{0}C for 1 min, and an extension step at 72 ^{0}C for 1 min, followed by a final extension step of 72 ^{0}C for 10 mins. The same primers were used for sequencing.

3.2.3.3 Amplification of *rbcL*

Amplification of *rbcL* gene was carried out by using the primer pairs S*rbcL*F and S*rbcL*R. Reaction mixture was prepared as mentioned in 3.2.3.1. Polymerase chain reaction was carried out targeting the gene using the respective primer pair. The thermalcycler was programmed with the following parameters: premelt at 94 ^{0}C for 5 min, 35 cycles each consisting of a denaturation step at 94 ^{0}C for 1 min, annealing step at 53 ^{0}C for 1 mins, and an extension step at 72 ^{0}C for 1 min, followed by a final extension step of 72 ^{0}C for 10 mins. Internal primers SrbF679 and SrbR791 were used for sequencing purpose.

3.2.3.4 Amplification of *matK*

Amplification of *matK* gene was carried out by using the primer pairs SmtkF and SmtkR. Reaction mixture was prepared as mentioned in 3.2.3.1. Polymerase chain reaction was carried out targeting the genes using the respective primer pair. The thermalcycler was programmed with the following parameters: premelt at 94 ^{0}C for 5 min, 35 cycles each consisting of a denaturation step at 94 ^{0}C for 1 min, annealing step at 64 ^{0}C for 1 mins, and an extension step at 72 ^{0}C for 1

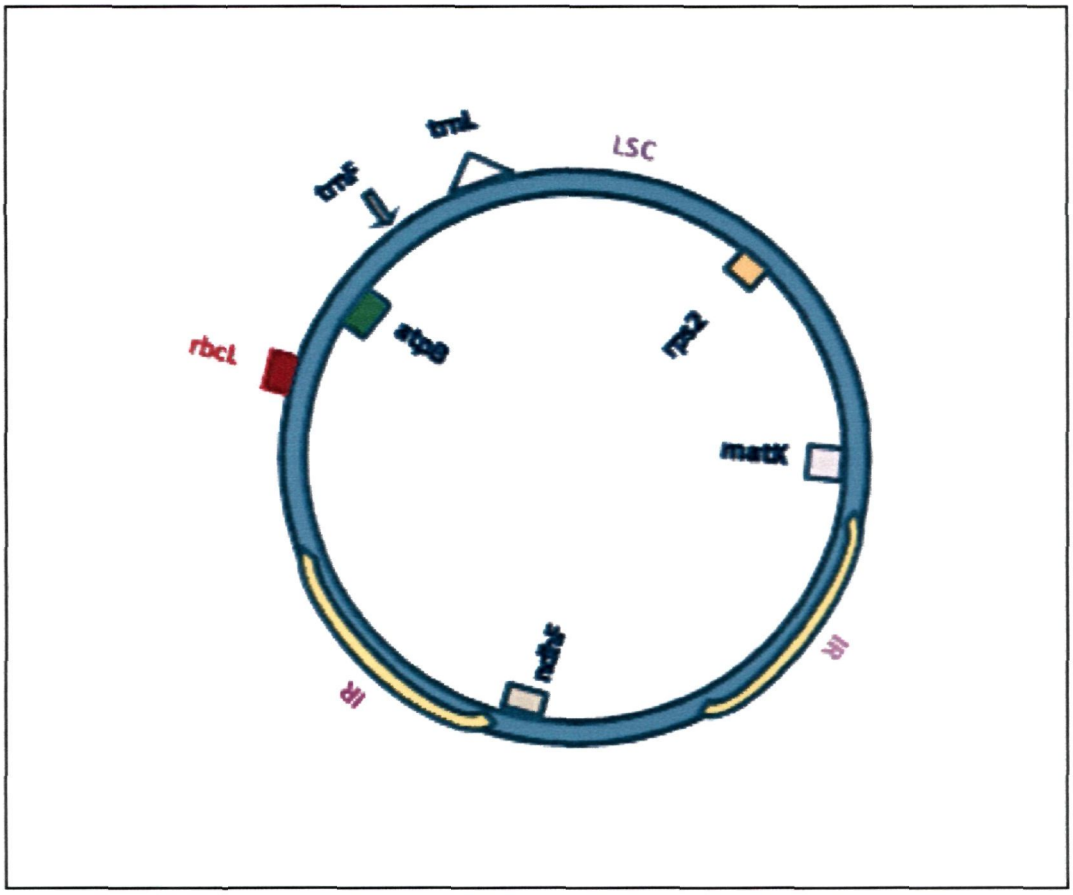


Fig. 3.6 Schematic diagram of chloroplast DNA showing *rbcL* and *matK* genes.

min, followed by a final extension step of 72 ^{0}C for 10 min. Internal primers Smtk1 and Smtk2 were used for sequencing purpose.

3.2.4 NUCLEOTIDE SEQUENCING

For sequencing the PCR products were purified using the following protocol:

3.2.4.1 Purification of PCR products:

- 5 µL of 3M Sodium acetate (pH 4.6) and 100 µL of 95% ethanol were added to each 50µL PCR reaction mixture.
- The tubes were vortexed and kept at -20^{0}C for 40 minutes to precipitate the PCR products.
- The solutions were then centrifuged for 20 minutes at 13000 rpm.
- After the spin the supernatant was pipetted out carefully and discarded.
- The pellet was rinsed with 300 µL of 70% ethanol and centrifuged for 5 minutes at 13000 rpm.
- The supernatant was pipetted out carefully and discarded.
- The pellet was vacuum dried and resuspended in 50 µL of water.

3.2.4.2 Cycle sequencing:

For sequencing the purified products were subjected to cycle sequencing. Only one primer was used in each cycle sequencing reaction so that the amplification of product was linear not exponential.

The chemicals listed below were mixed in a 0.2µL tube for cycle sequencing:

- Terminator Ready Reaction mix:

  Distilled water: 3.5μL

  5X buffer: 3.5μL

  Big dye: 1μL

- Cleaned PCR product: 150-300ng PCR product DNA.
- Primer 3.2-5pmole.

The final volume was made upto 20μL with MilliQ water.

Cycle sequencing was carried out with an initial denaturation at 96^0C for 5 minutes, followed by 25 cycles for 10 seconds at 96^0C, 10 seconds at 53^0C-63^0C (depending upon the annealing temperature of a particular primer), and for 4 minutes at 65^0C. No extra extension time was given.

3.2.4.3 Purification of the cycle sequencing product:

The cycle sequencing products were purified using the following protocol:

1. 2 μL of 125mM EDTA was added to each reaction tube and mixed well.
2. Then 2 μL of 3M sodium acetate (pH 4.6) was added.
3. The entire solution was transferred to a 0.5mL tube containing 50 μL 95% ethanol, mixed thoroughly, and kept in dark for 15 minutes incubation at room temperature.
4. Then the tubes were centrifuged at 12000 rpm for 20 minutes at room temperature.
5. The supernatant was discarded carefully so that pellet was not disturbed.
6. 250 μL of 70% ethanol was added to each tube and mixed gently.
7. The tubes were kept for 10 minutes in dark.

8. The tubes were centrifuged at 13000 rpm for 10 minutes at room temperature.
9. The supernatant was discarded carefully.
10. Step 6, 7, 8 were repeated.
11. The pellets were vacuum dried for 5 minutes.
12. These samples were stored at -20^{0}C for further use.

3.2.4.4 Loading the samples:

Sequencing was carried out in 3130 Genetic Analyzer (Applied Biosystems) at NEHU. Just before sequencing, 10μL of Hi-Di formamide was added to each tube. The samples were heated at 95^{0}C for 2 min and then the samples were loaded for sequencing.

3.2.5 RESTRICTION DIGESTION STUDIES

Amplicons were purified and then subjected to restriction digestion as per the conditions provided by the manufacturer for each restriction enzyme (Table 3.3).

3.2.5.1 Single digestion conditions:

The restriction digestion mixture was prepared in a 0.5 mL tube in which 10μL of the amplicon was digested with 5 units of the restriction enzyme in the appropriate buffer solution (2μL). Ultra pure distilled water was added to make the final volume to 20μL. The incubation temperatures for all the endonucleases used were 37^{0}C and were incubated overnight.

Sl. No.	Restriction enzyme	Cutting site	Source organism	Incubation temparature
1	*Nci*I	CC↓SGG S=C/G	*Neisseria cinerea*	37^{0}C
2	*Hpa*II	C↓CGG	*Haemophilus parainfluenza*	37^{0}C
3	*Dde*I	C↓TNAG N=A/T/C/G	*Bacillus stearothermophilus* DE	37^{0}C
4	*Scr*FI	CC↓NGG N=A/T/G/C	*Bacillus stearothermophilus* SC	37^{0}C
5	*Ava*II	G↓GWCC W=A/T	*Anabaena variabilis*	37^{0}C

Table 3.3: List of restriction enzymes used.

3.2.5.2 Double digestion conditions:

In case of double digestion, 10 μL of amplicons was digested with-

- Buffer H (10X): 2 μL
- Enzyme 1 (10U/μL): 1 μL
- Enzyme 2 (10U/μL): 1 μL
- Ultrapure water was added to make the volume 20 μL

Here Buffer H was taken as activity of both the enzymes *(Dde*I and *ScrF*I) were 100% in buffer H. The incubation temperature was 37^0C and the sample mixture was incubated overnight.

3.2.6 AGAROSE GEL ELECTROPHORESIS (AGE)

Genomic DNA, PCR products and digested fragments were separated on ethidium bromide (0.5μg/mL) stained agarose gel using horizontal electrophoresis in 1X Tris-borate (TBE) (5X TBE is 0.089M Tris-borate, 0.089M boric acid and 0.002M EDTA) buffer. Genomic DNA, PCR amplicons and restriction fragments were visualized on 0.8%, 1.5%, and 3% agarose gels respectively. Bromophenol blue (6X) (0.25% Bromophenol blue, 0.25% xylene cyanol and 30% glycerol in distilled water) was used as the gel loading dye.

The gels were then scanned and photographed using KODAK Gel Logic 1500 Imaging System and the profiles were analysed using the Kodak Molecular Imaging Software.

3.2.7 SEQUENCE ALIGNMENT

Sequences of all the four regions were subjected to Multiple Sequence Alignment together with the sequences retrieved from the Gene Bank using the CLUSTAL X (2.0) program (Thompson *et al.*, 1997) with default settings. Sequences from each genomic region were aligned separately and thus separate alignment matrices were generated for all the four regions. Clustal X generated alignments were further trimmed manually. A consensus sequence for each data set was constructed using Geneious Basic 5.3.6 (Drummond *et al.* 2010). Apart from the four separate data sets two other data sets were also aligned – one data matrix of nuclear 18S rDNA + ITS and another data matrix composed of chloroplast *rbcL* + *matK* . The data sets were saved as nexus files. Alignments of all genomic regions were then combined to a single nexus file.

3.2.8 SEQUENCE CHARACTERISTICS

Sequence characteristics of all genomic regions were calculated before performing phylogenetic analyses by using both MEGA version 5 (Tamura *et al.*, 2011) and Seqstate v.1.21 (Müller, 2005).

3.2.9 PHYLOGENETIC ANALYSES

The phylogenetic relationship were estimated by using both Neighbor joining method conducted in MEGA version5 (Tamura *et al.*, 2011) and Bayesian analysis by using MRBAYES v.3.1.2 (Ronquist and Huelsenbeck, 2003). The strong overall similarities of the topologies based on analyses of the individual genes and in combinations have been discussed. Phylogenetic trees were constructed with individual data sets, i.e. 18S rDNA , ITS, *rbcL, matK* using Neighbor joining method

conducted in MEGA version5 (Tamura *et al.*, 2011). Construction of nuclear and chloroplast trees and their tree topology comparison are very important as it can describe the role of hybridization in the phylogenetic history. By keeping that in mind two separate trees were constructed utilizing the nuclear data sets (18S rDNA + ITS) and chloroplast data sets (*rbcL* + *matK*) by conducting Bayesian analysis by using MRBAYES v.3.1.2 (Ronquist and Huelsenbeck, 2003). NJ analysis by using MEGA version5 (Tamura *et al.*, 2011). There might have issues of incongruence due to which a tree was also constructed by using all the four data sets in combination so that a fully resolved tree could be obtained with more explicit relationships by using MRBAYES v.3.1.2 (Ronquist and Huelsenbeck, 2003). Gaps were treated as missing data. The strength of support for individual tree branches was estimated using bootstrap values (BS) (Felsenstein, 1985) in case of NJ tree and posterior probability values in case of consensus trees generated by Bayesian analysis. Bootstrap analysis was carried out for 1000 replicates to examine the relative level of support for individual clades on the cladograms of each search.

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal trees with the sum of branch lengths were generated. The trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980) and were in the units of the number of base substitutions per sites.

3.2.10 SELECTION OF SUBSTITUTION MODEL AND BAYESIAN ANALYSES

Felsenstein (1988) stated that all phylogenetic methods make assumptions, whether explicit or implicit, about the process of DNA substitution. As all the methods of phylogenetic inference depend on the substitution models, it is always necessary to use the particular model of substitution which best fits the data sets. So appropriate models of substitutions were selected for the three data sets (nuclear, chloroplast and combined) using the jModelTest 0.1 (Pasoda, 2008). There are 88 models currently implemented in jModelTest, including 11 substitution schemes, equal or unequal base frequencies (+F), a proportion of invariable sites (+I) and rate variation among sites with a number of rate categories (+G). For all the three data sets Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were analyzed which compares several candidate models simultaneously. So there could be two models selected for each data set but the final model was selected on the basis of higher free parameters. The selected model for nuclear, chloroplast and combined data sets were TrN+G (Tamura-Nei plus Gamma), TVM+G (Transversion Model plus Gamma) and GTR+G (General Time Reversible plus Gamma) respectively (Table 3.4).

Besides NJ analysis, Bayesian inference (BI) of phylogeny was conducted for the above mentioned three dataset using MRBAYES v.3.1.2 (Ronquist and Huelsenbeck, 2003) utilising each selected model as mentioned above. BI analyses were performed for 1,000,000 generations applying the default settings (MCMC, two runs with four chains each, heating temperature 0.2, saving one tree

Regions	Model	Nucleotide frequencies				Substitution Rate Matrix					
		A	T	G	C	A-C	A-G	A-T	C-G	C-T	G-T
18SrRNA+ITS+rbcL+matK	GTR+G	0.2597	0.2149	0.2374	0.2879	2.0696	2.5181	0.9294	1.5903	3.9687	1.0000
rbcL+matK	TVM+G	0.2954	0.1904	0.1800	0.3341	1.7112	2.0312	0.5562	1.4020	2.0032	1.0000
18SrRNA+ITS	TrN+G	0.2323	0.2627	0.2876	0.2174	0.9373	1.5169	0.8920	0.8447	5.4881	1.0000

Table 3.4: Parameters of Maximum Likelihood analysis and selected models.

every 100 generations). All trees were viewed with the program FigTree v1.3.1 (Rambaut, 2009).

3.2.11 CLUSTER ANALYSIS

Cluster analysis is an exploratory data analysis tool for solving classification problems. The objective of cluster analysis is to assign observation to groups or clusters so that observations within each group are similar to one another with respect to variables. So within a cluster analysis result there are a number of heterogenous groups with homogenous contents. Thus cluster analysis sort cases into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Cluster analysis can either be agglomerative or divisive. Agglomerative method begins with as many clusters as there are observations and end with a single cluster containing all observations, whereas divisive method begins with a single cluster and ending with as many clusters as there are observations.

In the present study the banding patterns in ARP analysis and protein profiles were used to generate cluster dendrogram with the method SHAN (Sequential Agglomerative Hierarchical Nested cluster analysis) using NTSys software (2.1 version). That was based on Jaccard's coefficient of similarity (Jaccard, 1908). Jaccard coefficient is based on an algorithm that is used to estimate the level of similarity for multivariate data. Presence or absence of bands for each individual sample was scored for all the individual samples so that a descending pattern of values was obtained.

The Jaccard's coefficient value for each sample was calculated using the formula given below:

Jaccard's coefficient = Nab/[(Na+Nb)-Nab]

Where, Nab=Number of common bands in both samples

Na= Total number of bands present in the first sample

Nb= Total number of bands present in the second sample

3.2.12 TOTAL PROTEIN PROFILING

3.2.12.1 Total protein extraction

Total protein was extracted by using the following protocol-

1. 0.5-1.0 gm leaves were ground with liquid N_2 by using mortar and pestle.
2. When the ground leaves became fine powder, these were equally distributed in eppendorf tubes.
3. Prepared buffer (50mM Tris-Cl pH 6.8, 100mM NaCl, 10mM EDTA, 100mM Glycine) and 10% SDS were added.
4. The solutions were mixed carefully and kept in ice for 45 minutes.
5. Tubes were centrifuged at 10,000 rpm for 15 minutes at 4^0C.
6. The supernatants were collected carefully.
7. Repeated steps 5, 6.
8. The supernatant contained the protein.

3.2.12.2 Estimation of total protein

3.2.12.2.1 Preparation of Bradford's reagent

100mg of Commassie Brilliant Blue G-250 was taken in 50mL 95% ethanol and dissolved properly in a 250mL conical flask. To this 85% (w/v)

phosphoric acid was added to make up the volume (250mL). It was mixed properly and transferred to a dark bottle with tight cap. It was then stored in refrigerator for further use.

3.2.12.2.2 Preparation of working solution

15mL of Bradford's reagent (stock) was taken with 85mL distilled water and mixed thoroughly. It was filtered using whatman filter paper. The filtrate was stored in a glass reagent bottle at room temperature.

3.2.12.2.3 Assay of protein

1mg/mL BSA stock was prepared as Standard (BSA).

- 0.01, 0.02, 0.04, 0.06, 0.08 and 0.1mL of BSA stock was taken in different tubes for preparing standard curve.
- The protein extracted as above from all the samples and BSA solutions as above were then processed by making final volume to 0.1mL by adding requisite amount of water.
- To each tube 5mL of Bradford's working solution was added and vortex gently.
- The tubes were incubated at 25°C for 10 minutes.
- Absorbance was taken at 595nm between 10 minutes and 45 minutes.
- Unknown samples were compared with the standard BSA curve.

3.2.12.3 SDS-PAGE (SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS)

Total proteins were separated in a 12% SDS Polyacrylamide Gel Electrophoresis (PAGE). Resolving gels at 12% were prepared with 12 mL of 30% acrylamide, 7.5 mL of 1.5 M Tris pH 8.8, 0.3mL of 10% SDS, 0.3mL of 10%

ammonium persulfate, 0.012mL of TEMED and water until 30 mL. The 5% stacking gels were prepared with 0.83mL of 30% acrylamide, 0.63mL of 1.0 M Tris pH 6.8, 0.05mL of 10% SDS, 0.05mL of 10% ammonium persulfate, 0.005mL of TEMED and water until 5mL. A volume of 15 µL containing 40 µg of total proteins was mixed with an equal volume of 2X sample buffer (0.8 mL of 0.5 M tris-HCl pH 6.8, 2 mL glycerol, 0.2 mL of 0.01% bromophenol blue, final volume of 10 mL). The samples were heated at 100 °C for 5 minutes before loading. A 5X Tris-Glycine-SDS solution was prepared (94 g Glycine, 15.1 g Tris, 50mL 10% SDS for 1 litre, pH 8.8). It was diluted 1:5 to obtain the 1X running buffer. The electrophoretic run was at 100 volts for ~1 hour. The gel was stained by staining solution (0.25gm Coomasie blue dye, 40mL Methanol, 7mL Glacial Acetic acid and made the volume 100mL with distilled water) for overnight. The gel was destained with destaining solution (40mL Methanol, 7mL Glacial Acetic acid and made the volume 100mL with distilled water) till the bands were visible and distinct.

3.2.13. PHYTOCHEMICAL SCREENING

Fruits were sun-dried, pulverized and passed through a sieve (about 0.5 mm pore size) to obtain a fine dry powder. Aqueous extract of the sample was prepared by soaking 100 g of the powdered samples in 200 ml of distilled water for 12 hours. The extracts were filtered using Whatman filter paper No 42 (125 mm). Chemical tests were carried out on the aqueous extract and on the powdered samples to identify the constituents using standard procedures (Harborne, 1973; Sofowara, 1993; Trease and Evan, 1989). Colour intensity was used to categorize the presence of each phytochemical into copious, moderate or slight (trace).

3.2.13.1. Alkaloid test:

About 0.2 gm of the extract was warmed with 2% H_2SO_4 for two minutes. It was filtered and a few drops of Dragondorff reagent were added. Orange red precipitate indicated the presence of alkaloids.

3.2.13.2 Terpenoid test (Salkowski test):

5 mL of the extract was mixed in 2 mL of chloroform, and 3 mL concentrated H_2SO_4 was carefully added to form a layer. A reddish brown coloration of the inter face was formed to show positive results for the presence of terpenoids.

3.2.13.3 Flavonoid test:

The presence of flavonoids in the plant sample was determined by the methods described by Sofowara (1993) and Harborne (1973). 5 ml of dilute ammonia solution was added to a portion of the aqueous filtrate of the plant extract followed by addition of concentrated H_2SO_4. A yellow coloration observed in each extract indicated the presence of flavonoids. The yellow coloration disappeared on standing. Few drops of 1% aluminum solution were added to a portion of each filtrate. A yellow coloration was observed indicating the presence of flavonoids. A portion of the powdered plant sample was in each case heated with 10 mL of ethyl acetate over a steam bath for 3 min. The mixture was filtered and 4 mL of the filtrate was shaken with 1 mL of dilute ammonia solution. A yellow coloration is positive for flavonoids.

3.2.13.4 Cardiac glycosides test (Keller-Killani test):

5 ml of the extract was treated with 2 ml of glacial acetic acid containing one drop of ferric chloride solution. This was underlayed with 1 ml of concentrated sulphuric acid. A brown ring of the interface indicated a deoxysugar

characteristic of cardenolides. A violet ring may appear below the brown ring, while in the acetic acid layer, a greenish ring may form just gradually throughout thin layer.

3.2.13.5 Tannin test:

About 0.5 gm of the dried powdered samples was boiled in 20 mL of water in a test tube and then filtered. A few drops of 0.1% ferric chloride were added and observed for brownish green or a blue-black coloration.

3.2.13.6 Saponin test:

About 0.2 gm of the extract was shaken with 5 ml of distilled water and then heated to boil. Frothing shows the presence of saponin.

3.2.14 EXTRACTION OF TOTAL ALKALOID

Total alkaloid was extracted using Harborne's (1973) protocol .

- Fruits were oven dried and then ground in powdery form.
- 5 gm of fruit powder was mixed with 500 ml of methanol:water (4:1) and kept under stirring for 36 hours.
- The mixture was filtered using Whatman filter paper No 42 (125 mm) and the filtrate was evaporated to $1/10^{th}$ of volume.
- The filtrate was then acidified with 2M H_2SO_4 (pH 4.7).
- Then basified the extract again with NH_4 (pH 10).
- Then the extract was filtered.

- Filtrate was extracted thrice with Chloroform and the chloroform layer was collected.
- Evaporated the chloroform extract and used for TLC.
- Detection of alkaloids were carried out by using Dragendorff's reagent.

3.2.15 QUANTIFICATION OF SOLASODINE

3.2.15.1 Extraction and purification of solasodine

The method for the extraction and purification of solasodine was based on that of Lancaster and Mann (1975) as follows-

1. One g portions of dried, powdered tissue were gently shaken for 30 min in 50 mL of 2% aq. Oxalic acid in an automatic tilting apparatus.
2. Suspensions were vacuum filtered through Whatman filter paper with only the first 10-20 mL of clear extract being collected. Tubes were gently heated to 70^0C, 1 mL of 60% NaOH was added, and after heating for a further 5 min and the tubes were kept at room temperature overnight.
3. Tubes were centrifuged at 20^0C for 10min at 3000g. The supernatants were decanted and the pellets dissolved in 5 ml of 0.5 M HCl and hydrolysed by refluxing at 100^0C for 90 min.
4. Samples were cooled briefly and made alkaline with 1 ml 60% NaOH. Heating 100^0C was resumed for 10 min to complete the formation of the insoluble solasodine base.

3.2.15.2 Colorimetric determination of solasodine

Quantification of Solasodine was carried out using the following colorimetric method:

1. Solasodine was dissolved in 10 ml of alcohol free $CHCl_3$ solutions (0.2-2.0 mL) were transferred to 10 mL capped graduated polythene centrifuge tubes and made to 5.0 mL with alcohol free $CHCl_3$.

2. A 2.5mL aliquote of 200 μM bromothymol blue in borate buffer pH 8.0 was added to each tube and mixed for about 10 sec with a vortex stirrer. Formation of an emulsion was carefully avoided.

3. The upper aq. Phase was removed, and the lower $CHCl_3$ phase adjusted to 4.0 mL by the removal of excess $CHCl_3$. (Any droplets of aq. Dye still adhering to the walls of the tube were removed by swabbing with a cotton wool bud).

4. The dye which had been carried into the $CHCl_3$ solution as a complex with solasodine, was then converted to more strongly colored ion by adding 0.1 mL of methanolic 0.01 M NaOH to the $CHCl_3$.

5. The absorbance of solution was determined at 610 nm.

6. A standard curve was prepared using stock solution of 10 μg/mL, 20 μg/mL, 30 μg/mL, 40 μg/mL, 50 μg/mL, 60 μg/mL, 70 μg/mL, 80 μg/mL, 90 μg/mL, and 100 μg/mL pure solasodine in alcohol-free $CHCl_3$. 5 mL of chloroform solution and 2.5 mL dye buffer solution were mixed and treated as above.

RESULTS AND DISCUSSION

CHAPTER 4

RESULTS AND DISCUSSION

4.1 COLLECTION OF LEAVES

Young leaves of wild *Solanum khasianum* (presently known as *Solanum aculeatissimum*), *Solanum sisymbriifolium*, *Solanum torvum*, *Solanum kurzii*, *Solanum clavatum*, *Solanum nigrum* and *Solanum gilo* (presently known as *Solanum aethiopicum*) were collected from five different plants of each species representing different parts of Meghalaya i.e. NEHU campus, Shillong town, Smit, Umiam, Upper Shillong, Shora and Sohryngkham.

4.2 DNA EXTRACTION

The extracted DNA was monitored by running Agarose gel electrophoresis (AGE). All samples showed DNA bands of high molecular weight, though in some samples RNA was also seen. RNA was removed by treating the DNA samples with *RNase*. Agarose gel photographs of DNA of all the samples are shown in Fig. 4.1-4.2. Purity of the isolated DNA was assessed both by the Agarose gel electrophoresis (AGE) and spectrophotometrically using absorbance ratio of A260/A280. The A260/A280 ratio ranged between 1.8 to 2.0, indicating that the isolated DNA of the present study were pure and could be further used for subsequent amplification reactions.

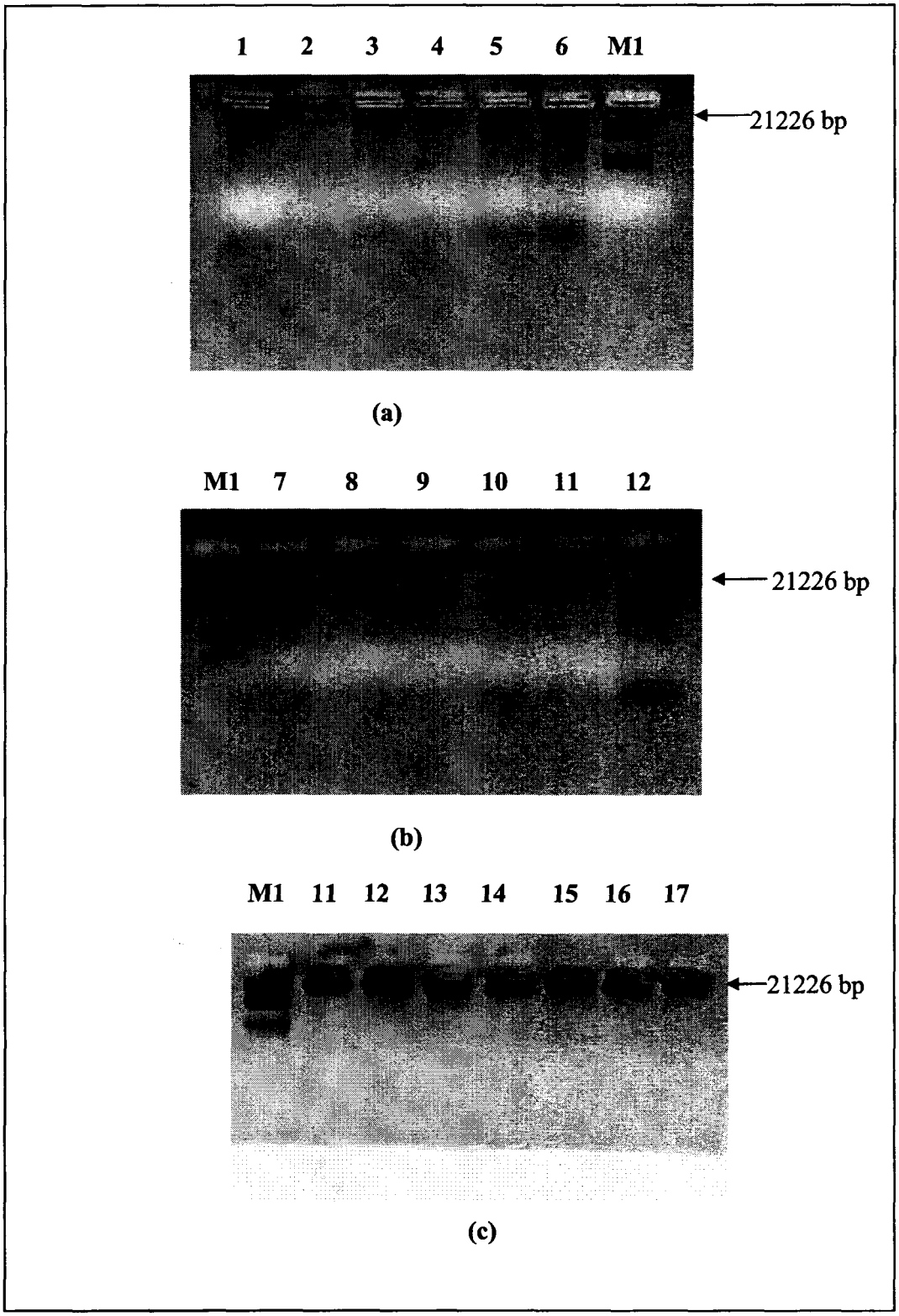


Fig. 4.1 (a-c): Agarose gel photograph of genomic DNA of *S. nigrum* (1-5), *S. clavatum* (6-10), *S. khasianum* (11-15), *S. gilo* (16-17). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

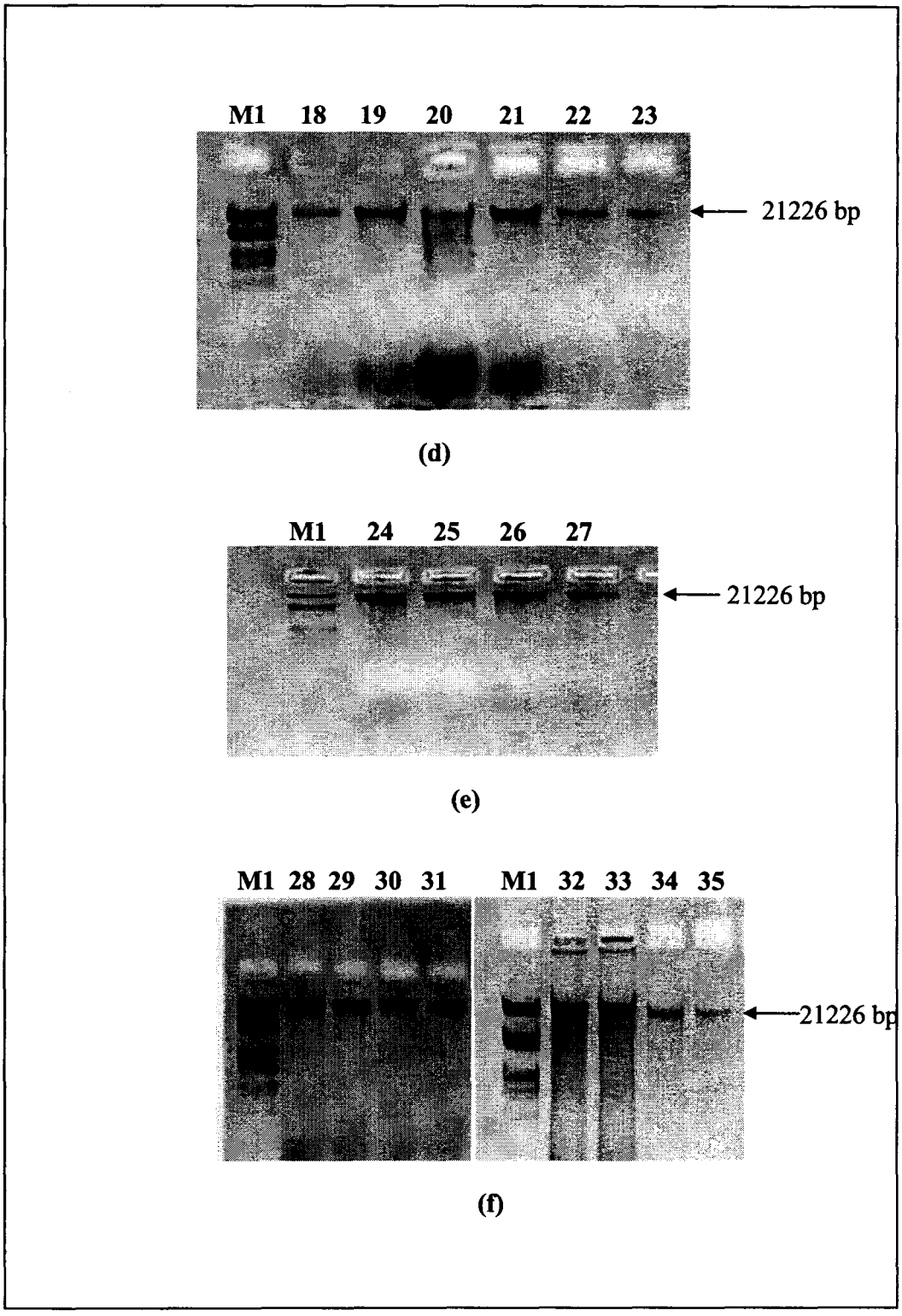


Fig. 4.2 (d-f): Agarose gel photograph of genomic DNA of *S. gilo* (18-20), *S. sisymbriifolium* (21-25), *S. torvum* (26-30), *S. kurzii* (31-35). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

4.3 PCR AMPLIFICATION

4.3.1 AMPLIFICATION OF *18S rRNA* GENE

Amplification of *18S rRNA* gene was carried out by using the primer pairs 18SF and 18SR (Table 3.2). The PCR reaction was carried out as described in section 3.2. The amplicons were run in 1% agarose gel. Amplification was tried at lower annealing temperatures which yielded multiple bands including the band of interest. The appearance of unwanted bands was due to the result of non-specific binding of primers at sites other than the targeted region. The best amplification with single band was obtained at annealing temperature of 55^0C. The amplification yielded a single band of size of approximately1.6Kb (Fig. 4.3-4.4). All the samples included in this study showed the same PCR profile in AGE.

4.3.2. AMPLIFICATION OF *rrn* ITS

The primers used for this amplification were SITSF and SITSR. During the initial optimization a number of annealing temperatures were tried to obtain the single band of expected size. While trying amplifications in lower temperatures multiple bands were generated due to non-specific binding of primers at sites other than the targeted site. The annealing temperature was optimized at 65^0C. The amplicons were observed in 1.5% agarose gel electrophoresis. Single fragments were observed for each species. The bands observed were approximately 600 bp (Fig. 4.5-4.6).

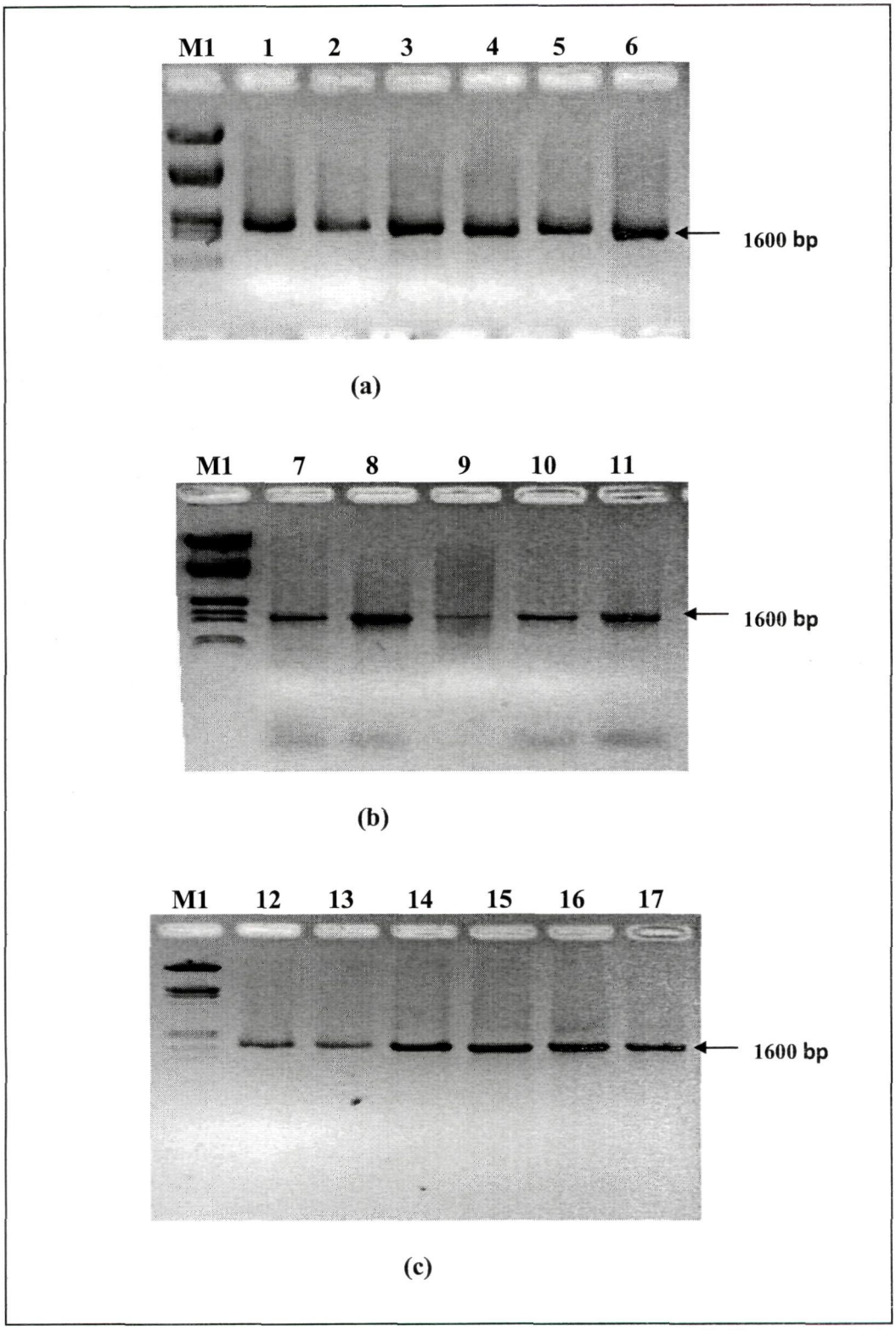


Fig. 4.3 (a-c) Amplification of the 18S rDNA region of *S. nigrum* (1-5), *S. clavatum* (6-10), *S. khasianum* (11-15), *S. gilo* (16-17). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

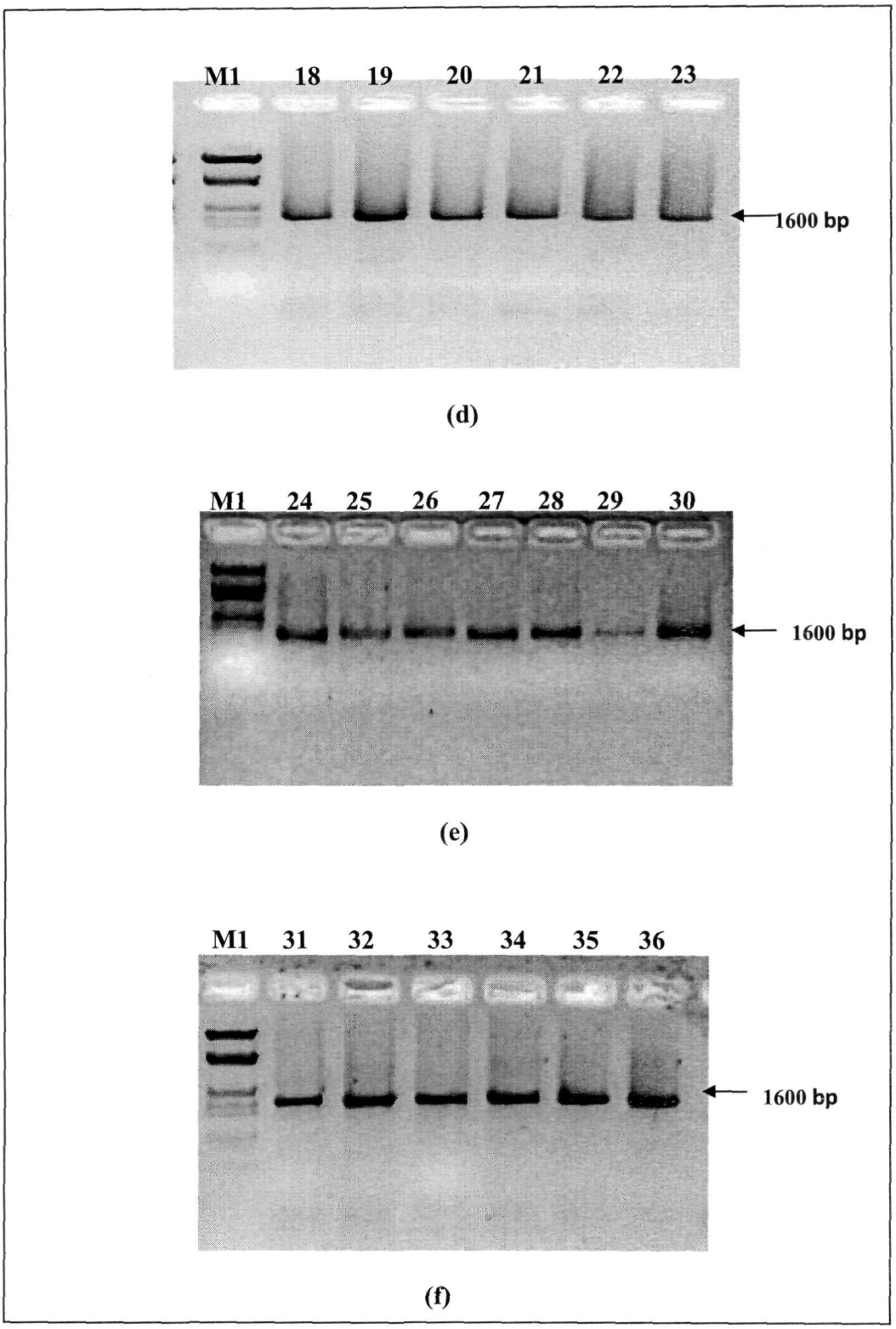


Fig. 4.4 (d-f) Amplification of the 18S rDNA region of *S. gilo* (18-20), *S. sisymbriifolium* (21-25), *S. torvum* (26-30), *S. kurzii* (31-35). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

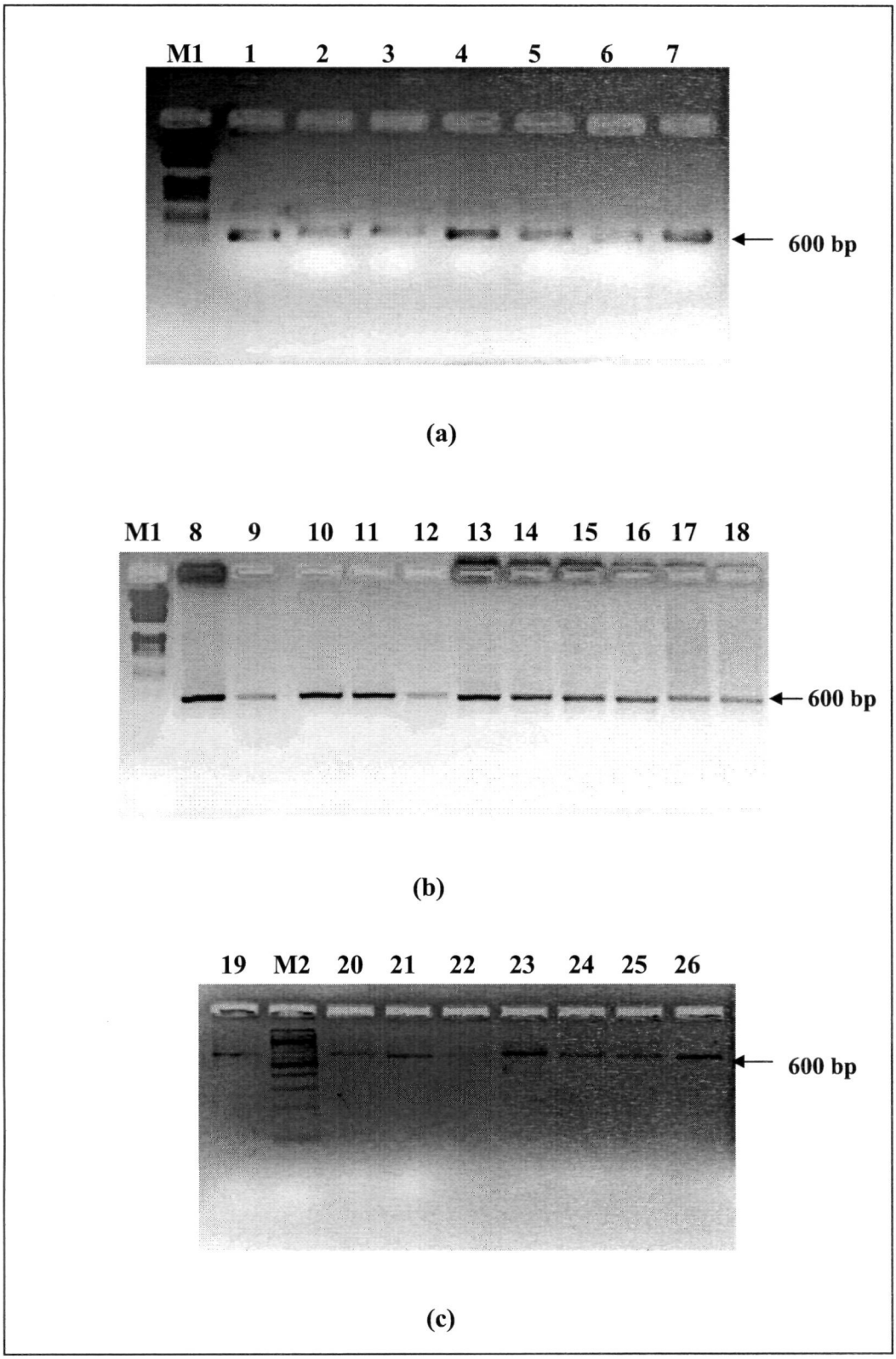


Fig. 4.5 (a-c) Amplification of the ITS region of *S. gilo* (1-5), *S. sisymbriifolium* (6-10), *S. torvum* (11-15), *S. kurzii* (16-20), *S. nigrum* (21-25), *S. khasianum* (26)
(M1= λDNA *Hind* III/*Eco*RI double digest marker, M2=100bp ladder)

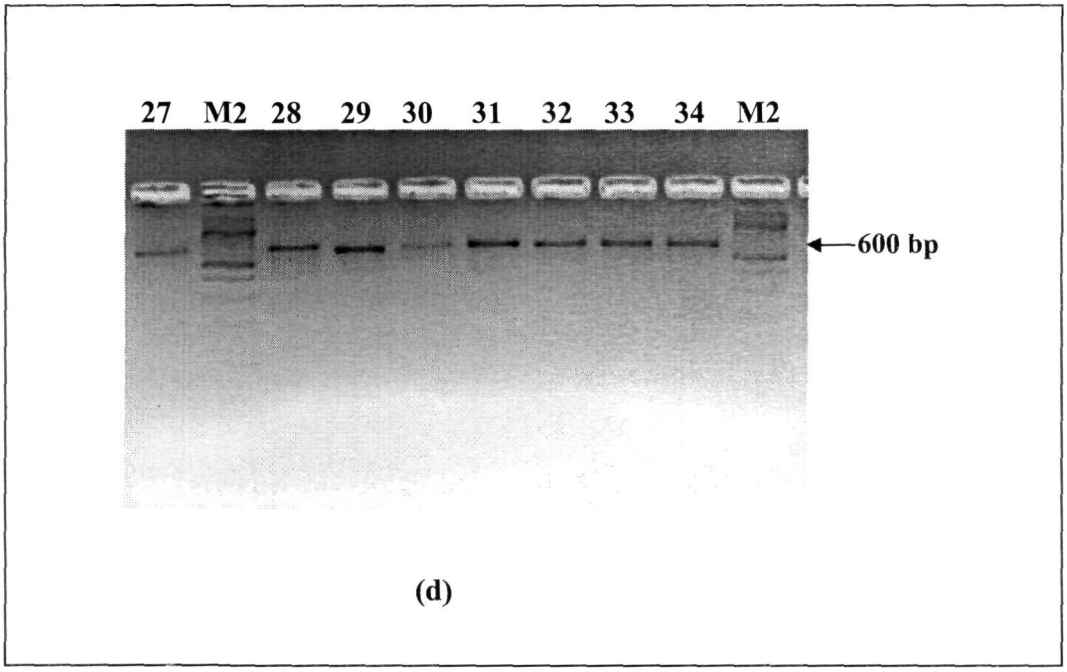


Fig. 4.6 (d) Amplification of the ITS region of *S. khasianum* (27-30), *S. clavatum*(31-34)
(M1= λDNA *Hind* III/*Eco*RI double digest marker, M2=100bp ladder)

4.3.3 AMPLIFICATION OF *rbcL*

The primers used for the amplification of *rbcL* region were S*rbcL*F and S*rbcL*R. PCR was practised in different annealing temperatures starting from 45^0C. Like other cases multiple bands were generated till the optimized temperature was obtained. The annealing temperature was optimized at 53^0C. The amplicons were loaded in 1% agarose gel (Fig.4.7-4.8). Single fragments of size approximately 1.4 Kb were observed in all cases.

4.3.4 AMPLIFICATION OF *matK*

The primers used for the amplification of *matK* region were SmtKF and SmtkR and the optimized annealing temperature was 64^0C. The amplicons were loaded in 1% agarose gel (Fig. 4.9-4.10). Single fragments of approximately 1.3 Kb were observed I n all samples.

4.4 NUCLEOTDE SEQUENCE ANALYSIS

Sequencing was carried out for all the above mentioned amplicons using 3130 Genetic Analyzer (Applied Biosystems). The respective sequences were analyzed with different computer programs as mentioned below. A comparative sequence analysis was carried out among the different genomic regions.

4.4.1 NUCLEOTIDE SEQUENCE ANALYSIS OF *18S rRNA*

The 18S rDNA amplicons of all the seven species were sequenced using the primers as described in section 3. The *18S rRNA* gene data set ranged from 1566-1571 bp, with an aligned length of 1595 characters. Sequences were subjected to

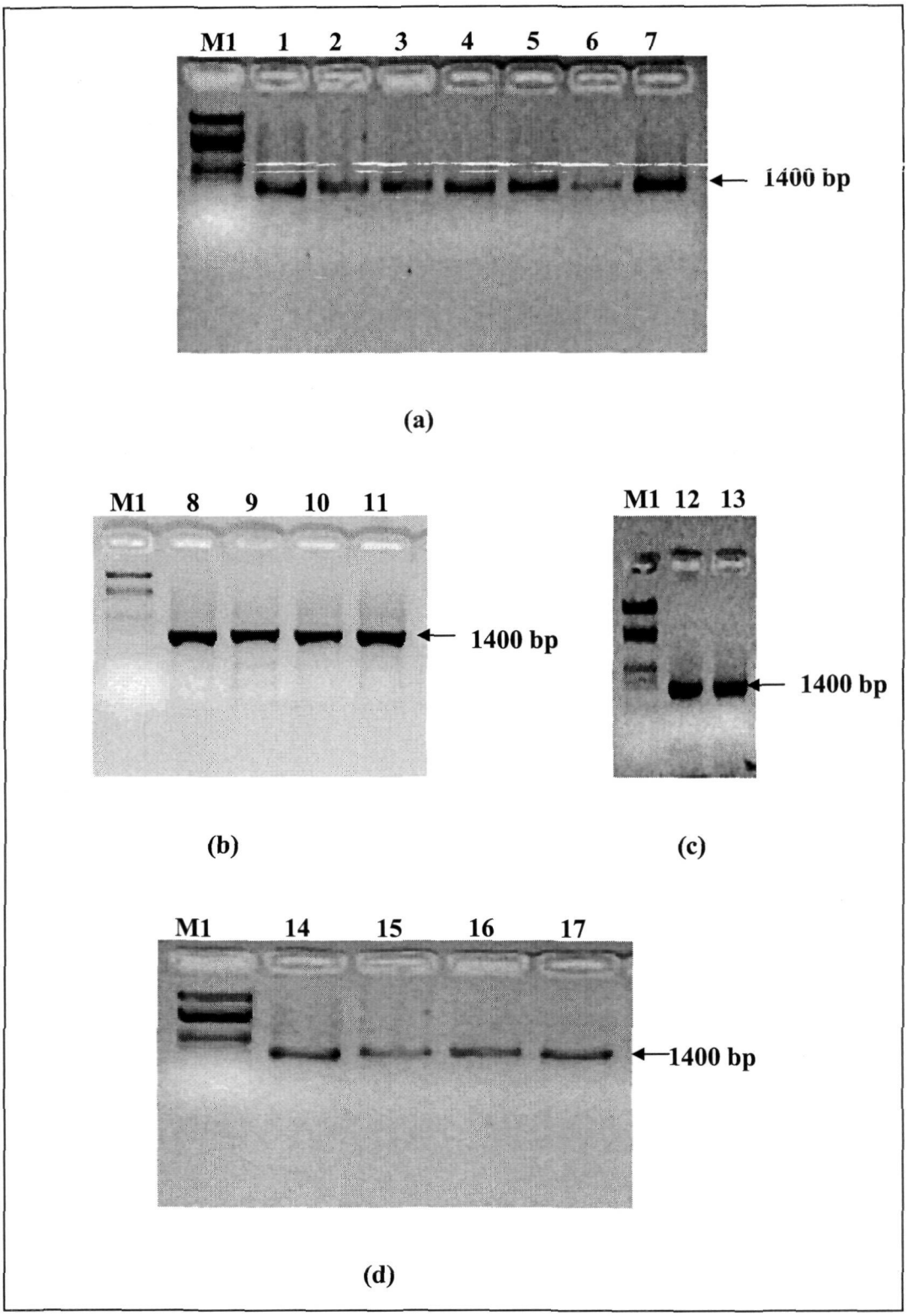
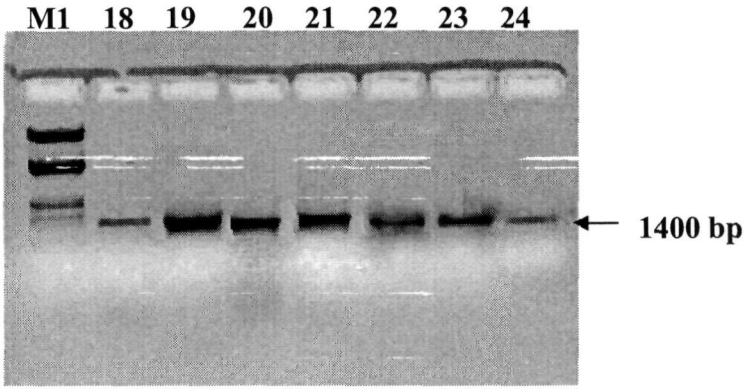
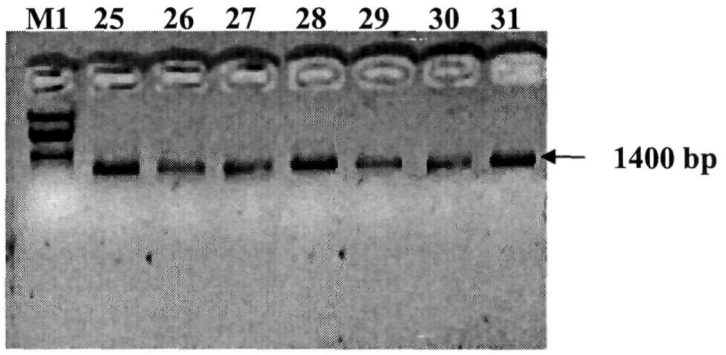


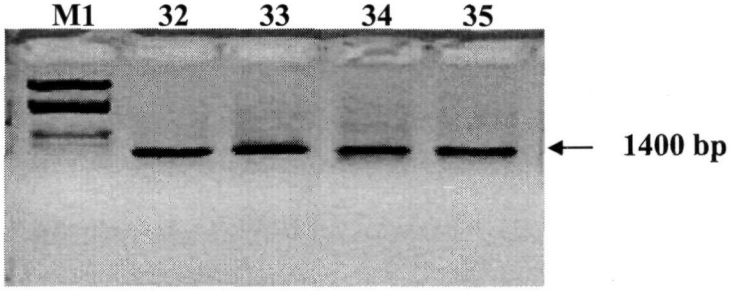
Fig. 4.7 (a-d) Amplification of *rbcL* region of *S. nigrum* (1-5), *S. khasianum* (6-10), *S. clavatum* (11-15), *S. gilo* (16-17). (M1= λDNA *Hind* III/*Eco*RI double digest marker)



(e)



(f)



(g)

Fig. 4.8 (e-g) Amplification of *rbcL* region of *S. gilo* (18-20), *S. sisymbriifolium* (21-25), *S. torvum* (26-30), *S. kurzii* (31-35). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

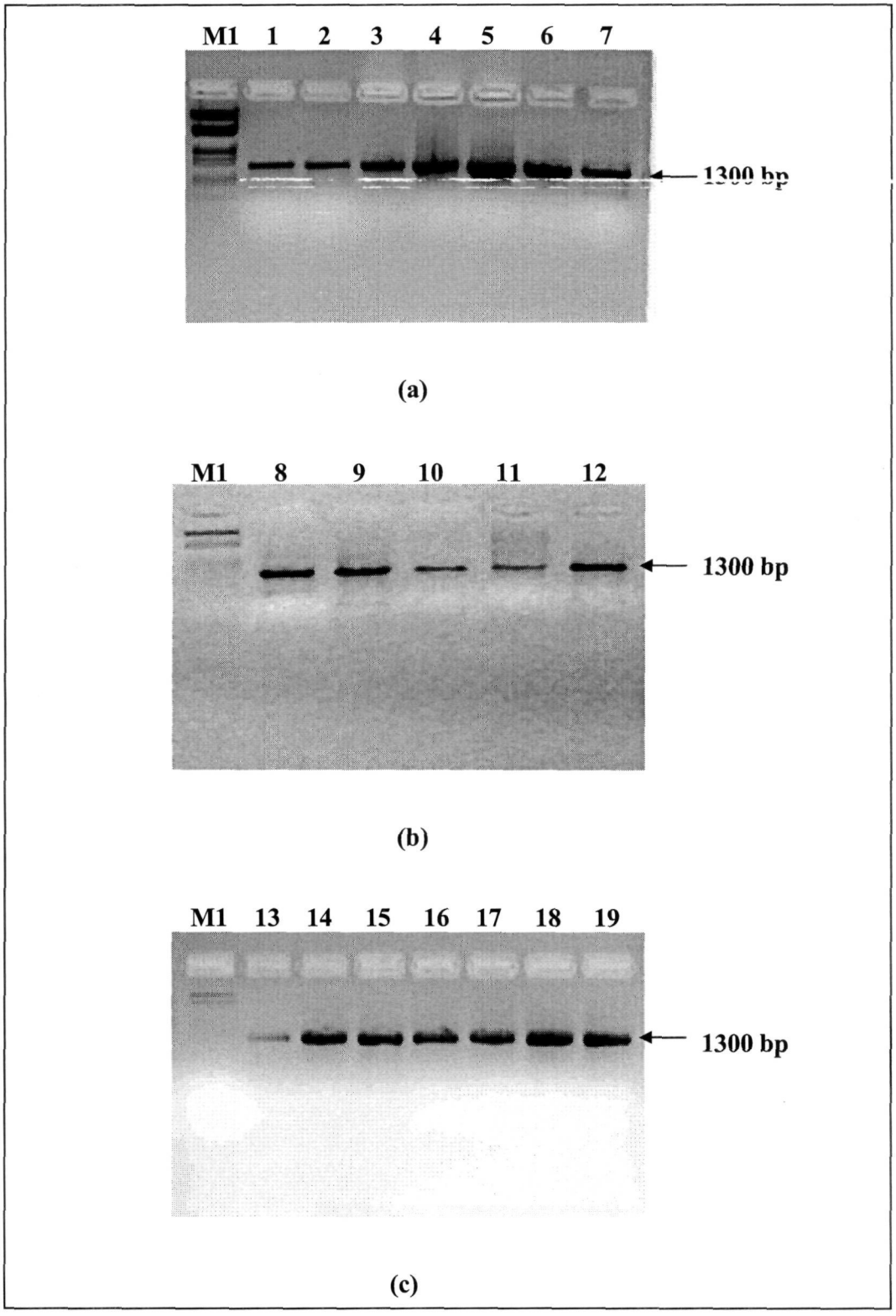


Fig. 4.9 (a-c) Amplification of *matK* region of *S. gilo* (1-5), *S. sisymbriifolium* (6-10), *S. torvum* (11-15), *S. kurzii* (16-19). (M1= λDNA *Hind* III/*Eco*RI double digest marker)

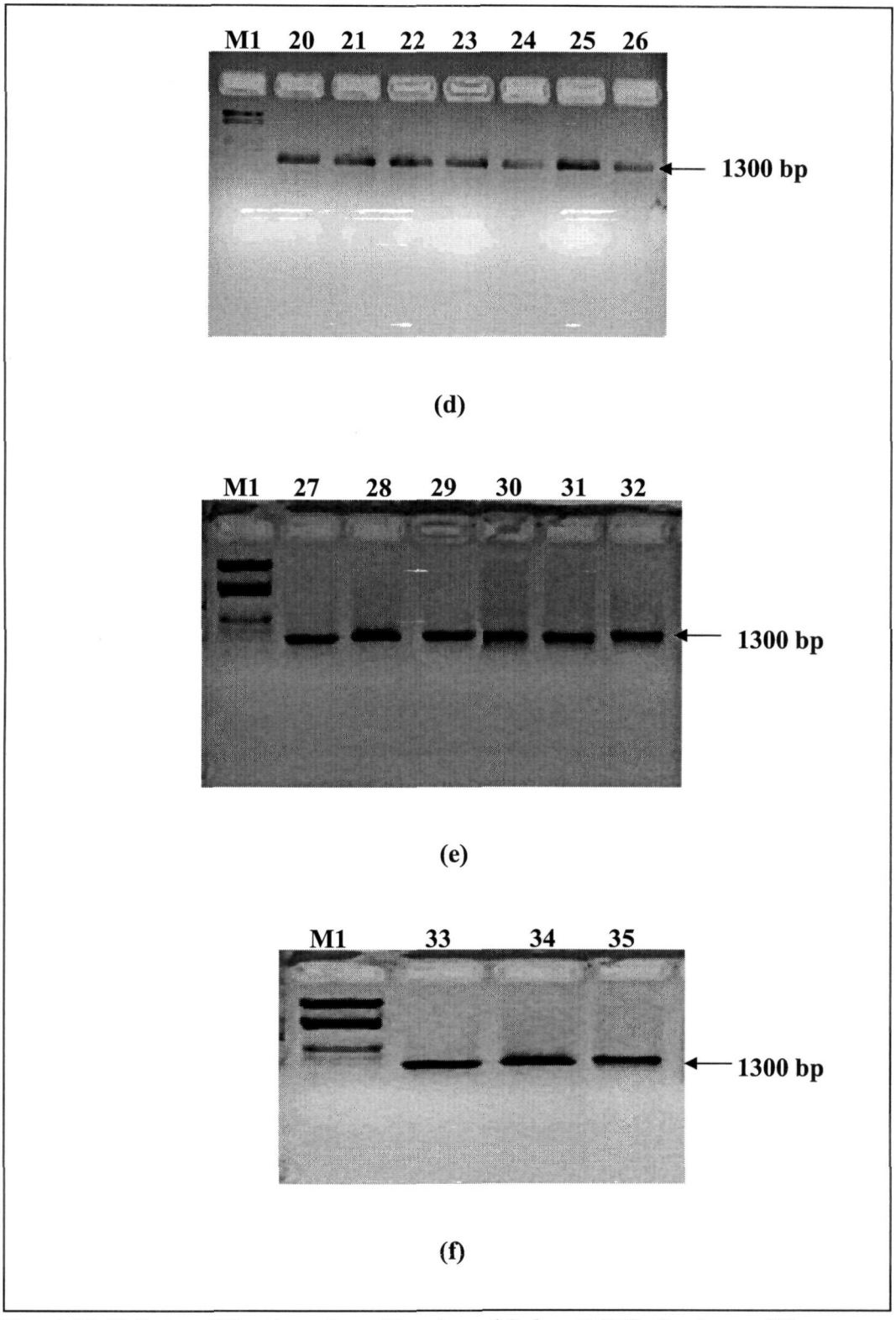


Fig. 4.10 (d-f) Amplification of *matK* region of *S. kurzii* (20), *S. nigrum* (20-25), *S. khasianum* (26-30), *S. clavatum* (31-35)
(M1= λDNA *Hind* III/*Eco*RI double digest marker)

Multiple Sequence Alignment using the CLUSTAL X (2.0) program (Thompson *et al.*, 1997). The sequence alignment of 18S rDNA region showed that the gene is highly conserved. A consensus sequence for each data set was constructed using Geneious Basic 5.3.6 (Drummond *et al.*, 2010). Sequence characteristics of this region were calculated by using both MEGA version 5 (Tamura *et al.*, 2011) and Seqstate v.1.21 (Müller, 2005) (Table 4.1). The sequence characteristics studies revealed that out of these 1595 characters 96.99% were conserved and 2.63% were variable sites and GC% of this region is 50.8%. Only 1.06% of this region was occupied by indels. The lower parsimony informative sites i.e. 0.75% indicates that though this region is phylogenetically informative, it is more conserved suggesting that it has not evolved rapidly.

4.4.2 NUCLEOTIDE SEQUENCE ANALYSIS OF *rrn* ITS

The *rrn* ITS amplicons of all the seven species were sequenced using the primers as described in section 3. The *rrn* ITS region includes ITS1, 5.8S *rRNA*, ITS2. The sequences include all these three regions. The ITS data set ranged from 575-619 bp, with an aligned length of 641 characters as it includes insertions and deletions. Sequences were subjected to Multiple Sequence Alignment using the CLUSTAL X (2.0) program (Thompson *et al.*, 1997). A consensus sequence for each data set was constructed using Geneious Basic 5.3.6 (Drummond *et al.*, 2010). Sequence characteristics of this region were calculated by using both MEGA version 5 (Tamura *et al.*, 2011) and Seqstate v.1.21 (Müller, 2005) (Table 4.1). The sequence characteristics studies revealed that out of 641 characters only 51.48% were conserved and a very considerable amount of region showed variation i.e. 46.49%. GC% of this region was 61.2%. Again a marked number of indels i.e. 19.81% were

Dataset characteristics	*18S rRNA*	ITS	*rbcL*	*matK*
Genome	Nuclear	Nuclear	Chloroplast	Chloroplast
Range of raw length	1566-1571	575-619	1385-1404	1282-1302
Aligned length	1595	641	1407	1356
Conserved sites (%)	96.99	51.48	95.8	85.69
Variable sites (%)	2.63	46.49	4.03	10.46
Parsimony informative sites (%)	0.75	17.32	1.51	1.80
GC (%)	50.8	61.2	43.9	32.4
Transitions	8	71	11	15
Transversions	4	47	7	19
Indels (%)	1.06	19.81	3.66	8
Avg. no. of base substitutions per site	0.008	0.030	0.011	0.058
Retention index (RI)	0.77	0.58	0.71	0.85
Consistency index (CI)	0.800	0.65	0.70	0.84

Table 4.1: Sequence information and comparison of data sets from two nuclear regions and two chloroplast regions.

observed in this region which indicates existence of considerably high amount of additions and deletions of sequences. These results suggested that ITS region is a highly variable region. This is expected as the *rrn* ITS1 and ITS2 are non-coding regions and are prone to random changes. The multiple sequence alignment showed that among the three partitions of ITS region (ITS1-5.8S rDNA-ITS2) only 5.8S rDNA showed conserved nature. It is expected as it is a coding region and is of only ~170 bp in length. ITS region showed a very high amount of Parsimony informative sites i.e. 17.32% together with considerably lower Consistency Index (0.65) and Retention Index (0.58) suggesting it's high level of homoplasy.

4.4.3 NUCLEOTIDE SEQUENCE ANALYSIS OF *rbcL*

The *rbcL* genes of all the seven species were sequenced using the primer pairs as described in section 3. The sequences ranged from 1385-1404 bp with an aligned length of 1407 bp as it counts the insertions and deletions. Sequences were subjected to Multiple Sequence Alignment using the CLUSTAL X (2.0) program (Thompson *et al.*, 1997). A consensus sequence for each data set was constructed using Geneious Basic 5.3.6 (Drummond *et al.*, 2010). Sequence characteristics of this region were calculated by using both MEGA version 5 (Tamura *et al.*, 2011) and Seqstate v.1.21 (Müller, 2005) (Table 4.1). The sequence characteristics showed that out of 1407 characters 95.8% were conserved, which indicates a highly conserved region. A very small amount of region was occupied by variable region i.e. 4.03% out of 1407bp aligned length. The GC% of this region was 43.9%. Indels were very rare as it occupies only 3.66%. It had a very small parsimony informative site i.e. 1.51% which suggests this region is not rapidly evolving and thus it is more conserved.

4.4.4 NUCLEOTIDE SEQUENCE ANALYSIS OF *matK*

The primer pairs that had been mentioned in section 3 were used to sequence the *matK* amplicons of all the seven species studied. The *matK* region ranged from 1282-1302 characters with an aligned length of 1356 bp as it includes indels. Sequences were subjected to Multiple Sequence Alignment using the CLUSTAL X (2.0) program (Thompson *et al.*, 1997). A consensus sequence for each data set was constructed using Geneious Basic 5.3.6 (Drummond *et al.*, 2010). Sequence characteristics of this region were calculated by using both MEGA version 5 (Tamura *et al.*, 2011) and Seqstate v.1.21 (Müller, 2005) (Table 4.1). The sequence characteristics of this region showed that out of 1356 characters 85.69% were conserved and 10.46% were variable regions. These results indicate that though *matK* is a coding region it still contains a high variability suggesting that it is a rapidly evolving coding region. The GC% of this region was 32.4%. Indels were also common in this region as they occupied 8% of the region. This region showed 1.8%, parsimony informative sites together with high consistency and retention indices i.e. 0.84 and 0.85 respectively.

4.5 DISCUSSION: SEQUENCE ALIGNMENT, STATISTICS AND COMPARATIVE UTILITY OF THE FOUR REGIONS

Among the four regions *18S rRNA* gene region was found to be most conserved (conserved sites =96.99%) as shown in the Table 2 followed by *rbcL* (95.80%). ITS is the most variable region (46.49%) generating highest number of parsimony informative sites (17.32%) followed by *matK* (10.46%) which provides 1.80% of parsimony informative sites. The GC% between the nuclear and chloroplast datasets as seen from table 4.1 indicated that the nuclear regions contained higher

GC content than the chloroplast regions. Among all the four studied regions transitions were higher than transversions, except for *matK* where transversions were seen to be more frequent compared to transitions. The highest number of indels were recorded in ITS followed by *matK*.

All the four data sets including two regions from nuclear genome and two from chloroplast genome resulted in almost similar topologies. However, in terms of phylogenetic utility, the chloroplast *matK* gene was found to be more informative having both a high percentage of Parsimony informative characters (1.8%) as well as high Consistency and Retention indices (0.84, 0.85) which suggests a low level of homoplasy (Table 4.1). The nuclear *18SrRNA* and chloroplast *rbcL* regions were also phylogenetically informative but were more conserved suggesting that they are not evolving rapidly, having a lower percentage of parsimony informative characters. The nuclear *ITS* data set had higher percentage of parsimony informative characters than *matk* (Table 4.1), which suggests its higher level of homoplasy.

The most common criticism of *18S rRNA* gene region as a source of phylogenetic information has been that it is not sufficiently variable for phylogenetic reconstruction within the angiosperms. It is previously reported that *18S rRNA* region typically evolves at one–third to one-half the rate of *rbcL* (Nickrent and Soltis, 1995). But in the present study it was seen that there was very little difference between *18S rRNA* and *rbcL* data sets in terms of conserved sites (96.99% and 95.8% respectively, Table 4.1).

4.6 PHYLOGENETIC ANALYSES

The phylogenetic trees were constructed to reveal the evolutionary relationship between all the taxa studied here. In all the analyses, *Nicotiana tabacum* was used as an outgroup.

4.6.1 PHYLOGENETIC ANALYSIS OF *18S rRNA* GENE

Multiple sequence alignment of the *18S rRNA* genes of all the seven species studied as well as sequences retrieved from GenBank (Table 3.1) were carried out using CLUSTAL W. The aligned sequences were used for phylogenetic analysis. Neighbour Joining tree was constructed from a Kimura two parameter (Kimura, 1980) distance matrix using MEGA version5 (Tamura *et al.*, 2011). The strength of support for individual tree branches was estimated using bootstrap values (BS) (Felsenstein, 1985). Bootstrap analysis was carried out for 1000 replicates to examine the relative level of support for individual clades on the cladograms of each search and were expressed in percentage. In the Neighbor joining tree the sum of branch length for optimal tree was=0.01980760.

The phylogenetic tree (Fig. 4.11) constructed by using *18S rRNA* dataset depicted that the tree consisted of three distinct clades. The monophyly of *S. gilo and S. kurzii* is supported by 99% bootstrap value. *S. torvum* emerged as a sister to the clade *S. gilo* + *S. kurzii* (BS=92%). *S. khasianum* and *S. sisymbriifolium* were grouped together and were supported by high BS (99%) showing their close relationship. *S. nigrum* and *S. clavatum* also showed very strong evolutionary relationship supported by high BS value (99%). Though *S. lycopersicum* and *S.*

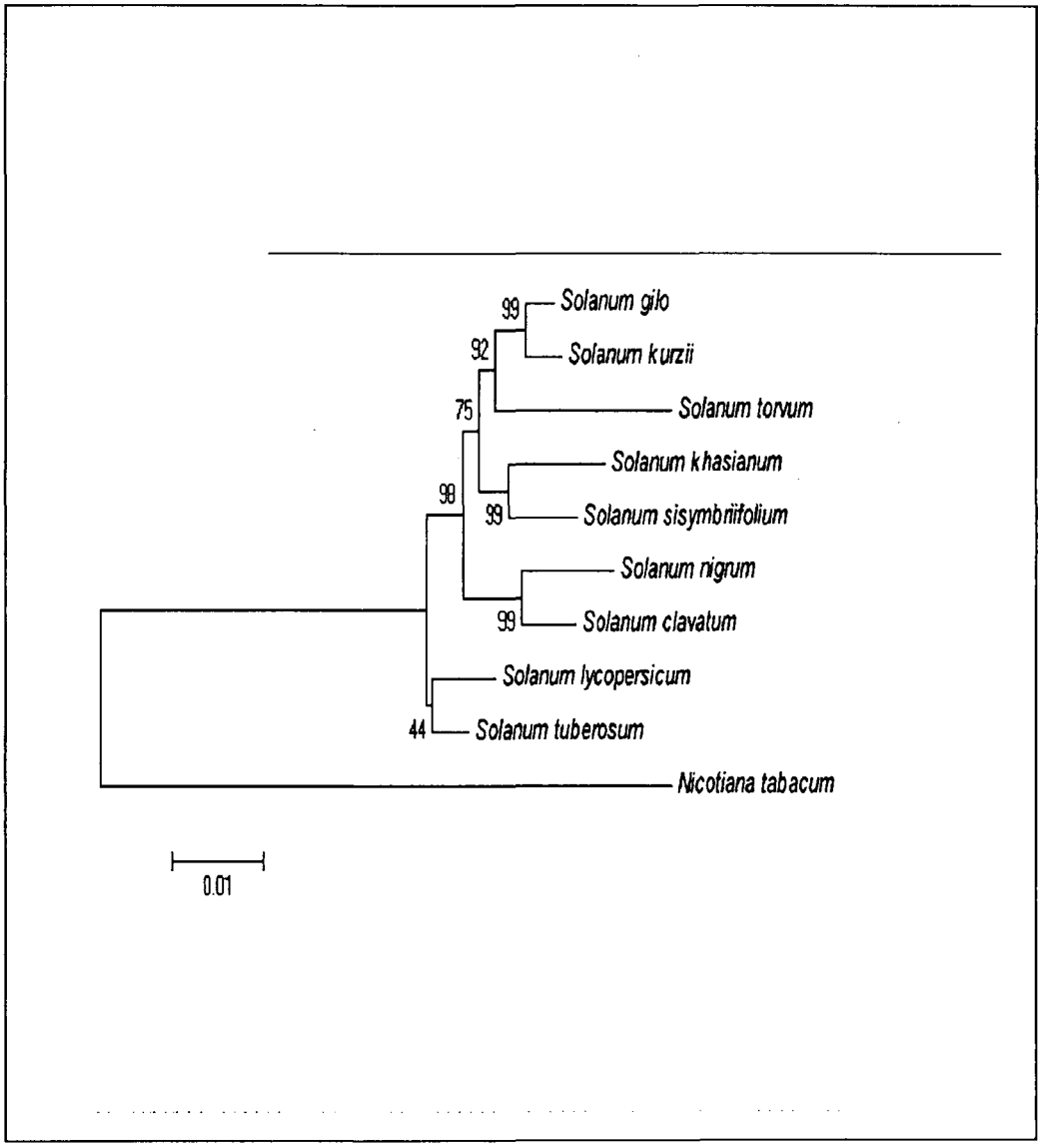


Fig. 4.11 NJ tree constructed using 18S rDNA data set.

tuberosum are clustered together but their relationship is poorly supported (BS=44%).

4.6.2 PHYLOGENETIC ANALYSIS OF *rrn* ITS

The *rrn* ITS sequences of all the seven species together with the sequences retrieved from GenBank (Table 3.1) were used for multiple sequence alignment. Multiple sequence alignment was carried out using CLUSTAL W. The aligned sequences were used for phylogenetic analysis. Neighbour Joining tree was constructed from a Kimura two parameter (Kimura, 1980) distance matrix using MEGA version5 (Tamura *et al.*, 2011). The strength of support for individual tree branches was estimated using bootstrap values (BS) (Felsenstein, 1985). To examine the relative level of support for individual clades on the cladograms of each search, Bootstrap analysis was carried out for 1000 replicates and were expressed in percentage. In the Neighbor joining tree the sum of branch length for optimal tree was = 0.54315707.

The phylogenetic analysis (Fig. 4.12) showed that *S. kurzii* and *S. gilo* were clustering together with a bootstrap support of 51%. *S. torvum* was again grouping with *S. kurzii* and *S. gilo* with a very high bootstrap value i.e. 99% which indicated that *S. torvum* had recently diversified from the group of *S. kurzii* and *S. gilo*. So *S. kurzii, S. gilo* and *S. torvum* together were making a subclade. Another subclade was made by *S. sisymbriifolium* and *S. khasianum* but their relationship was supported by low bootstrap value i.e. 42%. *S. nigrum* and *S. clavatum* were clustering together with a very high bootstrap value i.e. 100% which indicated their recent diversification and the formation of the clade was well supported. *S.*

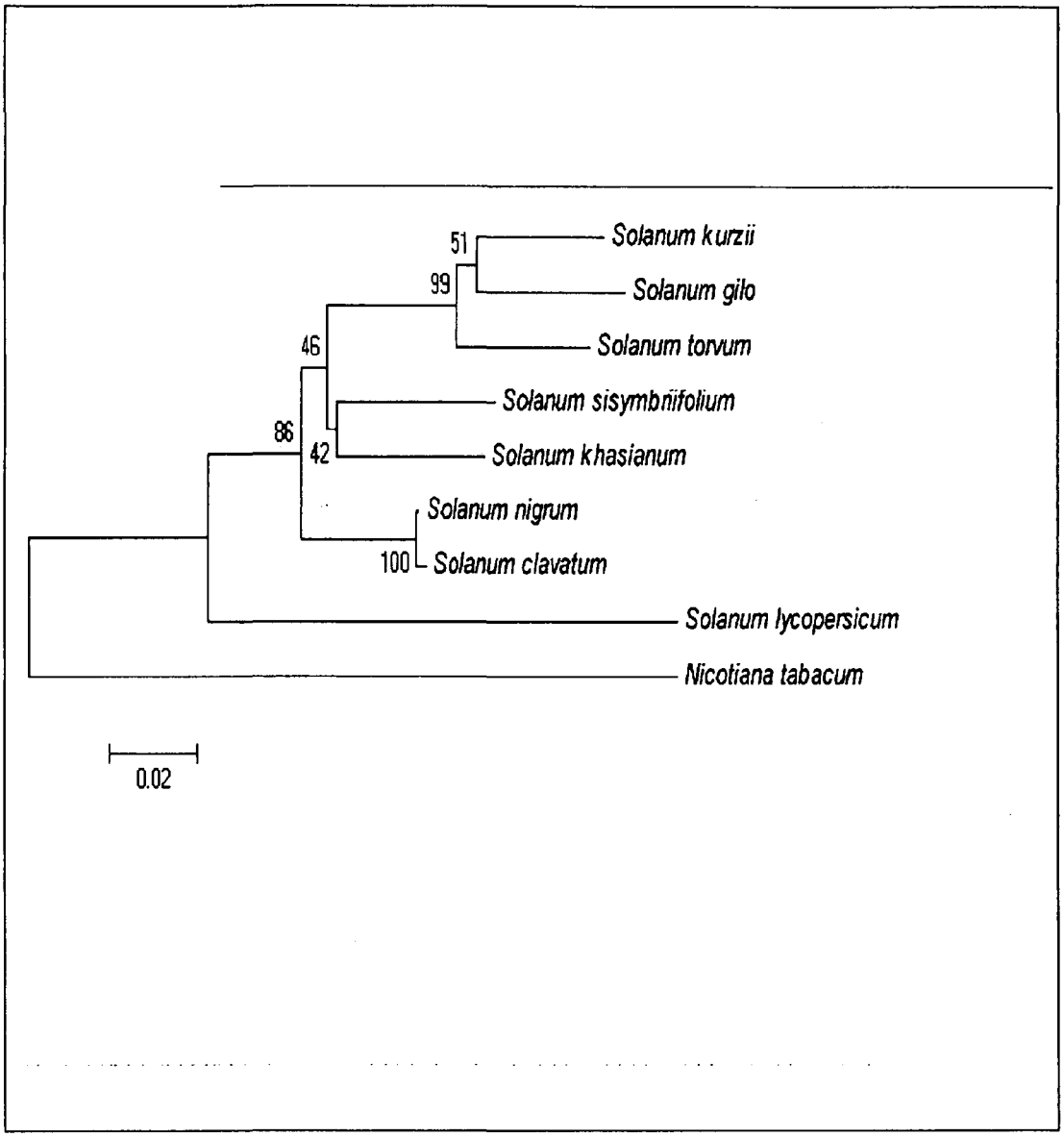


Fig. 4.12 NJ tree constructed using ITS data set.

lycopersicum emerged as a separate lineage. So, the phylogenetic relationship inferred by using ITS sequences and 18S rDNA sequences were congruent.

4.6.3 PHYLOGENETIC ANALYSIS OF *18S rRNA* and *rrn* ITS IN COMBINATION

The selected ML model which best fit the nuclear sequence data set including 18S rDNA sequences and ITS sequences was TrN+G (Table 3.1) as analysed with jModel test (Pasoda, 2008). The ML model parameters included nucleotide frequencies of A= 0.2323, C= 0.2627, G= 0.2876, T= 0.2174; a substitution rate matrix of A to C: 0.9373, A to G: 1.5169, A to T: 0.8920, C to G: 0.8447, C to T: 5.4881, G to T: 1.0000; a proportion of invariant sites = 0.1958; and a gamma rate distribution at variable sites with shape (alpha) = 0.0150.

The Bayesian analysis of the *18S rRNA* and ITS in combination resulted in 8 trees obtained from four Bayesian runs implementing the TrN+G (Tamura-Nei plus Gamma) model. The 50% majority-rule consensus tree in Fig.4.13 depicted the support for relationships between the taxa by posterior probability values. The tree topology obtained from the Bayesian inference of the combined data set of 18S rDNA + ITS were not congruent with that of the two data sets separately. The exceptional result obtained from the nuclear data set is the monophyly of the clade *S. lycopersicum* + *S. nigrum* + *S. clavatum* but are supported by very poor posterior probability value (58%). *S. sisymbriifolium* and *S. khasianum* emerged as two separate lineages which is also contradictory to the result obtained from the trees constructed by using ITS and 18S rDNA data sets separately. This difference might be due to the combination of one much conserved data set and another frequent variable data set.

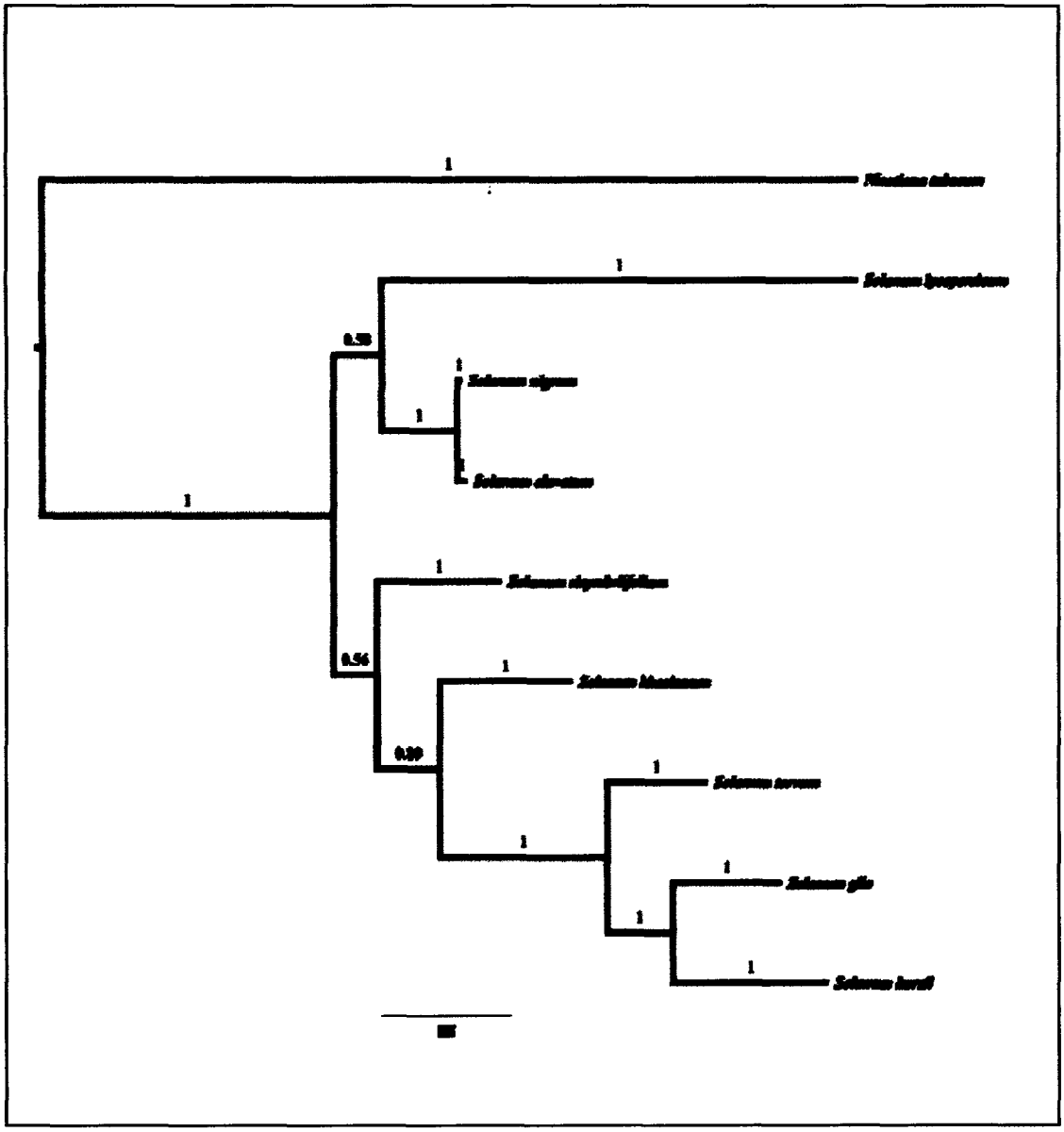


Fig. 4.13 The 50% majority rule consensus tree inferred from Bayesian analysis of the two nuclear data sets combined. Posterior clade probability (>50%) are given above the branches.

4.6.4 PHYLOGENETIC ANALYSIS OF *rbcL*

Multiple sequence alignment of the *rbcL* gene of all the seven species studied were carried out in conjunction with sequences retrieved from GenBank (Table 3.1) using CLUSTAL W. Phylogenetic analysis was performed by using the aligned sequences. Neighbour-Joining tree was constructed from a Kimura two parameter (Kimura, 1980) distance matrix using MEGA version5 (Tamura *et al.*, 2011). The strength of support for individual tree branches was estimated using bootstrap values (BS) (Felsenstein, 1985). Bootstrap analysis was carried out for 1000 replicates to examine the relative level of support for individual clades on the cladograms of each search and were expressed in percentage. In the Neighbor joining tree the sum of branch length for optimal tree was= 0.13780244.

The phylogenetic tree constructed by using *rbcL* dataset (Fig.4.14) revealed the same tree topology as the tree obtained by using the 18Sr DNA data set and ITS data set. The tree depicted that *S. nigrum* and *S. clavatum* were sister taxa as supported by high bootstrap value (100%) but the position of the clade *S. nigrum*+*S. clavatum* was contradictory with that of the topology obtained from nuclear dataset. However the terminal position of the clade *S. nigrum*+*S. clavatum* and their association with *S. torvum* was supported by very poor bootstrap value (BS=41%). *S. kurzii* and *S. gilo* also showed very strong evolutionary relationship (BS=94%) as in Fig. 4.11. The relationship between *S. khasianum* and *S. sisymbriifolium* was supported by intermediate to low BS (66%). *S. tuberosum* and *S. lycopersicum* emerged as separate branches suggesting their separate origins.

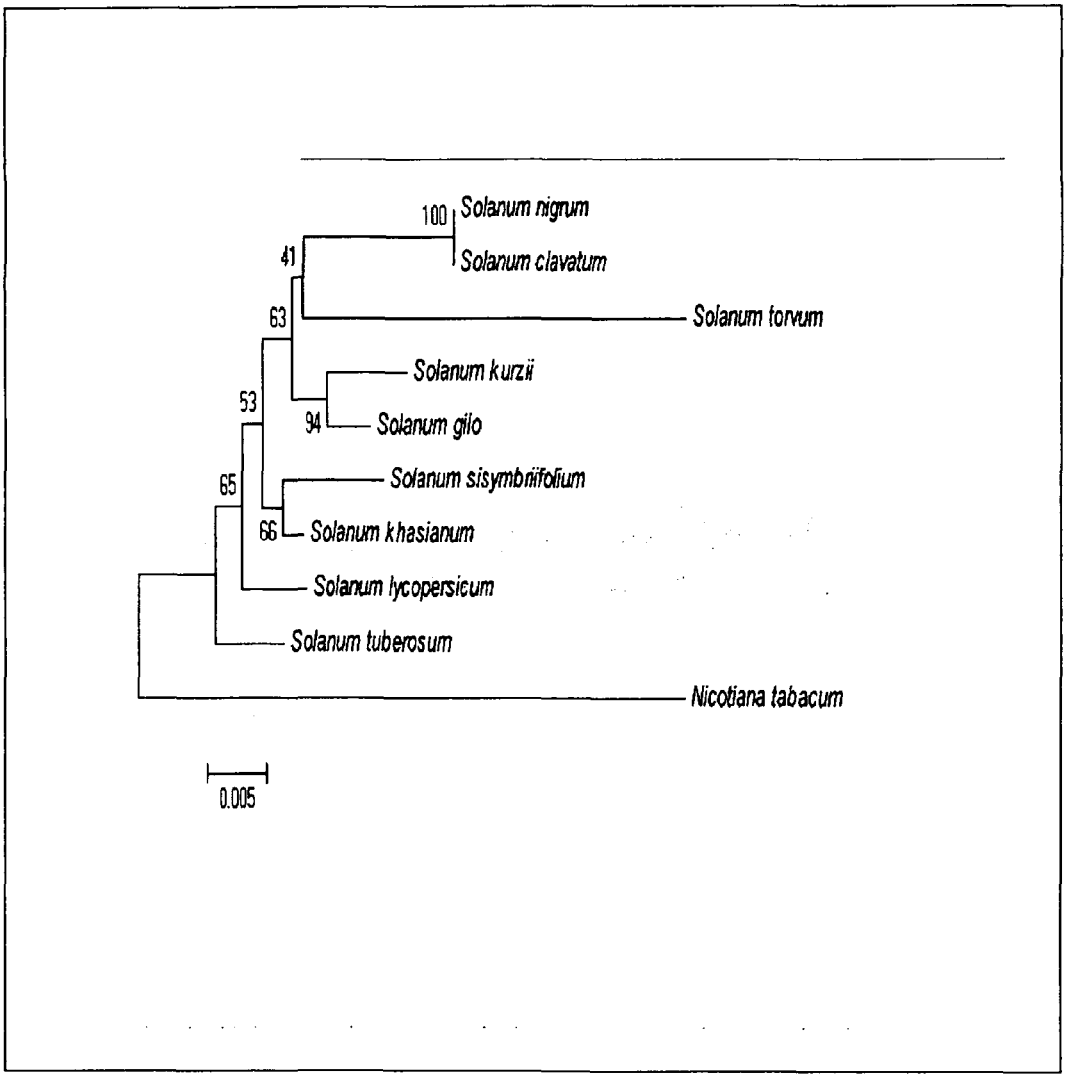


Fig. 4.14 NJ tree constructed by using *rbcL* dataset.

4.6.5 PHYLOGENETIC ANALYSIS OF *matK*

The *matK* sequences of all the seven species together with the sequences retrieved from GenBank (Table 3.1) were used for multiple sequence alignment. Multiple sequence alignment was carried out using CLUSTAL W. The aligned sequences were used for phylogenetic analysis. Neighbour Joining tree was constructed from a Kimura two parameter (Kimura, 1980) distance matrix using MEGA version5 (Tamura *et al.*, 2011). The strength of support for individual tree branches was estimated using bootstrap values (BS) (Felsenstein, 1985). To examine the Relative level of support for individual clades on the cladograms of each search Bootstrap analysis was carried out for 1000 replicates and were expressed in percentage. In the Neighbor joining tree the sum of branch length for optimal tree was = 0.23776973.

The tree topology as shown in Fig. 4.15 revealed close association of *S. gilo* and *S. kurzii* (BS=94%) and with 88% BS support *S. torvum* was closer to them. This was similar to the information obtained from nuclear data (Fig. 4.13). The position of *S. nigrum* and *S. clavatum* was contradictory with that of the previous trees. *S. lycopersicum* and *S. tuberosum* showed a recent diversification supported by intermediate BS (89%). On basis of this tree topology the evolutionary relationship between *S. khasianum* and *S. sisymbriifolium* was not strong (BS=37%).

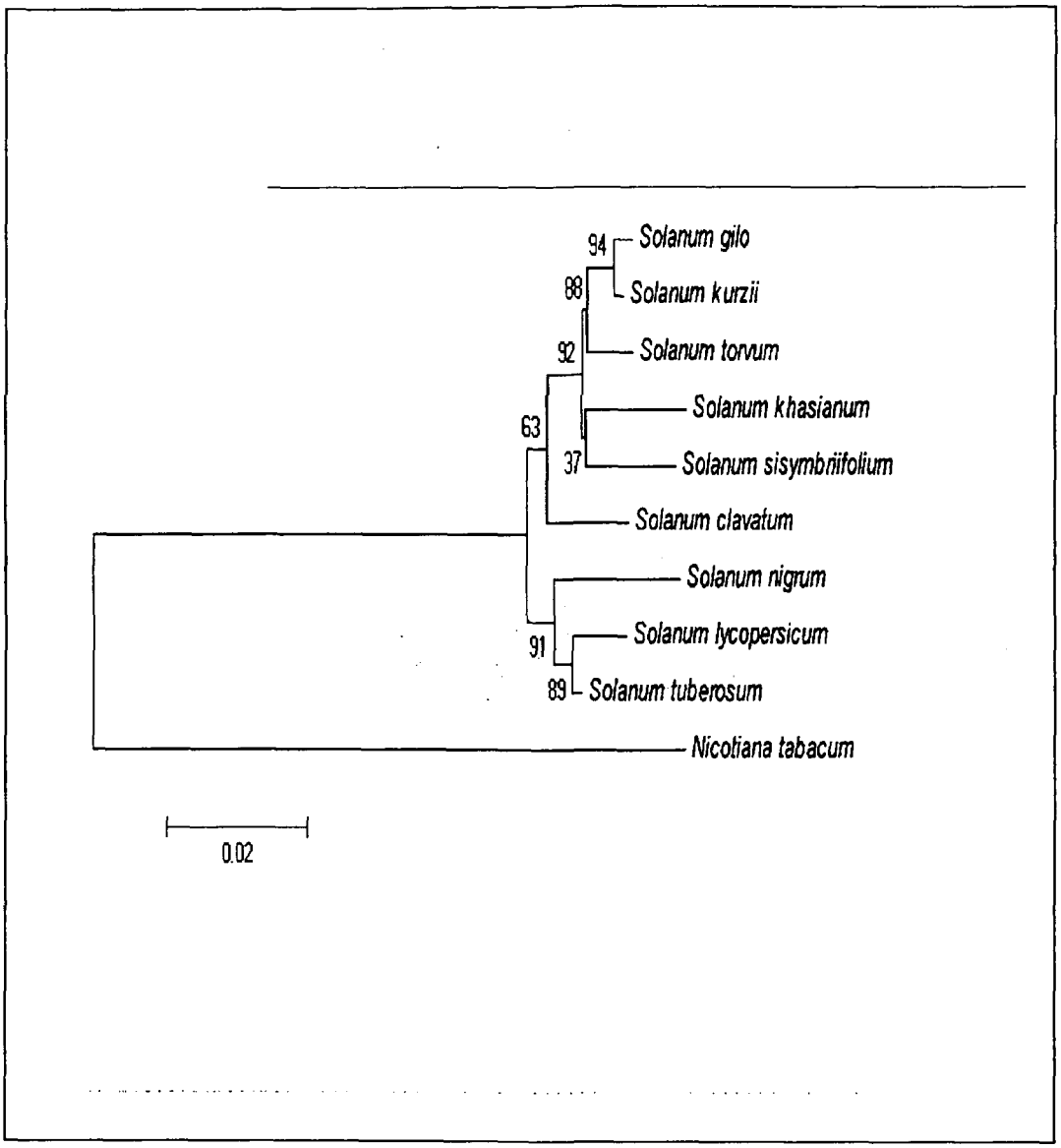


Fig. 4.15 NJ tree constructed by using *matK* dataset.

4.6.6 PHYLOGENETIC ANALYSIS OF *rbcL* AND *matK* IN COMBINATION

The ML model selected for the combined chloroplast data set which includes both *rbcL* and *matK* was TVM+G which best fit the data set by using jModel test (Pasoda, 2008). The ML model parameters included nucleotide frequencies of A= 0.2954, C= 0.1904, G= 0.1800, T= 0.3341; a substitution rate matrix of A to C: 1.7112, A to G: 2.0312, A to T: 0.5562, C to G: 1.4020, C to T: 2.0032, G to T: 1.0000; a proportion of invariant sites = 0.0008; and a gamma rate distribution at variable sites with shape (alpha) = 0.4700.

The chloroplast data sets *rbcL* and *matK* were combined in a single data matrix and Bayesian analysis was conducted implementing TVM+G (Transversion Model plus Gamma) model. It resulted in 11 trees for four Bayesian runs. The posterior probability values are shown (Fig. 4.16) above the branches in the 50% majority-rule consensus tree. The tree of this combined chloroplast data set was more resolved than the trees made with *rbcL* and *matK* separately (Fig. 4.14, 4.15) in terms of both clustering of taxa and branch supports. The tree topology was congruent with that of trees obtained from nuclear *18S rRNA* gene data set and ITS data set separately (Fig. 1, 2). It supported the clustering of *S. kurzii* +*S. gilo*, *S. torvum* + (*S. kurzii* +*S. gilo*), *S. nigrum* + *S. clavatum* by 100% posterior probability. The association of *S. khasianum* and *S. sisymbriifolium* was also supported by high posterior probability value (PP=96%). *S. lycopersicum* had emerged as a separate branch and was supported by 100% posterior probability. In this analysis *S. tuberosum* was not included (though there was sufficient information for both the chloroplast regions), so that the topologies between the two trees derived from the nuclear and chloroplast data sets remain comparable.

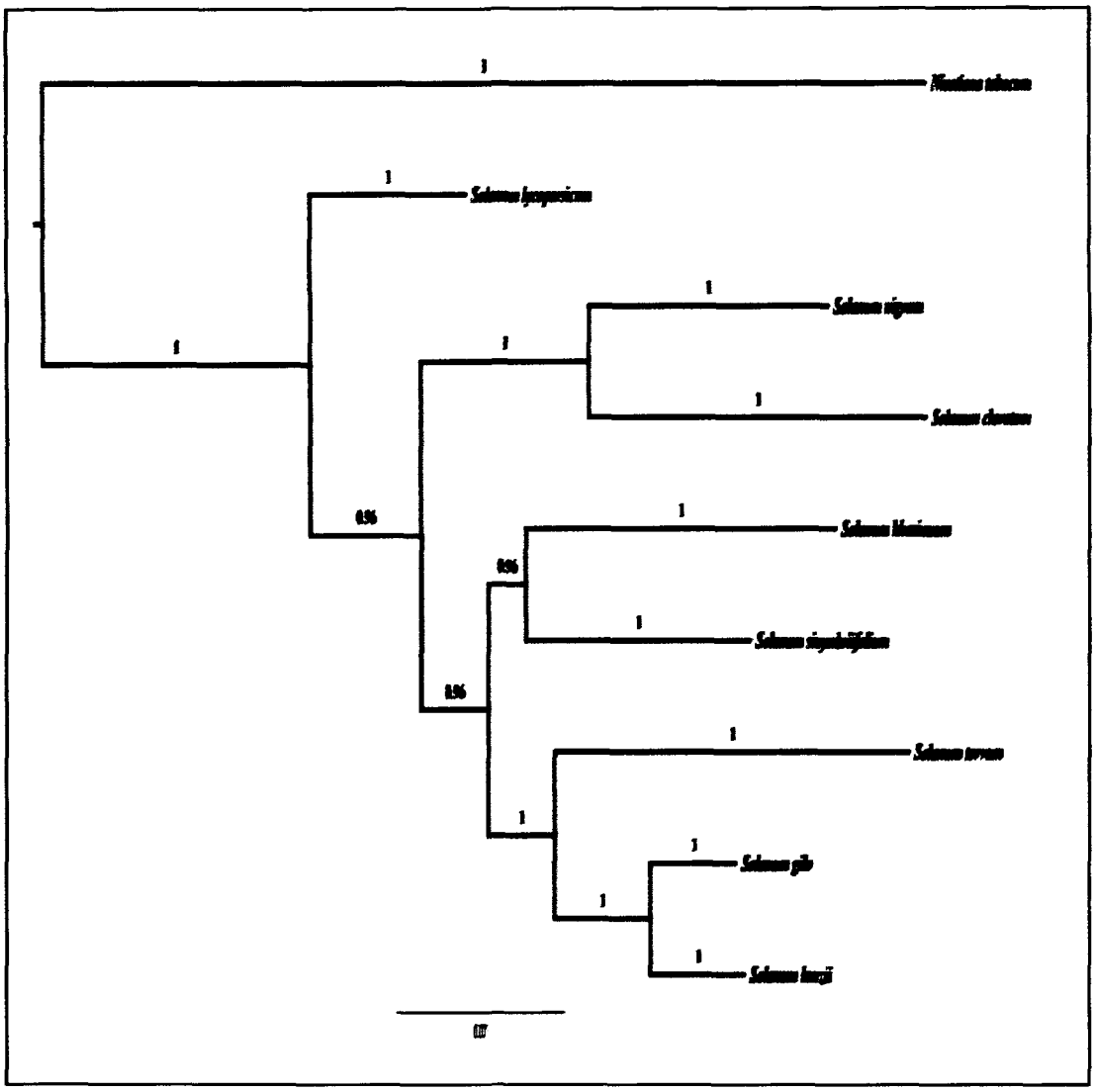


Fig. 4.16 The 50% majority rule consensus tree inferred from Bayesian analysis of the two chloroplast data sets combined. Posterior clade probability (>50%) are given above the branches.

4.6.7 PHYLOGENETIC ANALYSIS OF BOTH NUCLEAR AND CHLOROPLAST DATASETS IN COMBINATION

Maximum likelihood (ML) analysis of the combined data set of four regions was conducted with parameters estimated using jModeltest (Posada, 2008). This procedure indicated that the GTR+G (General Time Reversible plus Gamma) model best fit the data. The ML model parameters included nucleotide frequencies of A =0.2597, C =0.2149, G =0.2374, and T =0.2879; a substitution rate matrix of A to C: 2.0696, A to G: 2.5181, A to T: 0.9294, C to G: 1.5903, C to T: 3.9687, and G to T: 1.0000; a proportion of invariant sites = 0.0190; and a gamma rate distribution at variable sites with shape = 0.2440. The evolutionary history was also inferred using the Neighbor-Joining method. The Bayesian analysis generated 8 trees and the 50% majority-rule consensus tree is shown in Fig. 4.17 with posterior probability above the branches.

The phylogenetic tree constructed from the combined data set (Fig. 4.17) was more resolved. It was congruent with the tree topology revealed from the chloroplast data set (Fig. 4.16) more than the tree revealed by the use of nuclear data set (Fig. 4.13). The tree depicted the polyphyly of the three clades. Taxons of each clade were displayed with same colour (Fig. 4.17). The terminal clade which can be named as clade III is composed of two sub clades which are supported by high posterior probability values - *S. kurzii* + *S. gilo* + *S. torvum* (PP=100%) and *S. khasianum* + *S. sisymbriifolium* (PP=100%). The clade II consists of *S. nigrum* + *S. clavatum* and their monophyly is supported by 100% posterior probability. Clade I comprise of *S. lycopersicum* and its emergence as a separate branch was supported by 100% posterior probability. The tree topology also confirmed the close

relationship between *S. kurzii* and *S. gilo* (PP=100%), *S. nigrum* and *S. clavatum* (PP=100%) as well as *S. khasianum* and *S. sisymbriifolium* (PP=100%).

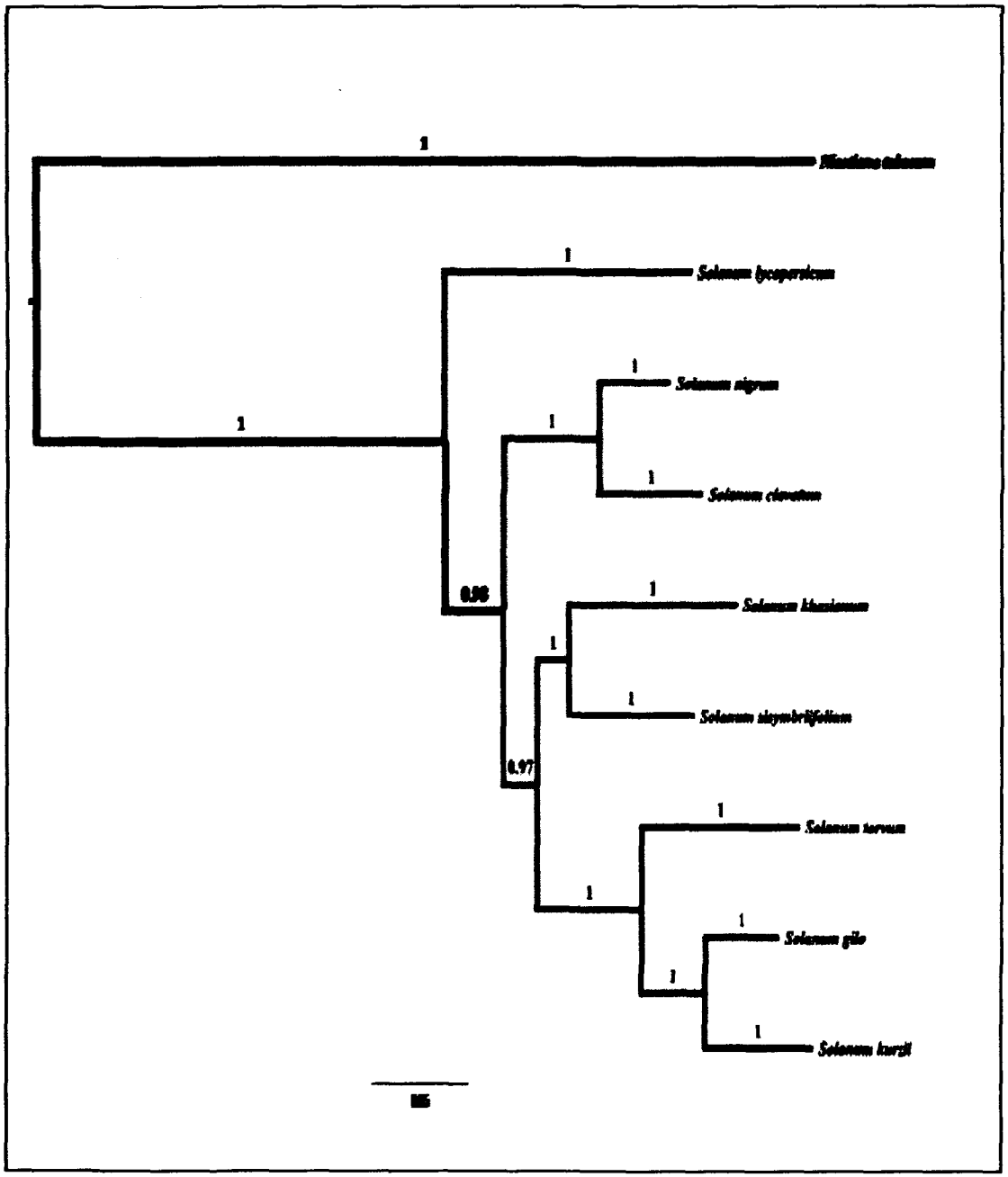


Fig. 4.17 The 50% majority rule consensus tree inferred from Bayesian analysis of the two nuclear data sets and two chloroplast data sets in combinations. Posterior clade probability (>50%) are given above the branches.

4.7 AMPLICON RESTRICTION PATTERN (ARP)/ PCR- RESTRICTION FRAGMENT LENGTH PROFILE (PCR-RFLP)

The amplicons of ITS and *matK* were digested with different restriction enzymes so that polymorphic restriction patterns could be obtained. Restriction digestion was carried out on all the five replicates of each seven species studied.

4.7.1 AMPLICON RESTRICTION PATTERNS (ARP) WITH *NciI*

Restriction enzyme *Nci*I is obtained from the bacterium *Neisseria cinerea.* This is a five base cutter restriction enzyme with a recognition site of "CCSGG" where'S' represents either 'C' or 'G'. The enzyme slices between 'C' and 'S'. Altogether four different profiles were generated when this enzyme was used to digest the amplicons of *rrn* ITS of the samples studied.

The four different profiles were named as PN1, PN2, PN3 and PN4. PN1 profile was obtained for all the five replicates of *Solanum khasianum* (SKh1, SKh2, Skh3, SKh4, and SKh5). PN2 profile comprised of all the five replicates of *Solanum sisymbriifolium* (SS1, SS2, SS3, SS4, and SS5). PN3 profile was obtained for all the representatives of *Solanum clavatum* (SC1, SC2, SC3, SC4, and SC5). Profile 4 was obtained for all the five representatives of each *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5); *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5); *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4, and SKu5) and *Solanum nigrum* (SN1, SN2, SN3, SN4, and SN5). One of the interesting outcomes of this restriction digestion is that-there was no intraspecific polymorphism within the species. This restriction enzyme could successfully differentiate between *Solanum khasianum*, *Solanum*

sisymbriifolium, *Solanum clavatum* and each of these three species from the group of *Solanum torvum*, *Solanum gilo*, *Solanum kurzii* and *Solanum nigrum* as a whole. So it is also revealed from this study that this particular restriction enzyme failed to differentiate between *Solanum torvum*, *Solanum gilo*, *Solanum kurzii* and *Solanum nigrum* as no marked polymorphism in the restriction banding pattern between these species were observed.

4.7.1.1 Restriction digestion analysis of profile PN1

Five replicates of *Solanum khasianum* namely SKh1, SKh2, Skh3, SKh4, and SKh5 were digested with restriction enzyme *Nci*I. Computer simulated restriction digestion analysis of the *rrn* ITS sequence of *Solanum khasianum* identified one restriction site at position 141 bp which generated two bands of sizes 478 bp and 141 bp (Table 4.3). All the five replicates of *Solanum khasianum* showed the profile PN1 which comprised of two bands of sizes ~480 bp and ~140 bp. So it is seen that both the restriction bands generated by the restriction digestion and the computer simulated restriction bands tally with each other.

4.7.1.2 Restriction digestion analysis of profile PN2

Five replicates of *Solanum sisymbriifolium* namely SS1, SS2, SS3, SS4, and SS5 were digested with restriction enzyme *Nci*I and showed the profile PN2. Computer simulated restriction digestion analysis of the *rrn* ITS sequence of *Solanum sisymbriifolium* detected three restriction sites at position 65th bp, 141st bp and 211th bp which would generate fragments of sizes 408 bp, 77 bp, 69 bp and 64bp (Table 4.3). But the restriction digestion of the amplicons with this enzyme generated three fragments of sizes ~400 bp, ~140 bp and ~50 bp. Here 408 bp and 64 bp bands

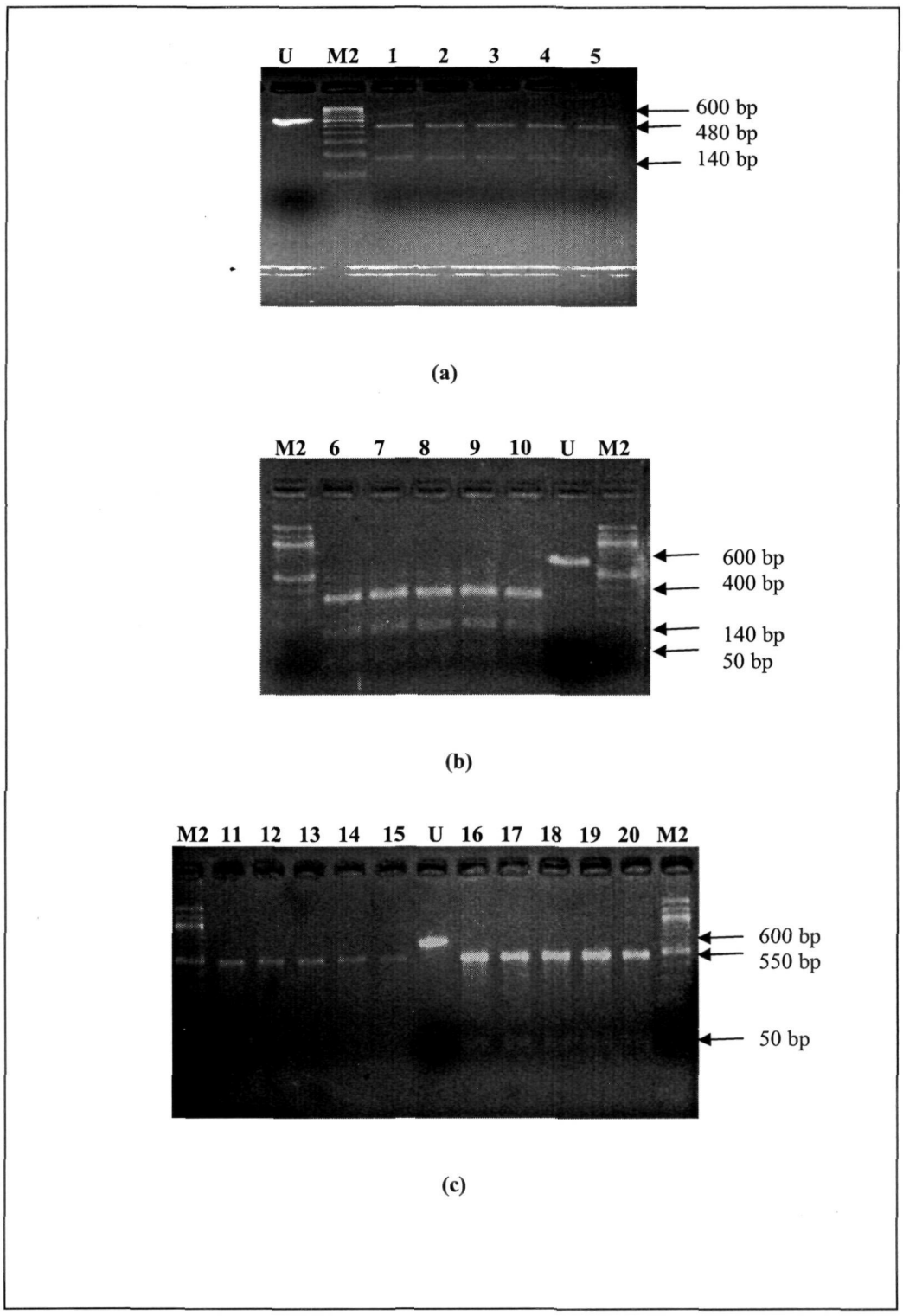
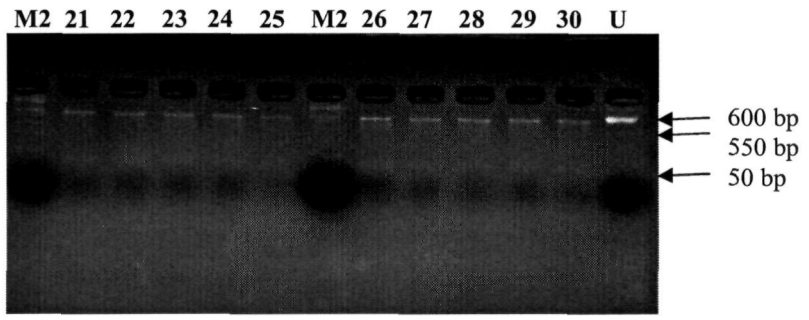
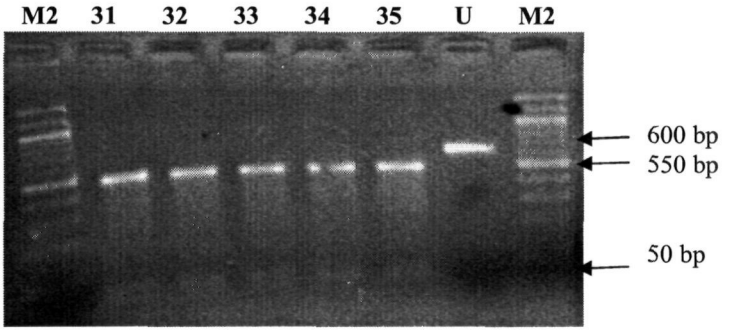


Fig. 4.18 (a-c) PCR-RFLP profile of ITS amplicons using *NciI*.
(M2= 100 bp ladder)



(d)



(e)

Fig. 4.19 (a-c) PCR-RFLP profile of ITS amplicons using *Nci*I.
(M2= 100 bp ladder)

Profiles for *NciI*	Samples	Size of restriction fragments				
		~550	~480	~400	~140	~50
PN1	SKh1, SKh2, Skh3, SKh4, SKh5	-	+	-	+	-
PN2	SS1, SS2, SS3, SS4, SS5	-	-	+	+	+
PN3	SC1, SC2, SC3, SC4, SC5	+	-	-	-	-
PN4	ST1, ST2, ST3, ST4, ST5; SG1, SG2, SG3, SG4, SG5; SKu1, SKu2, SKu3, SKu4, SKu5; SN1, SN2, SN3, SN4, SN5	+	-	-	-	+

Table 4.2: ARP profile with their respective restriction fragments using *Nci*I

Profile	Bands visible in the gel (bp)	Bands produced by webcutter (bp)
PN1	~480	478
	~140	141
PN2	~400	408
	~140	77
		69
	~50	64
PN3	~550	567
	N.V.	38
PN4	~550	ST1-5, no C.S.
		543
		550
		568
	~50	ST1-5, no C.S.
		33
		25
		38

Table 4.3: ARP profile with their respective restriction fragments visible in the gel using *Nci*I and the band sizes generated by webcutter. N.V.- Not visible; C.S.-Cutting site; ST1-5: *Solanum torvum* (1-5)

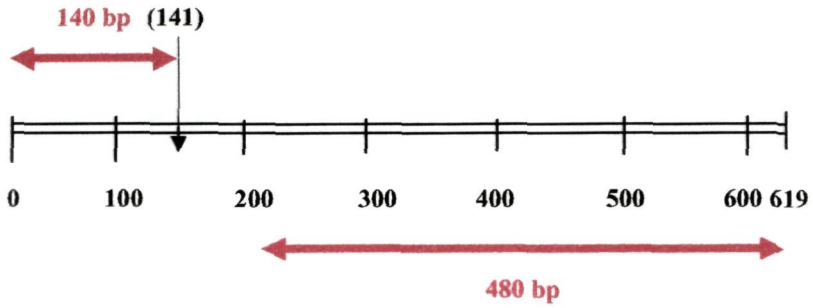


Fig. 4.19 Map of Profile PN1 showing the restriction site using *Nci*I. The red lines show the length of bands as appeared in gel. (Distances not to scale)

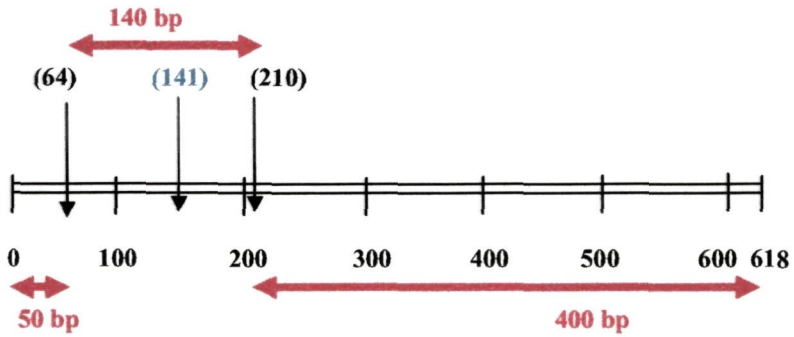


Fig. 4.20 Map of Profile PN2 showing the restriction site using *Nci*I. The red lines show the length of bands as appeared in gel. The blue coloured digit show the point of probable mutation. (Distances not to scale)

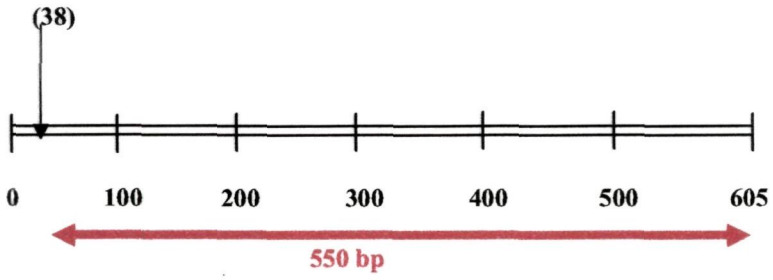


Fig. 4.21 Map of Profile PN3 showing the restriction site using *Nci*I. The red lines show the length of bands as appeared in gel. (Distances not to scale)

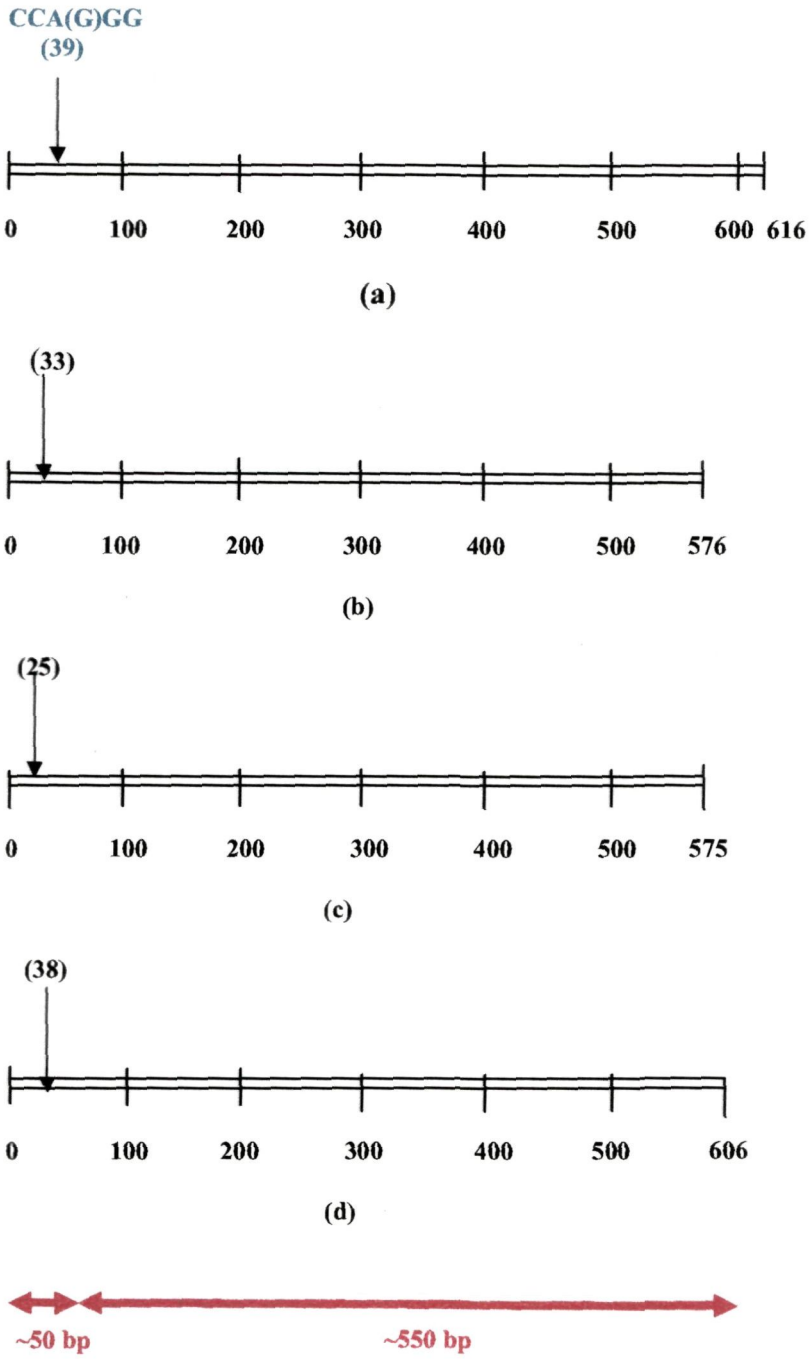


Fig. 4.22 (a-d) Map of Profile PN4 showing the restriction site using *NciI*. The red lines show the length of bands as appeared in gel. a- *S. torvum* with probable mutation site in blue colour. (Distances not to scale)

might have appeared as ~400 bp and ~50 bp respectively in the gel. The computer simulated restriction analysis identified the cutting site at position 141 bp (Fig. 4.20) which would generate two fragments of sizes 77 bp and 69 bp. The sum total of these two bands is 146 bp which is almost similar to the band size of 140 bp which is obtained in the gel. So it can be hypothesized that one of the nucleotide base in this particular recognition site might have altered by mutation due to which the restriction enzyme could not slice in this site. The other reason for obtaining this result might be incomplete digestion. However, we feel it was the former than the latter.

4.7.1.3 Restriction digestion analysis of profile PN3

All the five replicates of *Solanum clavatum* named SC1, SC2, SC3, SC4, and SC5 showed the restriction profile PN3. Computer simulated restriction analysis identified one recognition site for restriction digestion at position 38 and thus two fragments of sizes 567 bp and 38 bp were generated. But the profile PN3 consisted of only one band of size ~550 bp. The computer generated fragment of size 567 bp corresponds to the 550 bp of the gel. The 38 bp long fragment might have disappeared due to the long run of electrophoresis.

4.7.1.4 Restriction digestion analysis of profile PN4

All the five replicates of each *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5), *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5), *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4, and SKu5) and *Solanum nigrum* (SN1, SN2, SN3, SN4, and SN5) showed the restriction profile PN4 when the amplicons of *rrn* ITS region were digested with restriction enzyme *Nci*I. The profile PN4 consisted of two bands of~550 bp and~50 bp respectively.

When the computer simulated restriction analysis was performed using software DS Gene (version 1.1) with the ITS sequence of *Solanum torvum*, the software failed to locate even one restriction site in the whole stretch of the sequence. However, manual check of the sequence base by base revealed a site where substitution of one nucleotide base could have produced restriction site for the enzyme. There is a stretch of nucleotide bases as "5'CCAGG3'" which is similar to the recognition site of *Nci*I except one nucleotide base 'A' at the middle. The nucleotide base 'A' is located at 39th position in the sequence. If this 'A' is substituted by nucleotide base 'G', then this would create a recognition site for the restriction enzyme *Nci*I and this restriction enzyme would slice the DNA at this recognition site and would generate two bands of sizes 577 bp and 39 bp which corresponds to the two bands of sizes ~550 bp and ~50 bp. Therefore, due to the substitution of 'A' by 'G', the restriction enzyme might have sliced the amplicons and as a result two bands of respective sizes were observed in the profile.

The computer simulated restriction analysis of *rrn* ITS sequences of each five replicates of *Solanum gilo*, *Solanum kurzii*, *Solanum nigrum* recognized one recognition site at each ITS sequences at positions 33, 25 and 38respectively. So two bands for each sequence would be generated for *Solanum gilo* (543bp and 33 bp), *Solanum kurzii* (550bp and 25bp) *Solanum nigrum* (568bp and 38bp). Restriction patterns of the *rrn* ITS region using *Nci*I for all the seven species studied showed that this restriction enzyme could successfully differentiate between *Solanum khasianum*, *Solanum sisymbriifolium* and *Solanum clavatum*. But it failed to differentiate between *Solanum torvum*, *Solanum gilo*, *Solanum kurzii* and *Solanum nigrum* as all of these samples generated the same restriction profiles. The restriction

pattern also revealed that there was no intraspecific polymorphism between the species with respect to the ITS region for this particular enzyme.

4.7.2 AMPLICON RESTRICTION PATTERNS (ARP) WITH *Hpa*II

The source of the restriction enzyme *Hpa*II is the bacterium *Haemophilus parainfluenzae*. This is a four base cutter restriction enzyme. The recognition site of this enzyme is "CCGG". The enzyme slices between the two 'C's. *Hpa*II is a methylation sensitive restriction enzyme. So if a DNA strand contains methylated cytosine then this restriction enzyme will not be able to cleave within the same recognition sequence.

Six different profiles were generated when ITS amplicons of the seven species were digested by the restriction enzyme *Hpa*II. The six different profiles were named as PH1, PH2, PH3, PH4, PH5, and PH6 (Table 4.4). All the five replicates of each *Solanum clavatum* (SC1, SC2, SC3, SC4 and SC5) and *Solanum nigrum* (SN1, SN2, SN3, SN4 and SN5) were categorized as profile PH1. Profile PH2 included all the representatives of *Solanum torvum* (ST1, ST2, ST3, ST4 and ST5). The five representatives of *Solanum khasianum* (SKh1, SKh2, SKh3, SKh4 and SKh5) were included in profile PH3. The five representatives of *Solanum sisymbriifolium* (SS1, SS2, SS3, SS4, and SS5) were included in profile PH4. The profile PH5 comprised of all the five replicates of *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4 and SKu5). The five replicates of *Solanum gilo* (SG1, SG2, SG3, SG4 and SG5) were categorized as profile PH6.

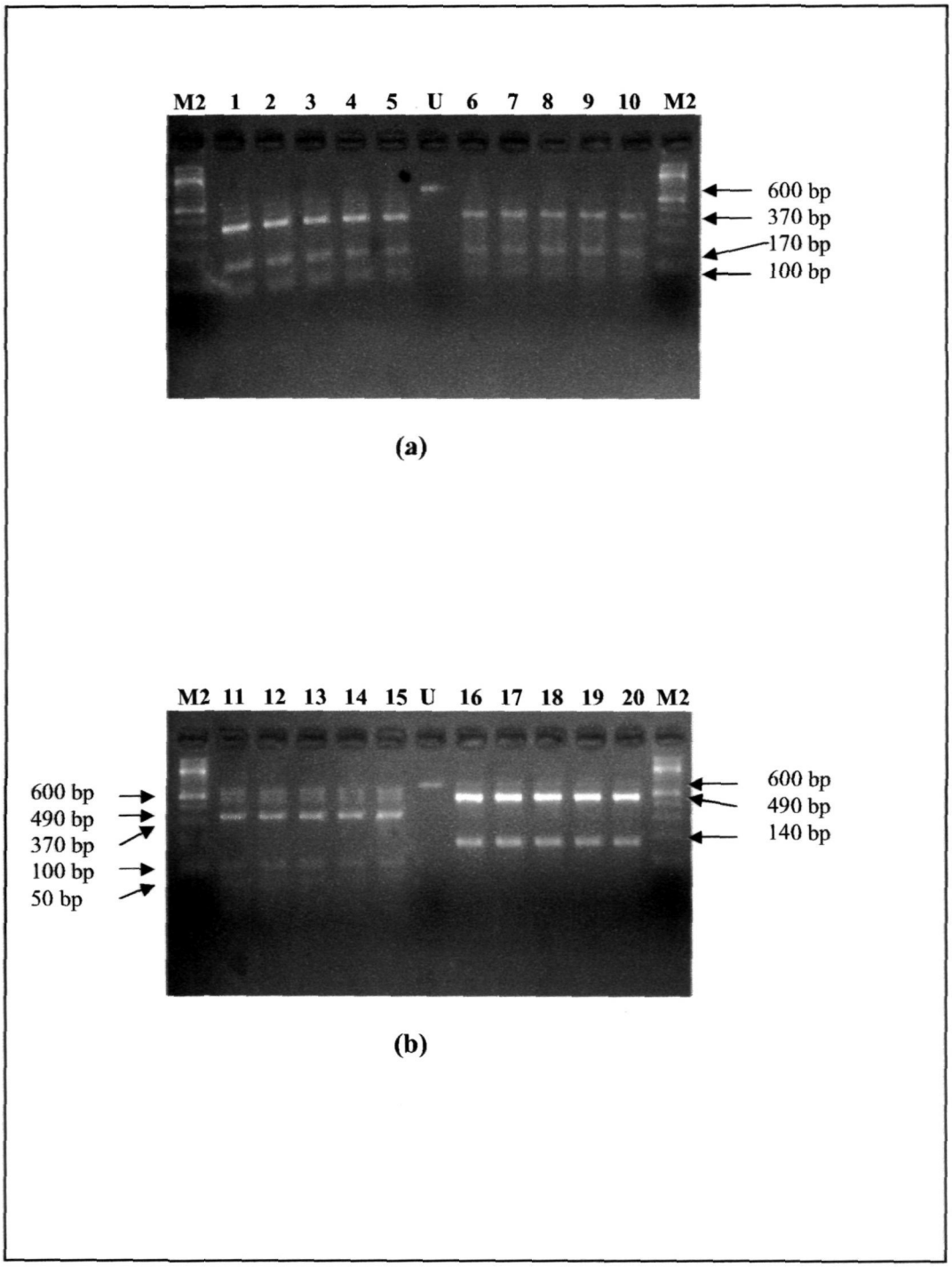


Fig. 4.23 (a-b) PCR-RFLP profile of ITS amplicons using *Hpa*II.
(M2= 100 bp DNA Ladder)

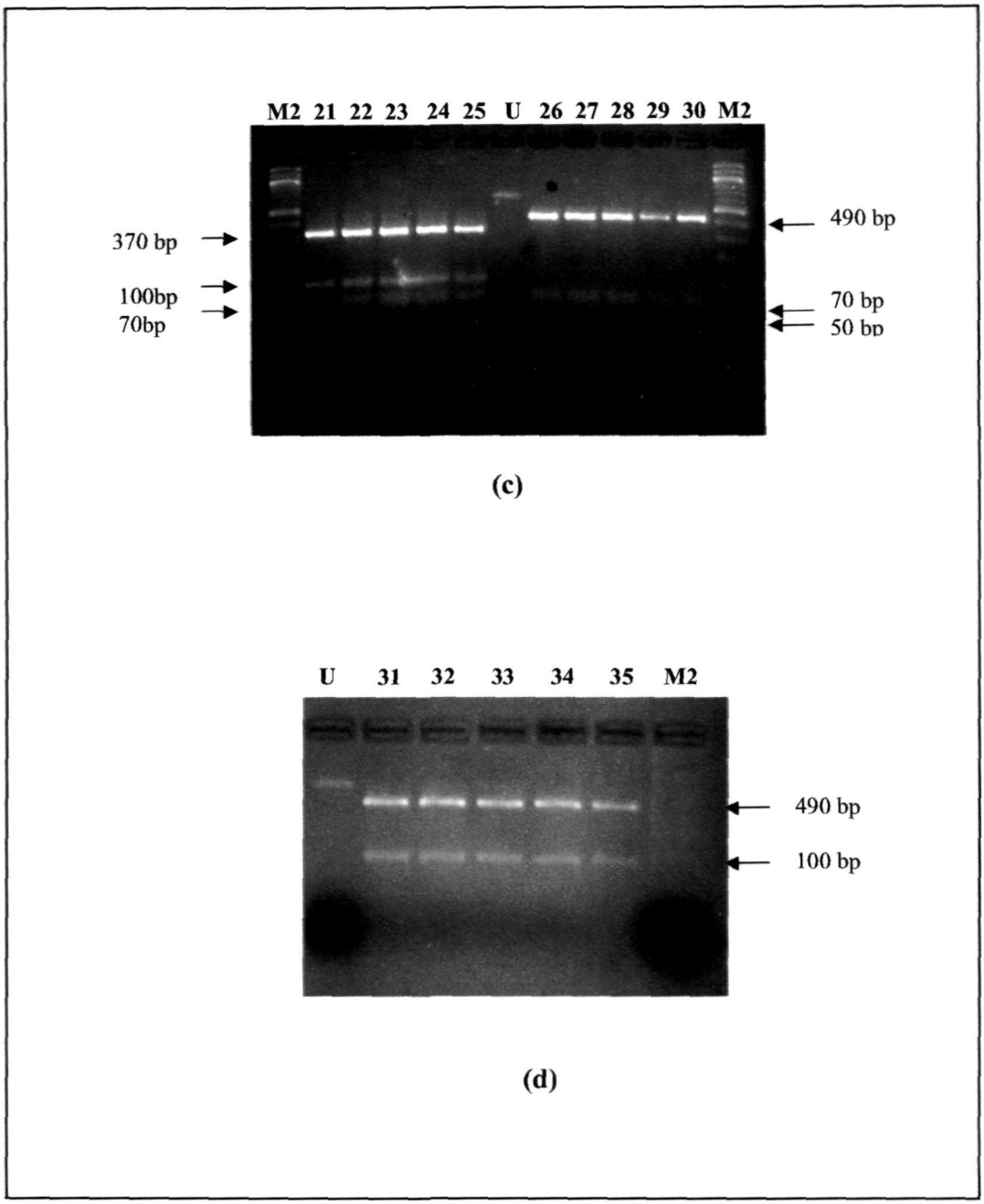


Fig. 4.24 (c-d) PCR-RFLP profile of ITS amplicons using *Hpa*II.
(M2= 100 bp DNA Ladder)

Profiles for *Hpa*II	Samples	Size of restriction fragments						
		~600	~490	~370	~140	~100	~70	~50
PH1	SC1; SC2; SC3; SC4; SC5; SN1; SN2; SN3; SN4; SN5	-	-	+	+	+	-	-
PH2	ST1; ST2; ST3; ST4; ST5	+	+	+	-	+	-	+
PH3	SKh1; SKh2; SKh3; SKh4; SKh5	+	+	-	+	-	-	-
PH4	SS1; SS2; SS3; SS4; SS5	-	+	-	+	-	-	-
PH5	SKu1; SKu2; SKu3; SKu4; SKu5	-	-	+	-	+	+	-
PH6	SG1; SG2; SG3; SG4; SG5	-	+	-	-	-	+	+

Table 4.4: ARP profile with their respective restriction fragments using *Hpa*II

Profile	Bands visible in the gel (bp)	Bands produced by webcutter (bp)
PH1	370	372
	140	136
	100	60
		37
PH2	600	U.D.
	490	416
		71
	370	N.P.
	100	99
	50	30
PH3	600	U.D.
	490	478
	140	141
PH4	490	408
		69
	140	77
		64
PH5	370	376
	100	110
	70	**64**
	N.V.	25
PH6	440	444
	70	99
	N.V.	33

Table 4.5: PCR-RFLP profile with their respective restriction fragments visible in the gel using *Hpa*II and the band sizes generated by webcutter.

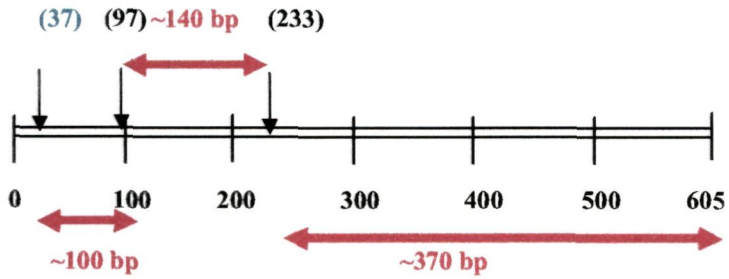


Fig. 4.25 Map of Profile PH1 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. Digit in blue colour shows point of mutation. (Distances not to scale)

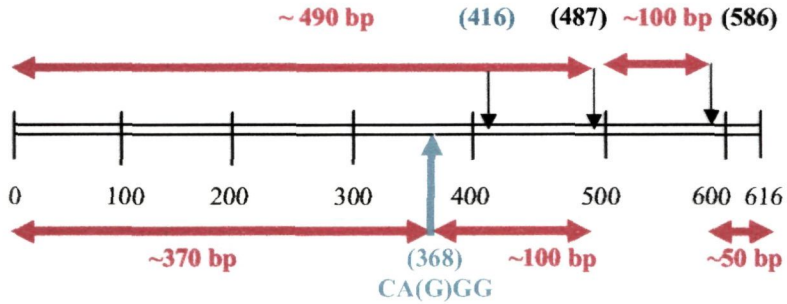


Fig. 4.26 Map of Profile PH2 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. The blue arrow indicates the possible restriction site (Distances not to scale)

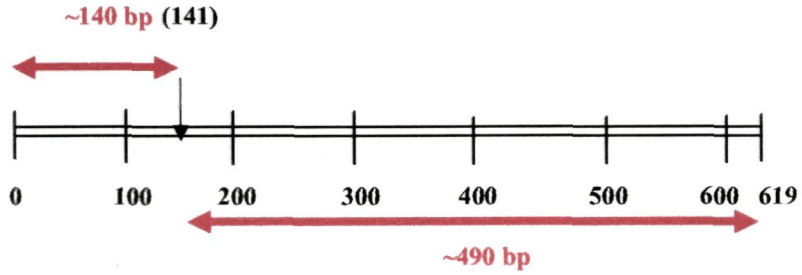


Fig. 4.27 Map of Profile PH3 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

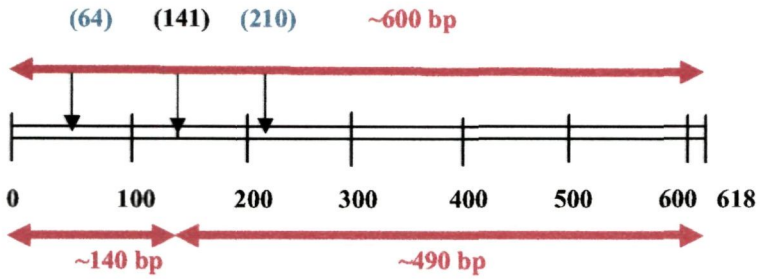


Fig. 4.28 Map of Profile PH4 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

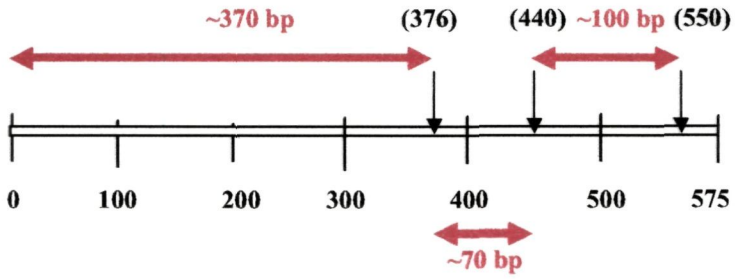


Fig. 4.29 Map of Profile PH5 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

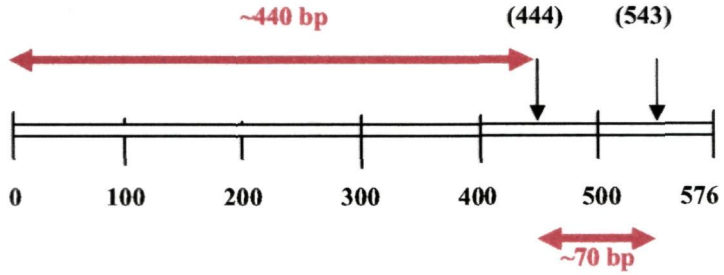


Fig. 4.30 Map of Profile PH6 showing the restriction site using *Hpa*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

4.7.2.1 Restriction digestion analysis of profile PH1

The amplicons of *rrn* ITS region of all the five replicates of each of *Solanum nigrum* (SN1, SN2, SN3, SN4 and SN5) and *Solanum clavatum* (SC1, SC2, SC3, SC4 and SC5) were digested with *Hpa*II restriction enzyme which generated the profile PH1. The profile PH1 comprised of three bands of sizes ~370 bp, ~140 bp and ~100 bp. Computer simulated restriction analysis via DS Gene (Version 1.1) of the sequences of ITS region of both *Solanum nigrum* and *Solanum clavatum* recognized three recognition sites for this restriction enzyme. The cutting sites were located at positions 37, 97 and 233. According to the web cutter, the probable bands should be 37 bp, 60bp, 136 bp and 372 bp respectively. The restriction fragment of 136 bp and 372 bp might have appeared as bands of sizes ~140 bp and ~370 bp respectively in the agarose gel. Instead of two bands of sizes 37 bp and 60 bp there was only one band of size ~100 bp. The sum of the two bands (37 bp + 60 bp) is equal to 97 bp. So, there are two possibilities regarding this difference. Firstly, it might be a case of incomplete digestion where 100 bp band did not get digested by the enzyme. Alternately, one of the nucleotide bases within the recognition sequence of the enzyme might have got mutated due to which the enzyme could not slice the fragment. In that case the 97 bp band might have appeared as ~100 bp in the gel. Since we did a prolonged restriction digestion, the second case seems to be more probable.

4.7.2.2 Restriction digestion analysis of profile PH2

All the five replicates of *Solanum torvum* namely ST1, ST2, ST3, ST4 and ST5 showed the profile PH2 when the ITS amplicons of each samples were digested with *Hpa*II. The profile consisted of 5 restriction bands of sizes ~490 bp,

~370 bp, ~100 bp, ~50 bp along with undigested ~600 bp band. Computer simulated restriction digestion analysis of the ITS sequences using the software DS Gene (Version 1.1) of *Solanum torvum* identified three recognition sites for the restriction enzyme *Hpa*II at positions 416th bp, 487th bp and 586th bp. The restriction bands thus generated according to the computer simulated analysis are of sizes 416 bp, 71 bp, 99 bp and 30 bp respectively (Table 4.5). So it was observed that the bands generated by the actual digestion of the amplicons and computer simulated digestion analysis of the sequences were different from each other. Considering that one base at the 416th position might have undergone mutation due to which the restriction enzyme might have not sliced in that position and thus a single band of size 487 bp could have appeared. The band of size 487 bp appeared as ~490 bp in the gel. Other two bands of sizes 99 bp and 30 bp appeared as ~100 bp and ~50 bp respectively in the agarose gel. However a manual check of the sequence base by base within 1 bp-490 bp revealed a site where the substitution of one nucleotide base could have produced restriction site for the enzyme. There is a stretch of nucleotide bases – "CAGG" at position 367th – 371th bp within which if the nucleotide base 'A' is substituted by 'C' (Fig. 4.26) then the restriction enzyme would identify it as a recognition site and slice at that position due to which another two fragments of sizes 370 bp and 120 bp would be produced. So we can assume that the 490 bp fragment have a cutting site for the restriction enzyme and due to the digestion of 490 bp, the 370 bp band have appeared and 120 bp band have appeared as ~100 bp band along with the undigested ~490 bp fragment. Considering that one of the copies of gene had methylated cytosines in the recognition site at 367^{th} and 377^{th} positions due to which *Hpa*II could not have sliced in that particular recognition site and as a result instead of two bands of sizes ~370 bp and ~100bp, only one band of ~490 bp have

appeared in agarose gel. Absence of recognition sites in one of the copies may have produced the ~600 bp band as observed in the agarose gel. It may be mentioned that this enzyme is methylation sensitive.

4.7.2.3 Restriction digestion analysis of profile PH3

Profile PH3 was showed by all the representatives of *Solanum khasianum* namely SKh1, SKh2, SKh3, SKh4 and SKh5 when the amplicons of ITS region of each of the samples were digested using the restriction enzyme *Hpa*II. The profile showed three bands of sizes ~600 bp, ~490 and ~140 bp. The computer simulated restriction analysis of the ITS sequences of *Solanum khasianum* recognized only one recognition site at 141st position and thus the probable sizes of bands were expected to be 478 bp and 141 bp. This might explain the appearance of ~490 bp and the ~141 bp bands. Absence of recognition site in one of the copies may have produced the ~600 bp band as observed in the gel.

4.7.2.4 Restriction digestion analysis of profile PH4

The amplicons of each of the replicates of *Solanum sisymbriifolium* namely SS1, SS2, SS3, SS4, and SS5 were digested using the restriction enzyme *Hpa* II and the restriction bands obtained were categorized in the profile PH4. Computer simulated restriction analysis of ITS sequences of *Solanum sisymbriifolium* identified three recognition sites for the restriction enzyme *Hpa*II at positions 64th bp, 141th bp and 210th bp respectively. Thus the probable sizes of the bands were expected to be 64 bp, 77 bp, 69 bp and 408 bp. The fragments generated after digestion of the amplicons were ~490 bp and ~140 bp. Considering that one of the nucleotide bases at each of the recognition sites at positions 64th bp and 210th bp have undergone

change, the recognition sites for the restriction enzyme might get altered. As a result only one restriction site position 210th bp would remain. When in such condition the restriction enzyme sliced the amplicons at position 210 bp, two bands of sizes 478 bp and 141 bp would be generated. This might explain the appearance of~490 bp and ~140 bp bands respectively in the agarose gel.

4.7.2.5 Restriction digestion analysis of profile PH5

All the five replicates of *Solanum kurzii* namely SKu1, SKu2, SKu3, SKu4 and SKu5 were digested with the restriction enzyme *Hpa*II and they showed the restriction profile PH5. The profile PH5 comprised of three bands of sizes ~370 bp, ~100 bp and ~70 bp. Computer simulated restriction analysis of the ITS sequences of *Solanum kurzii* recognized three cutting sites at positions 376 bp, 440 bp and 550 bp. So computer generated restriction fragments were expected to be 376 bp, 110 bp, 64 bp and 25 bp. The 25 bp band was not visible in the agarose gel due to it being too smaller it could have been lost during prolonged electrophoresis.

4.7.2.6 Restriction digestion analysis of profile PH6

All the five replicates of *Solanum gilo* namely SG1, SG2, SG3, SG4 and SG5 were digested with the restriction enzyme *Hpa*II and they showed the restriction profile PH6. The profile PH6 comprised of two bands of sizes ~440 bp,~~70 bp. Computer simulated restriction analysis of the ITS sequences of *Solanum gilo* recognized two cutting sites at positions 444 bp and 543 bp which would generate three bands of sizes 444 bp, 99 bp, and 33 bp fragments. The 33 bp sized band was not visible in the agarose gel due to its small size or due to run off in the prolonged electrophoresis.

From the above result it is seen that all the five replicates of each species showed same corresponding type of profiles with same recognition sites. Therefore intraspecific polymorphism was not detected. Interspecific polymorphism could be successfully resolved using *Hpa*II restriction enzyme except for *Solanum nigrum* and *Solanum clavatum* as these two species showed same restriction profiles. From the results we can also say that as *Hpa*II is a methylation sensitive restriction enzyme so it is not a reliable restriction enzyme in terms of barcode generation.

4.7.3 AMPLICON RESTRICTION PATTERNS (ARP) WITH *DdeI/ScrFI*

*Dde*I is a five base cutter restriction enzyme which is isolated from the bacterium *Bacillus stearothermophilus* DE. The recognition site of this enzyme is "CTNAG" and the enzyme slices between 'C' and 'T'. Likewise *Scr*FI is also a five base cutter restriction enzyme. This enzyme is obtained from the bacterium *Bacillus stearothermophilus* SC. The recognition site of this particular restriction enzyme is "CCNGG" and the enzyme slices between 'C' and 'N'. Both these restriction enzymes were used in the same reaction mixture to undergo the double digestion of *mat K* amplicons.

Two different profiles were obtained when *matK* amplicons were digested with *Dde*I/ *Scr*FI. The profiles were named as PDS1 and PDS2. PDS1 profile was generated by all the five replicates of *Solanum khasianum* (SKh1, SKh2, Skh3, SKh4, and SKh5). PDS2 profile was obtained for all other remaining *Solanum spp* studied i.e. *Solanum sisymbriifolium* (SS1, SS2, SS3, SS4, and SS5), *Solanum clavatum* (SC1, SC2, SC3, SC4, and SC5), *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5), *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5), *Solanum kurzii* (SKu1,

SKu2, SKu3, SKu4, and SKu5) and *Solanum nigrum* (SN1, SN2, SN3, SN4, and SN5).

4.7.3.1 Restriction digestion analysis of profile PDS1

The profile PDS1 comprised of bands of sizes ~400 bp,~370 bp,~220 bp,~190 bp and~100 bp. Computer simulated restriction digestion analysis of the *matK* sequences identified total of three recognition sites for both the enzymes i.e. *Dde*I and *Scr*FI. So the expected band sizes generated by the computer simulated restriction digestion were 615 bp, 369 bp, 189 bp and 109 bp. Here 369 bp, 189 bp and 109 bp fragments might have appeared as ~370 bp, ~190 bp and ~100 bp respectively. The difference between the computer simulated restriction digestion and the real restriction digestion of amplicons is the presence of two fragments of sizes ~400 bp, ~220 bp in the digestion profile instead of only one fragment of size 615 bp as generated by computer simulated program. When the sequences of *matK* amplicons were manually monitored carefully, an array of nucleotide sequences were found starting from 1061 bp i.e. "ACCGG". If the nucleotide 'A' would get mutated to 'C', then the nucleotide sequence would be "CCCGG" which is a recognition site of the restriction enzyme *Scr*FI. If the enzyme would slice at this position then the resulting fragments would be of sizes 395 bp and 238 bp. Thus these fragments of sizes 395 bp and 238 bp might have appeared as bands of sizes ~400 bp, ~220 bp respectively in the gel and we feel that the said site was altered in our samples.

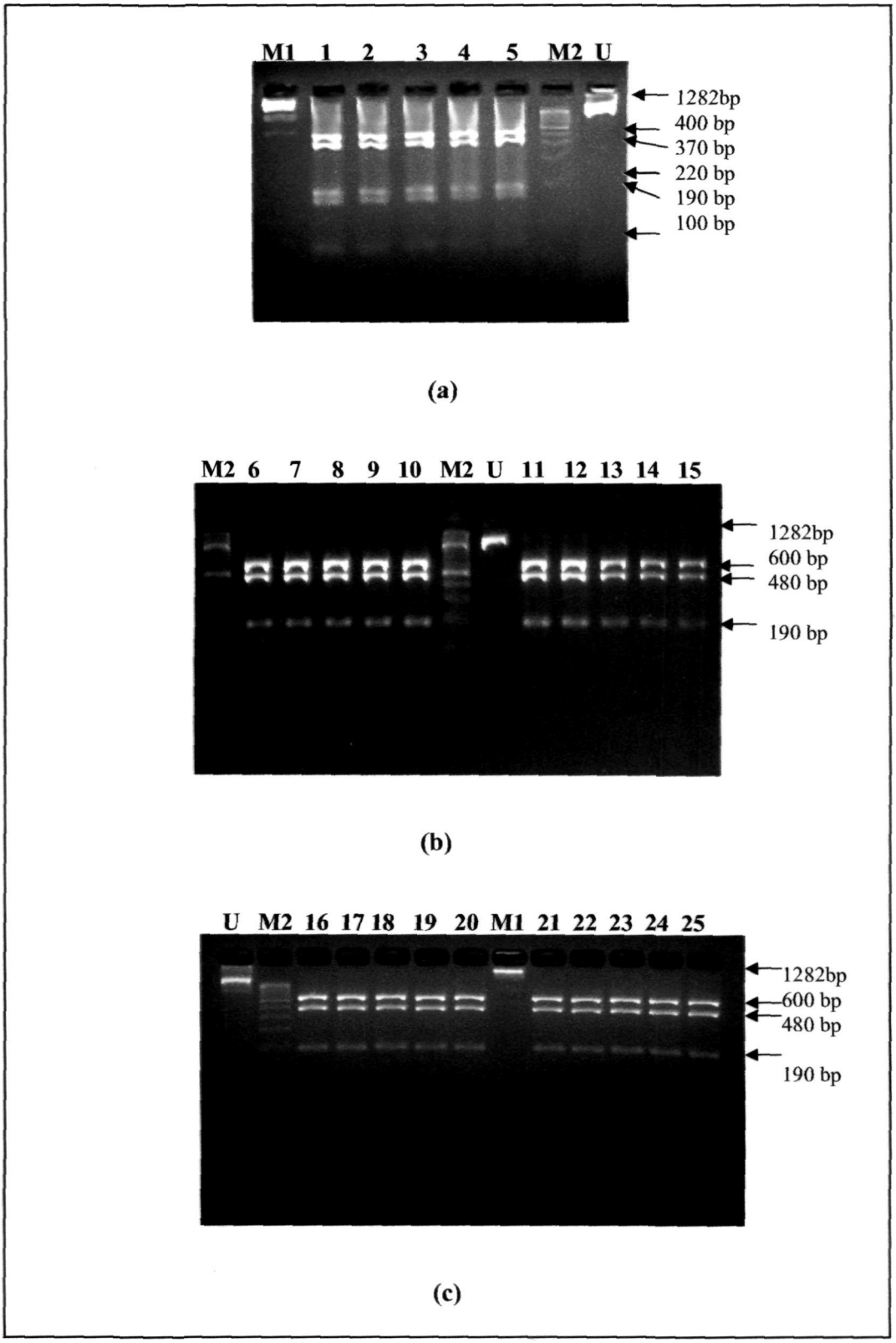


Fig. 4.31(a-c) PCR-RFLP profile of *matK* region using *Dde*I/*Scr*FI.
(M1= λDNA *Hind* III/*Eco*RI double digest marker, M2= 100 bp ladder)

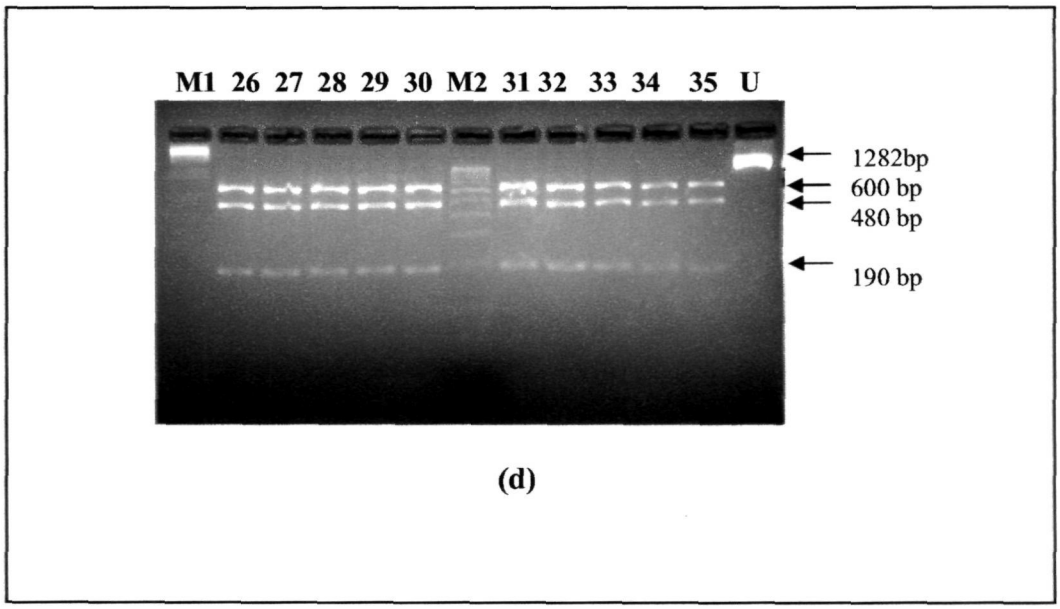


Fig. 4.32(d) PCR-RFLP profile of *matK* region using *Dde*I/*Scr*FI.
(M1= λDNA *Hind* III/*Eco*RI double digest marker,
M2= 100 bp ladder)

Profile for *DdeI/ ScrFI*	Samples	Size of restriction fragments (~ bp)						
		600	480	400	370	220	190	100
PDS1	SKh1, SKh2, SKh3, SKh4, SKh5	-	-	+	+	+	+	+
PDS2	SC1, SC2, SC3, SC4, SC5; SN1, SN2, SN3, SN4, SN5; ST1, ST2, ST3, ST4, ST5; SS1, SS2, SS3, SS4, SS5; SKu1, SKu2, SKu3, SKu4, SKu5; SG1, SG2, SG3, SG4,SG5	+	+	-	-	-	+	-

Table 4.6: PCR-RFLP profile with their respective restriction fragments using *DdeI/ScrFI*

Profile	Bands visible in the gel (~bp)	Bands produced by webcutter (bp)
PDS 1	400	
	220	615
	370	369
	190	189
	100	109
PDS 2	600	605
	480	478
	190	189

Table 4.7: PCR-RFLP profile with their respective restriction fragments visible in the gel using *DdeI/ScrFI* and the band sizes generated by webcutter.

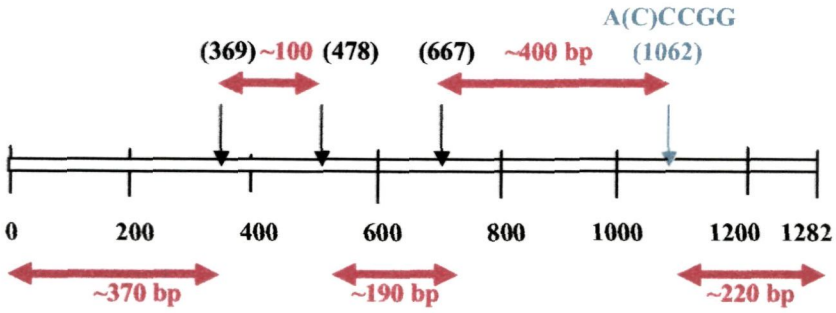


Fig. 4.33 Map of Profile PDS1 showing the restriction site using *Dde*I/ *Scr*FI. The red lines show the length of bands as appeared in gel. Blue arrow indicates the possible restriction site. (Distances not to scale)

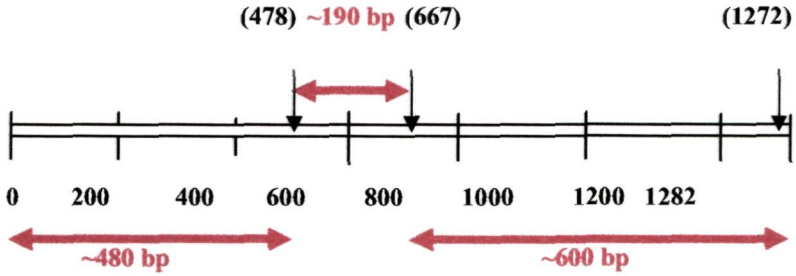


Fig. 4.34 Map of Profile PDS2 showing the restriction site using *Dde*I/ *Scr*FI. The red lines show the length of bands as appeared in gel. (Distances not to scale)

4.7.3.2 Restriction digestion analysis of profile PDS2

The PDS2 profile was generated by all the five replicates of each of *Solanum sisymbriifolium* (SS1, SS2, SS3, SS4, and SS5), *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4, and SKu5), *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5), *Solanum clavatum* (SC1, SC2, SC3, SC4, and SC5), *Solanum nigrum* (SN1, SN2, SN3, SN4, and SN5) and *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5) when the amplicons of *matK* region were digested with restriction enzymes *DdeI/Scr*FI. The fragments generated after the digestion were of sizes ~600 bp, ~480 bp and ~190 bp. Computer simulated restriction digestion analysis of *matK* sequences generated two recognition sites at positions 478th bp and 667th bp for *Scr*FI and *Dde*I respectively. Thus the expected band sizes from the computer simulated restriction digestion are of sizes 189 bp, 478 bp and 615 bp, which were similar to the ones seen in the gel. Intraspecific polymorphism was not monitored in the digestion profile. This double digestion could only differentiate between *Solanum khasianum* from rest of the species studied.

4.7.4 AMPLICON RESTRICTION PATTERNS (ARP) WITH *Ava*II

The restriction enzyme *Ava*II is isolated from *Anabaena variabilis*. This is a five base cutter restriction enzyme. The recognition sequence of this enzyme is "GGWCC" where W is either A or T. The enzyme nicks the DNA strand between the two 'G's.

MatK amplicons of each of the five replicates of all the seven species studied here were digested with restriction enzyme *Ava*II. Four different restriction profiles were observed and were named as PA1, PA2, and PA3. Profile PA1 was observed for all the five replicates of *Solanum nigrum* (SN2, SN3, SN4, and SN5), profile PA2 was shown by all the five replicates of each *Solanum clavatum* (SC1, SC2, SC3, SC4, SC5), *Solanum sisymbriifolium* (SS1, SS2, SS3, SS4, and SS5), *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5), *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5) and *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4, and SKu5) and profile PA3 was generated by all the five replicates of *Solanum khasianum* (SKh1, SKh2, Skh3, SKh4, and SKh5).

4.7.4.1 Restriction digestion analysis of profile PA1

The profile consisted of three bands of sizes ~700 bp, ~390 bp and ~200 bp. The computer simulated restriction digestion of the *matK* sequences showed four bands of sizes 701 bp, 368 bp, 207 bp and 6 bp. Thus the DNA fragments generated by restriction digestion of *matK* amplicons were almost similar with the profile generated by the computer simulated restriction digestion of *matK* sequences. The 6 bp band was not visible in the gel due to its small size.

4.7.4.2 Restriction digestion analysis of profile PA2

Profile PA2 was generated when *matK* amplicons of all the five replicates of each *Solanum clavatum* (SC1, SC2, SC3, SC4, SC5), *Solanum sisymbriifolium*(SS1, SS2, SS3, SS4, and SS5), *Solanum torvum* (ST1, ST2, ST3, ST4, and ST5), *Solanum gilo* (SG1, SG2, SG3, SG4, and SG5) and *Solanum kurzii* (SKu1, SKu2, SKu3, SKu4, and SKu5). When the *matK* amplicons were digested

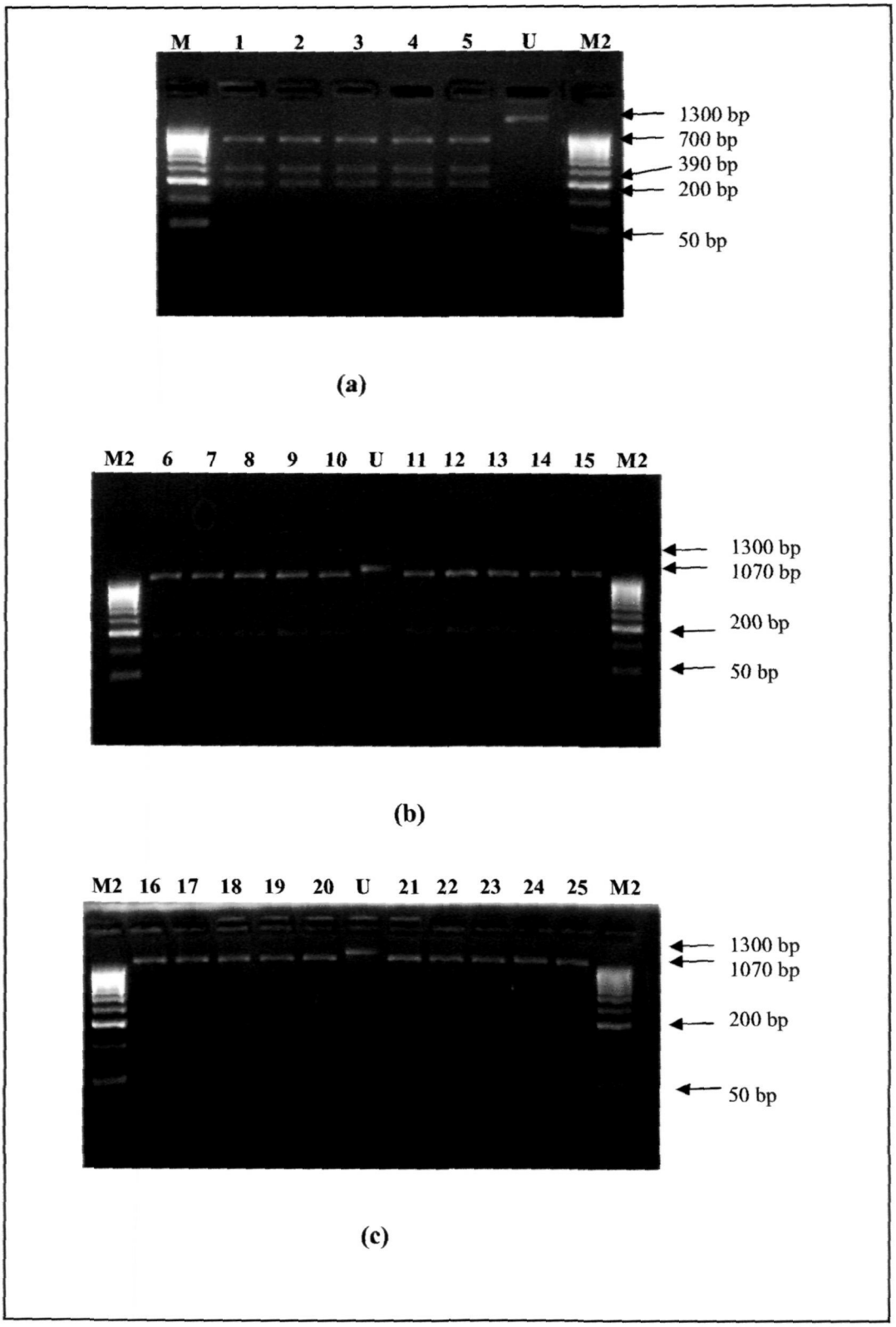


Fig. 4.35(a-c) PCR-RFLP profile of *matK* amplicon using *Ava*II.
(M2= 100 bp ladder)

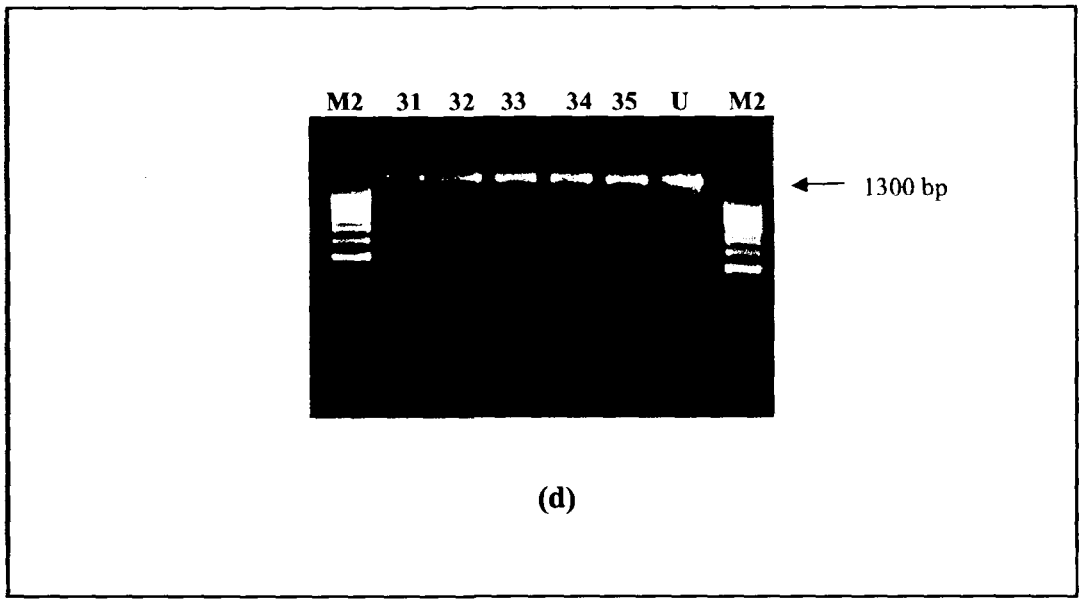


Fig. 4.36(d) PCR-RFLP profile of *matK* amplicon using *Ava*II.

Profile For *Ava* II	Sample	Size of restriction fragments (~ bp)				
		1300	1070	700	390	200
PA1	SN1, SN2, SN3, SN4, SN5	-	-	+	+	+
PA2	ST1, ST2,ST3, ST4, ST5;SS1, SS2, SS3, SS4, SS5;SKu1,SKu2, SKu3, SKu4, SKu5;SG1, SG2, SG3, SG4, SG5; SC1, SC2, SC3, SC4, SC5	-	+	-	-	+
PA3	SKh1, SKh2, SKh3, SKh4, SKh5	+	-	-	-	-

Table 4.8: PCR-RFLP profile with their respective restriction fragments using *Ava*II.

Profile	Bands visible in the gel (~ bp)	Bands produced by webcutter (bp)
PA1	700	→ 701
	390	→ 368
	200	→ 207
	N.V.	→ 6
PA2	1070	→ 1069
	200	→ 213
PA3	1300	→ No cutting site

Table 4.9: PCR-RFLP profile with their respective restriction fragments visible in the gel using *Ava*II and the band sizes generated by webcutter.

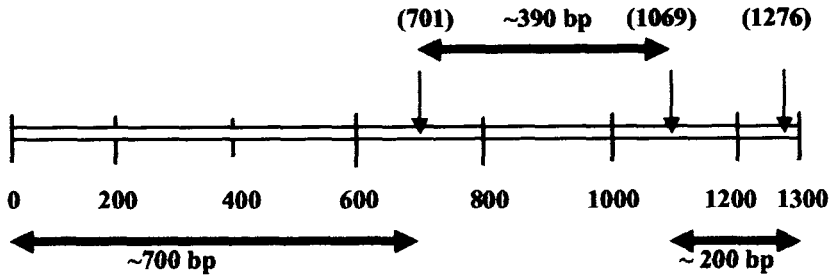


Fig. 4.37 Map of Profile PA1 showing the restriction site using *Ava*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

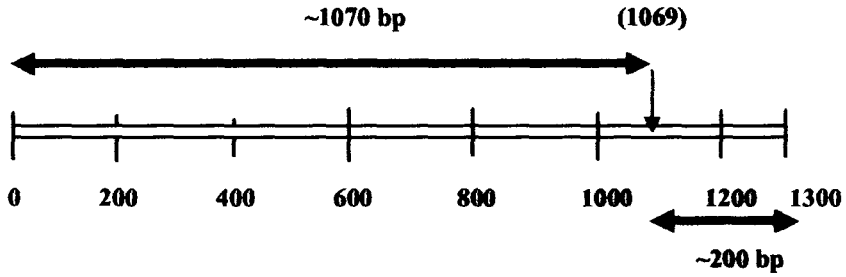


Fig. 4.38 Map of Profile PA2 showing the restriction site using *Ava*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

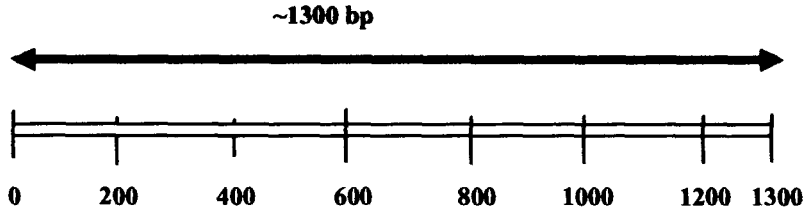


Fig. 4.39 Map of Profile PA3 showing the restriction site using *Ava*II. The red lines show the length of bands as appeared in gel. (Distances not to scale)

with restriction enzyme *Ava*II two bands of sizes ~1070 bp and ~200 bp were generated. The computer simulated restriction digestion of *matK* sequences also showed two bands of sizes 1069 bp and 218 bp. The band of sizes 1069 bp and 218 bp might have appeared as bands of sizes ~1070 bp and ~200 bp respectively in the gel.

4.7.4.3 Restriction digestion analysis of profile PA3

The PA3 profile was shown by all the five replicates of *Solanum khasianum* (SKh1, SKh2, Skh3, SKh4, and SKh5). The *matK* amplicons of this species did not undergone any digestion, due to which the undigested bands of size ~1300 bp were visible in the gel. The computer simulated restriction digestion of the *matK* amplicons also gave the same result as the sequences do not posses any recognition site for the restriction enzyme *Ava*II.

Thus intraspecific polymorphism was not generated when *Ava*II was used. However, this enzyme could successfully differentiate between *Solanum clavatum*, *Solanum nigrum*, *Solanum khasianum*, whereas this enzyme failed to generate polymorphism between *Solanum sisymbriifolium*, *Solanum torvum*, *Solanum gilo* and *Solanum kurzii.*

4.8 CLUSTER ANALYSIS

Cluster dendrogram was constructed by using the ARP profiles of ITS and *matK* amplicons using different restriction enzymes. A single UPGMA tree was constructed by scoring the bands of all the Amplicon Restriction Profiles. The phylogenetic tree showed two major clusters. Cluster I comprised of only *Solanum khasianum* and cluster II comprised of *Solanum sisymbriifolium*, *Solanum torvum*,

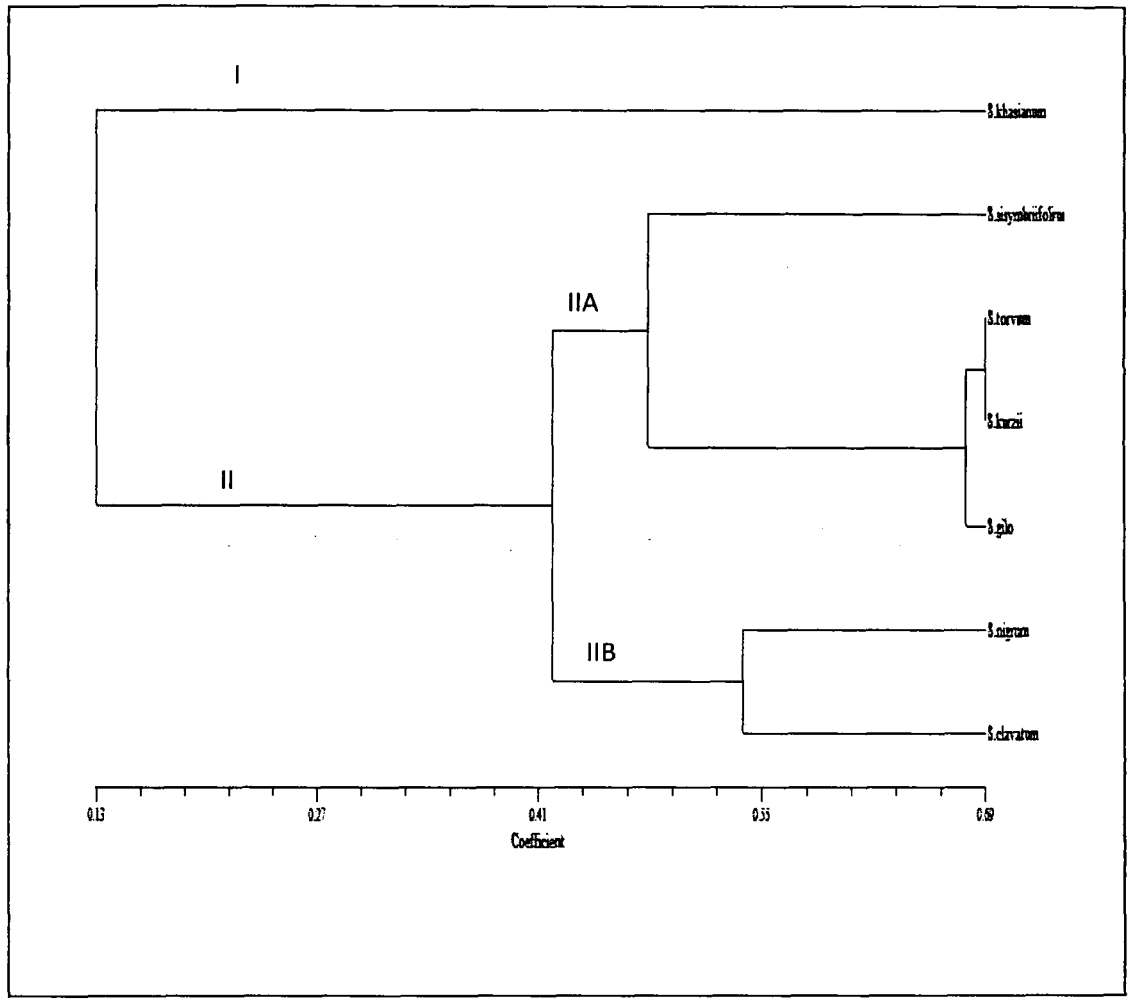


Fig. 4.40 Phylogenetic tree constructed using PCR-RFLP profile.

Solanum kurzii, Solanum gilo, Solanum nigrum and *Solanum clavatum.* Cluster II was subdivided into two major Subclusters and were named as IIA and IIB. IIA comprised of two clades, Clade I included *Solanum sisymbriifolium* and clade II showed clustering of *Solanum torvum, Solanum kurzii, Solanum gilo*. Within clade II *Solanum torvum* and *Solanum kurzii* were associated more closely. *Solanum nigrum* and *Solanum clavatum* clustered together and emerged as the subcluster II in the dendrogram. This phylogenetic tree is almost congruent to the phylogenetic tree constructed using genomic sequences as seen in Fig 4.17, the only major difference is emergence of *Solanum khasianum* as a different branch which was seen as a close relative of *Solanum sisymbriifolium* and clustered together in case of the phylogenetic tree constructed using gene sequences of four genomic regions in combinations. The emergence of *Solanum khasianum* as a separate branch from *Solanum sisymbriifolium* is due to the difference between the Amplicon Restriction Profiles of *Solanum khasianum* and *Solanum sisymbriifolium.* In case of ARP of ITS region of *Solanum khasianum*, we have seen bands of sizes ~600 bp, ~480 bp (Table 4.10) which are absent in case of *Solanum sisymbriifolium.* Again restriction fragments of ~400 bp and ~50 bp have been seen in case of *Solanum sisymbriifolium* which are absent in case of *Solanum khasianum.* The ARP of *matK* region also showed various differences in case of *Solanum khasianum* and *Solanum sisymbriifolium* e.g. restriction fragments of ~1070 bp, ~600 bp, ~480 bp and ~200 bp have been monitored in *Solanum sisymbriifolium* but not in *Solanum khasianum* and restriction fragments of ~400 bp, ~370 bp, 200 bp and ~100 bp are monitored in ARP of *Solanum khasianum* which are absent in *Solanum sisymbriifolium.* The closer association of *Solanum gilo, Solanum kurzii* and *Solanum torvum* is similar in both the dendrograms made of ARP and combined sequences. The only difference is

the clustering of *Solanum gilo* and *Solanum kurzii* indicating their recent diversification in case of phylogenetic tree constructed using combined gene sequences whereas in case of phylogenetic tree constructed using the amplicon restriction profile, *Solanum torvum* and *Solanum kurzii* are clustered together. *Solanum khasianum* has emerged as a separate branch due to the difference in the amplicon restriction patterns from all other species studied. The uncommon restriction fragments seen in case of digestion of the ITS and *matK* regions of *Solanum khasianum* are ~480 bp, ~600 bp and ~400 bp, ~370 bp, ~200 bp, ~100 bp respectively due to which *Solanum khasianum* has emerged as a separate branch.

Thus it is seen that the dendrograms obtained from the combined gene sequences and amplicon restriction patterns were quite similar. The few differences that have been found can be easily explained as the dendrogram made from ARP depends on presence or absence of restriction fragments and generation of restriction fragments again depends on recognition sequences for respective enzymes. Mutation of a single nucleotide may create a recognition site for a particular enzyme or might change a recognition site to a non-recognition site.

4.9 SPECIES SPECIFIC AMPLICON RESTRICTION PROFILE

The restriction fragments that have been generated while digesting the ITS amplicons with *Nci*I and *Hpa*II as a whole are 12 in number and total number of restriction fragments produced while digesting *matK* amplicons with *Ava*II and *DdeI/Scr*FI are also 12. So as a whole there are total 24 restriction fragments monitored. In the table 4.10 all the species with their corresponding banding patterns are tabulated. So from this table it can be seen that each species has a unique restriction fragment profile. From this data it can be said that if these particular

REGION	R. E.	BAND SIZE (~ bp)	NAME OF SPECIES						
			S. khasianum	*S. sisymbriifolium*	*S. torvum*	*S. nigrum*	*S. clavatum*	*S. kurzii*	*S. gilo*
ITS	*Hpa* II	600	+	-	+	-	-	-	-
		490	+	+	+	-	-	-	+
		370	-	-	+	+	+	+	-
		140	+	+	-	+	+	-	-
		100	-	-	+	+	+	+	-
		70	-	-	-	-	-	+	+
		50	-	-	+	-	-	-	+
	Nci I	550	-	-	+	+	+	+	+
		480	+	-	-	-	-	-	-
		400	-	+	-	-	-	-	-
		140	+	+	-	-	-	-	-
		50	-	+	+	+	-	+	+
MatK	*Ava* II	1300	+	-	-	-	-	-	-
		1070	-	+	+	-	+	+	+
		700	-	-	-	+	-	-	-
		390	-	-	-	+	-	-	-
		200	-	+	+	+	+	+	+
	Dde I/*Scr*FI	600	-	+	+	+	+	+	+
		480	-	+	+	+	+	+	+
		400	+	-	-	-	-	-	-
		370	+	-	-	-	-	-	-
		200	+	-	-	-	-	-	-
		190	+	+	+	+	+	+	+
		100	+	-	-	-	-	-	-

Table 4.10: Total amplicon restriction profile.

primers are used for amplification of the corresponding genes or DNA regions and the amplicons generated are digested with the same restriction enzymes used in the study, then the restriction pattern that would generate can be used for the identification of the species. So each amplicon restriction pattern corresponding to each species can be used as barcode for identification of the particular species.

4.9 DISCUSSION

Among all the trees constructed, the tree obtained from Bayesian analysis of the four combined datasets (Fig. 4.17) was most resolved. All the wild solanums studied here fell under three monophyletic clades. The numbering of clades as clade I, clade II and clade III are based on the position of the clades as basal, middle and terminal positions respectively. Each clade corresponds to three subgenera which is congruent to that described by Olmstead & Palmer (1992). Here clade I includes representatives of subgenus *Potatoe*, clade II includes members of subgenus *Solanum* and clade III includes representatives of subgenus *Leptostemonum*. This result dovetails with that of Olmstead & Palmer (1997) where they reported three major clades within *Solanum* based on chloroplast DNA restriction site variation but only difference in their result was that the clade I comprised of representatives from four subgenera – *Archaesolanum*, *Potatoe*, *Minon*, and *Solanum* whereas there is only one subgenus recognized in the present study. The members of subgenera *Archaesolanum*, *Minon* are not sampled in this study. This difference in presence of only one subgenus in this result is due to the small sample size. The only difference between the results of present study and that of Olmstead & Palmer (1997) is that in our result clade II comprised of subgenus

Solanum and clade III comprised of subgenus *Potatoe* which is vice versa in case of Olmstead & Plamer (1997).

The subgenus *Leptostemonum* is large morphologically distinctive group (ca. 250-450 species) within *Solanum* comprising almost one third of the genus (D'Arcy, 1972, 1991). Members of this group are characterized by presence of spines and stellate hairs. 20 sections within subgenus *Leptostemonum* were reported by Whalen (1984). Levin *et al.* (2006) have defined 10 clades within subgenus *Leptostemonum* based on two nuclear regions (ITS and granule bound starch synthase gene i.e. GBSSI or *waxy*) and one chloroplast spacer region (*trnS-trnG*). To make this study comparable nomenclature correction for two species is necessary: *Solanum aethiopicum* in studies of Olmstead & Palmer (1997) is same with *Solanum gilo* of the present study as the later is the cultivar group of the former and *Solanum khasianum* is the synonym of *Solanum aculeatissimum* in the study of Levin *et al.* (2006). In our study the representatives of the subgenus *Leptostemonum* are *Solanum kurzii*, *Solanum sisymbriifolium*, *Solanum gilo*, *Solanum torvum, Solanum khasianum*, and these species are member of the single monophyletic clade III. In the present study *Solanum gilo* and *Solanum kurzii* are clustering together with 100% posterior probability (Fig. 4.17) within clade III and both fall under the sectionion *Oliganthes* which is also placed in the same clade of sectionion *Oliganthes* as reported by Olmstead & Palmer (1997). The leaves of these two species are not prickly but have stellate hairs (Isshiki *et al.*, 2008). In the studies of Levin *et al.* (2006), *Solanum sisymbriifolium* was placed in the sectionion *Androceras/Crinitum* supported by a posterior probability of 97% but this analysis was not supported in the combined parsimony analysis, where *Solanum sisymbriifolium* occupied an

unresolved position within *Leptostemonum*. In the present study we have seen that *Solanum sisymbriifolium* was clustered with *Solanum khasianum* (Fig.4.17). The fully resolved tree resulted from the combined data matrix (Fig.4.17) supported the close relationship between these two species with high posterior probability of 100%. Levin *et al.* (2006) placed *Solanum aculeatissimum* (*Solanum khasianum*) in the subgenus *Leptostemonum* as a representative of the sectionion *Acanthophora*. Levin *et al.* (2005) have extensively studied the sectionion *Acanthophora* on basis of two nuclear regions (ITS and the granule-bound starch synthase gene [GBSSI or *waxy*]) and two chloroplast regions (*trnT-trnF* and *trnS-trnG*) and they found that the sectionion *Acanthophora* was not monophyletic. In the present study clustering of *Solanum sisymbriifolium* (section. *Androceras/ Critinum*) with *Solanum khasianum* (section. *Acanthophora*) together formed a subclade within the subgenus *Leptostemonum* which also does not support the monophyly of sectionion *Acanthophora*. *Solanum torvum* emerged as a separate branch in the subgenus *Leptostemonum* indicating it as a separate sectionion. Olmstead & Palmer (1997) and Bohs (2005) placed *Solanum torvum* within the sectionion *Torva* of subgenus *Leptostemonum*.

D'Arcy (1972, 1991) reported the subgenus *Solanum* with 200 species within it and a worldwide distribution. Bohs (2005) placed this sectionion in the *Morelloid* clade in his 12 major clades within *Solanum* on basis of *ndhF* sequences. Spooner *et al* (1993) placed *Solanum nigrum* at the base of the lineage leading to clade I. Olmstead & Palmer (1997) placed *Solanum ptychanthum,* which is related to *Solanum nigrum,* in the subgenus *Solanum* leading to clade I. Poczai & Hyvonen (2011) stated that much less was known regarding the relationships of the species

Solanum nigrum. In the present study *Solanum nigrum* grouped with *Solanum clavatum* with 100% posterior probability values (Fig. 4.17). Both these species formed a monophyletic clade II, which suggests their close evolutionary relationship and their recent diversification. From these results it can be suggested that *Solanum clavatum* may share the same subgenus with *Solanum nigrum*.

In all the phylogenetic trees constructed (Fig.4.11-4.17) *Solanum lycopersicum* has taken the basal position as a separate lineage. D'Arcy (1972, 1991) placed *Solanum lycopersicum* in the sectionion *Lycopersicon* under the subgenus *Potatoe*. Olmstead & Palmer (1997) also placed *Solanum lycopersicum* in the same position as D'Arcy (1972, 1991). In fig. 4.11 and fig. 4.15 it is seen that *Solanum lycopersicum* and *Solanum tuberosum* are grouped together, but their association is supported by poor BS in case of *18S rRNA* data set and moderate BS in case of *matK* data set. So the strength of the evolutionary relationship between these two cultivated species remains unresolved in the present study. But D'Arcy (1972, 1991) and Olmstead & Palmer (1997) placed *Solanum tuberosum* and *Solanum lycopersicum* together in the same subgenus *Potatoe*. *Solanum tuberosum* is the member of the sectionion *Petota*. The monophyly of the clade *Potatoe* is not properly resolved in the present study.

4.10 BIOCHEMICAL ANALYSES

Biochemical analysis together with molecular studies provides more insight into evolutionary history of a species. It helps in assessing the environmental effects on a particular genotype. Comparison among different species on basis of different biochemical parameters helps in understanding the diversity of respective species in relation to each other.

4.10.1 PROTEIN PROFILING

One of the biochemical methods extensively used in taxonomy and assessment of genetic diversity is the protein profiling. The relevant method for protein profiling is electrophoretic analysis of total proteins by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). Keeping in mind that total protein profiling would provide more information about the genetic diversity and phylogenetic relationship among the species, we have included this parameter in our study.

Total protein was isolated and was run in SDS-PAGE (Fig. 4.39). Protein bands generated by SDS-PAGE were scored. Total 28 polymorphic bands were observed. UPGMA tree was constructed using the programme NTSys.

The tree (Fig. 4.40) generated by using the total protein profile was also almost congruent to the tree constructed using gene sequences in combination. In this phylogenetic tree *Solanum khasianum* and *Solanum sisymbriifolium* grouped together showing their recent diversification. *Solanum gilo*, *Solanum kurzii* and *Solanum torvum* clustered together proving that they are closely related and among

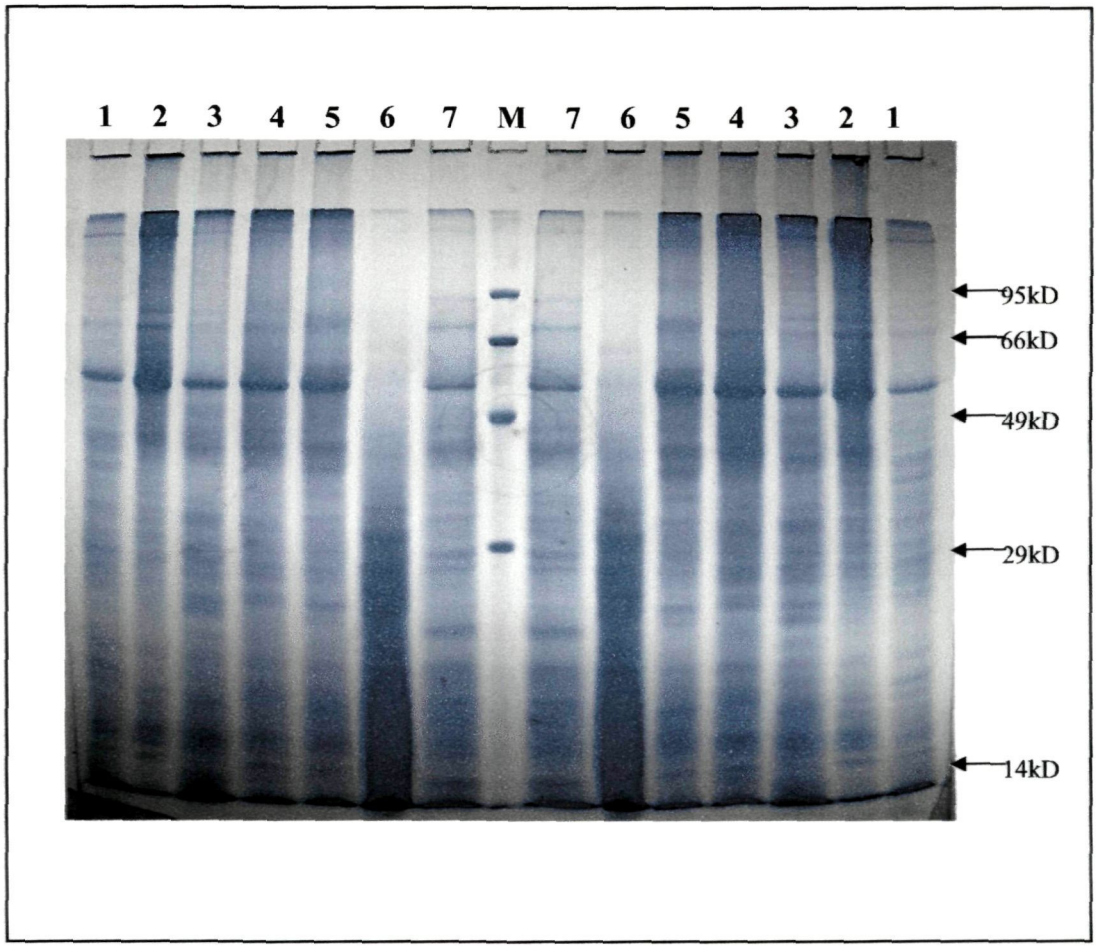


Fig. 4.41 SDS-PAGE generated total protein profile.

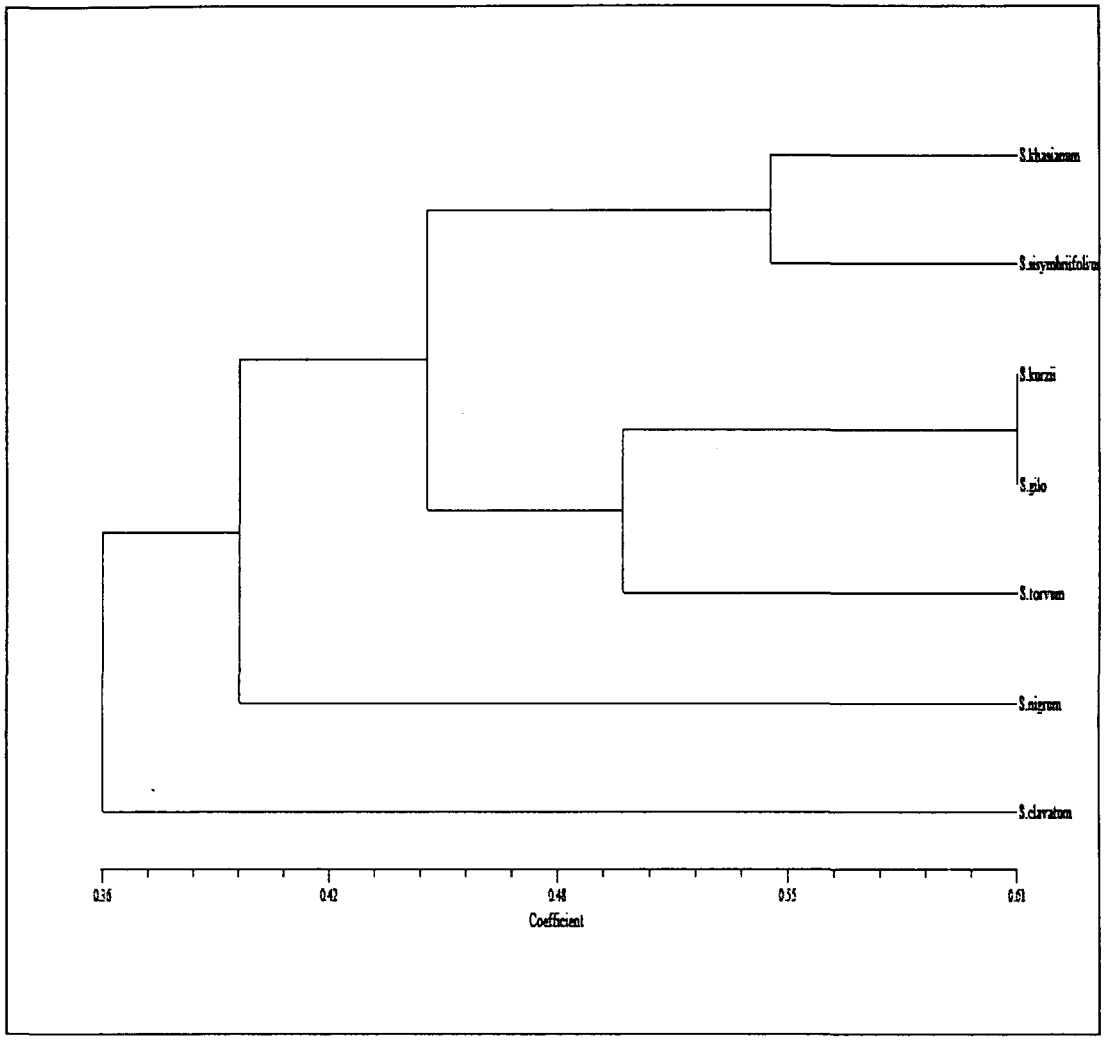


Fig. 4.42 Phylogenetic tree constructed using total protein profile.

which *Solanum gilo* and *Solanum kurzii* are more recently diversified from each other. *Solanum nigrum* and *Solanum clavatum* emerged as separate branches which is contradictory to the result obtained from combined gene sequence analyses. This result is quite expectable as the production of proteins depends also on environmental conditions other than the genetic makeup of the plant.

4.10.2 PHYTOCHEMICAL SCREENING

Many of the secondary metabolites, such as alkaloids, terpenoids, flavonoids etc. are the characteristics of many plants and these compounds are responsible for pharmacological properties of a particular plant species. The genus *Solanum* also includes many species with pharmacological importance. In this study, qualitative phytochemical screening was carried out in all the seven species of *Solanum* to ascertain the presence or absence of secondary metabolites such as alkaloids, flavonoids, tannins, saponins etc.

The results of phytochemical screening of the fruits of the wild *Solanums* showed the presence of many of the important secondary metabolites like- alkaloids, terpenoids, flavanoids, cardiac glycosides, tannins and saponins (Table.4.11). Alkaloids were present copiously in all the species studied here. Terpenoids were present copiously in *Solanum khasianum* and *Solanum sisymbriifolium* and moderately in *Solanum clavatum* and *Solanum gilo*. Only marginal quantities were found in *Solanum nigrum* and *Solanum torvum*. They were completely absent in *Solanum kurzii*. In case of flavanoids it was seen that *Solanum khasianum* and *Solanum torvum* contained copious quantities, while *Solanum nigrum*, *Solanum clavatum*, *Solanum kurzii*, *Solanum sisymbriifolium* contained moderately and *Solanum gilo* contained very low quantities. One of the most

important secondary metabolite called cardiac glycoside was found present adequately in *Solanum khasianum*, while in *Solanum clavatum* and *Solanum kurzii* it was present moderately. Among the plants studied here which contained trace amounts of cardiac glycosides are *Solanum sisymbriifolium* and *Solanum torvum* while *Solanum nigrum* and *Solanum gilo* did not contain the same. In case of tannins *Solanum khasianum* and *Solanum torvum* were found to be source of adequate amount of tannins. *Solanum nigrum*, *Solanum kurzii*, *Solanum gilo* and *Solanum sisymbriifolium* contained moderate quantities of tannins while *Solanum clavatum* showed the presence of tannins copiously. *Solanum kurzii* contained saponins copiously while *Solanum gilo*, *Solanum khasianum* and *Solanum sisymbriifolium* contained tannins moderately. *Solanum nigrum*, *Solanum clavatum* and *Solanum torvum* contained only trace s of tannins. Thus among the seven species studied *Solanum khasianum* is a rich source of many valuable secondary metabolites studied here.

Phytochemicals	*S. nigrum*	*S. clavatum*	*S. kurzii*	*S. gilo*	*S. khasianum*	*S. sisymbriifolium*	*S. torvum*
Alkaloids	+++	+++	+++	+++	+++	+++	+++
Terpenoids	+	++	-	++	+++	+++	+
Flavonoids	++	++	++	+	+++	++	+++
Cardiac glycosides	-	++	++	-	+++	+	+
Tannins	++	+	++	++	+++	++	+++
Saponins	+	+	+++	++	++	++	+

Table 4.11: Phytochemical screening of *Solanum spp.* fruits.
+++ (copiously present), ++ (moderately present), + (slightly present/trace), - (absent)

4.10.3 QUANTIFICATION OF GLYCOALKALOID SOLASODINE

Solasodine has potential as a substitute for diosgenin in the semi-synthetic production of steroidal hormones for use in pharmaceuticals. Therefore, steroidal glycoalkaloid from Solanaceae plants have become increasingly important as a starting material for the production of steroidal hormone. Because of the high demand of solasodine, it is very important to ascertain the presence of solasodine in the species studied here and quantification of solasodine is also a very important component of a strategy to improve yield.

Presence of glycoalkaloid solasodine in all the seven species of *Solanum* was assessed by running the plant extracts in Thin Layer Chromatography (TLC) by the procedure mentioned in Chapter 3. Presence of alkaloid in the TLC plate is detected by using Dragendorff's reagent. If alkali is present then the spot coloured orange colour. Pure solasodine was also run together with the extract for detection of the solasodine spot in the extract (Fig.4.40). The Rf value of the pure solasodine spot was 0.46. Among different spots in the extract lane one particular spot coincide with the standard spot with the same Rf value 0.46. So, it can be confirmed that the particular spot corresponds to the compound solasodine.

Quantification of solasodine was performed by method standardised by Lancaster and Mann (1974) via colorimetric determination. The colorimetric quantification of the glycoalkaloid solasodine in the fruit powders of all the seven plant species studied here revealed that one gram dried fruit powder of each *Solanum nigrum*, *Solanum kurzii*, *Solanum clavatum*, *Solanum gilo*, *Solanum khasianum*, *Solanum sisymbriifolium* and *Solanum torvum* contained 75.575±1.43 µg, 60.725±0.72 µg, 65.615±1.09 µg, 60.61±1.03 µg, 90.83±0.59 µg, 89.34±0.64 µg and

87.17±1.05 µg of solasodine respectively. Therefore, it seems that fruits of *Solanum khasianum* contain significantly higher amount of solasodine in comparison to the other species studied (Table 4.12). *Solanum sisymbriifolium* can be placed after *Solanum khasianum* in terms of amount of solasodine present in one gram of fruit powder. *Solanum kurzii* and *Solanum gilo* contained least amount of solasodine in comparison to other species studied.

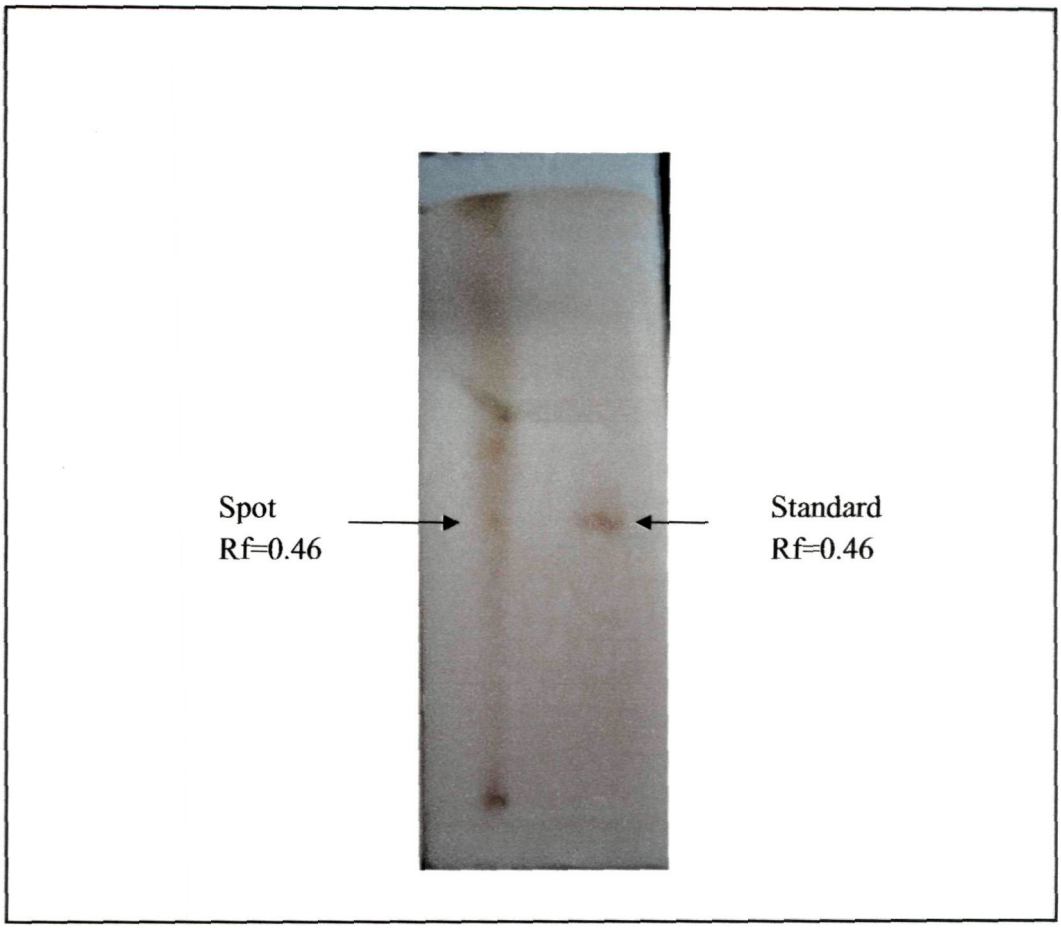


Fig. 4.43 TLC plate showing solasodine spot.

Species name	*S. nigrum*	*S. kurzii*	*S. clavatum*	*S. gilo*	*S. khasianum*	*S. sisymbriifolium*	*S. torvum*
Concentration (µg/gm)	75.575±1.43[b]	60.725±0.72[a]	65.615±1.09[b]	60.61±1.03[a]	90.83±0.59[b]	89.34±0.64[b]	87.17±1.05[b]

Table 4.12: Concentration of Solasodine.

Mean values (mean±SD) sharing 'a' donot differ significantly and sharing 'b' differ significantly by Tukey's test at $p \leq 0.05$.

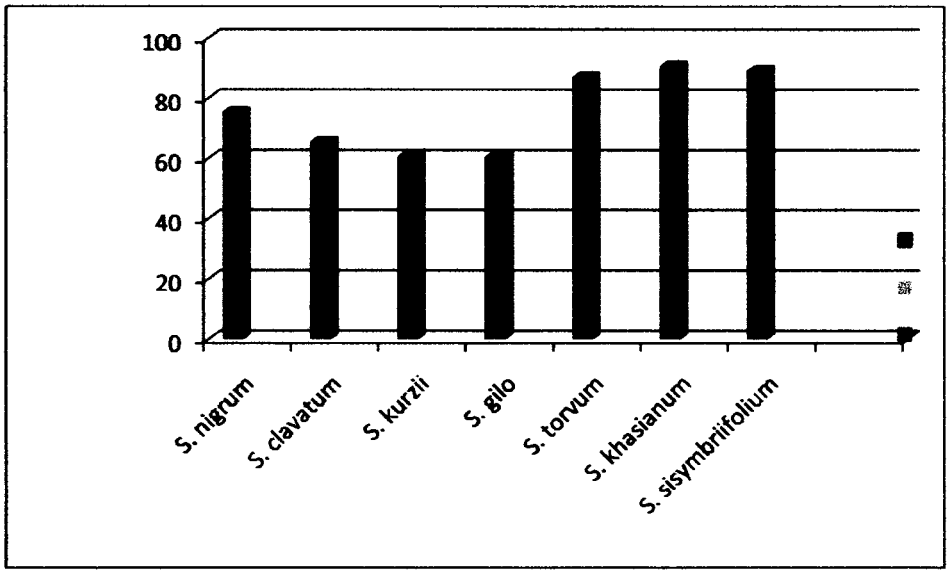


Fig. 4.44 Graphical representation of concentration of Solasodine in each species.

4.11 DISCUSSION

Qualitative phytochemical screening of all the species proves the presence of alkaloids, saponins, terpenoids, flavonoids, tannins etc. Presence of these phytochemicals is indicative of the therapeutical importance of these plants and it can be said that these plants have the potential of providing precursors for the synthesis of useful drugs. Similar phytochemical analyses was conducted by Chindeu *et al.* (2011) in two indigenous eggplants of Africa, *Solanum aethiopicum* and *Solanum macrocarpon*. According to their results fruits of these two plants contain significant amount of alkaloids, saponins, flavonoids, tannins and ascorbic acid; terpenoids were present in trace amounts. They found that *Solanum aethiopicum* contains higher level of beneficial agents than *Solanum macrocarpon*. *Solanum gilo* which is synonym of *Solanum aethiopicum*, also contains almost all the beneficial agents except for cardiac glycosides in our study. Cardiac glycoside is also seem to be absent in *Solanum nigrum* in the present study. It has been said that most of the observed effects of eggplants are due to their phytochemical contents. Bitterness of eggplants is due to the presence of alkaloids, mainly glycoalkaloids. According to Vohora *et al.* (1984) alkaloidal extracts of *Solanum melongena* leaves showed analgesic effects and some CNS depression. Saponins found in the fruits are known to exhibit antimicrobial activities and protect plants from microbial pathogens (Sczkowski *et al*, 1988). Asl and Hossein (2008) showed that saponins present in traditional medicine preparations cause hydrolysis of glycoside from terpenoid to avert the toxicity associated with the intact molecule. All the seven species contain more or less equal amount of flavonoids which is an effective antioxidant (Bagehi,

1999 and Vinson, 1998). Sudheesh (1997) showed that flavonoids extracted from fruits of *Solanum melongena* showed significant hypolipidemic actions in normal and cholesterol fed rats. *In-vitro* studies have also shown that flavonoids have anti-allergic, anti-inflammatory, anti-microbial and anti-cancer activities (Cushine and Lamb, 2005; Yamamoto and Gaynor, 2001). *Solanum* have antiviral, anticancer, anticonvulsant and anti-infective effects due to phytochemicals they contain (Gbile and Adesina, 1988).

Among the glycoalkaloids present in *Solanum*, solasodine is always a topic of interest for researchers due to its medicinal use in steroidal precursor for the supplementary source of the commercial synthesis of several steroidal drugs (Rodriguez *et al.*, 1979; Sree *et al.*, 1982). According to Crabbe and Fryer (1982), solasodine can be primarily obtained from various plants of genus *Solanum*. Nearly 264 species of *Solanum* have so far been investigated for solasodine and about 100 of these are reported to possess it in varying quantities (Mann, 1978). Isolation procedure of solasodine has been reported from berries, leaves or stems of *Solanum marginatum* L. (Guerrero, 1976), *Solanum asperum* Vahl. (Bhattacharyya, 1984), *Solanum paludosum* Moric (Bhattacharyya, 1984), *Solanum eleagnifolium* Cav. (Rodriguez *et al.*, 1979), and *Solanum laciniatum* Ait. (Kangaroo apple), an indigenous vegetable plant of Australia and New Zealand (Conner, 1987). Perez-Aador (2007) made a comparison between plants of *Solanum torvum* that grow in Chiapas, Mexico, and plants of the same species originating from India. They tried to establish either similarities or differences between these plants in total alkaloid contents and presence of solasodine. The total alkaloid content (0.12%) of the plants coming from Chiapas and India was similar, but they found solasodine only in the

plants of Chiapas. But this result of Perez-Aador (2007) is contradictory to our result as we have confirmed the presence of solasodine in *Solanum torvum* and it was quantified that one gram of fruit powder of *Solanum torvum* contains 87.17±1.05 µg of solasodine. A comparative study was carried out by Eltayeb *et al.* (1997) regarding changes in solasodine during development of *Solanum nigrum* and *Solanum incanum*. According to them all tested organs of both the species elaborated solasodine, but the levels varied widely. In both species, the smallest leaves showed the highest alkaloid concentration. Maximum levels in *Solanum incanum* leaves were greater than twice those in *Solanum nigrum*. They also established that small unripe fruits of *Solanum nigrum* had a high concentration of solasodine, but both the concentration and the absolute amount per fruit decreased with fruit maturation. It is difficult to compare our result with that of Eltayeb *et al.* (1997) because fruits collected for this study was irrespective of their developmental stages.

As many of the research works concentrate attention mainly on solasodine due to its high value in pharmaceutical industries, we have also tried to contribute little information regarding the presence of solasodine in the seven *Solanum spp.* of Meghalaya and it can be established from our study that all the seven species of *Solanum* contain solasodine, out of which *Solanum khasianum* contains highest amount of solasodine per gram of fruit powder.

CONCLUSION

CHAPTER 5

CONCLUSION

From the results and discussions of the present study it can be concluded that-

- The nuclear *18SrRNA* and chloroplast *rbcL* are phylogenetically informative but are much conserved suggesting that they are not rapidly evolving, having a lower percentage of parsimony informative characters.
- The nuclear *ITS* data set and chloroplast *matk* had higher percentage of variable sites which suggests that they are rapidly evolving.
- Results suggest that instead of analyzing the individual genes or loci separately it is always better to use multiple data matrix in combination which adds more resolution in phylogenetic interpretation.
- There is no significant difference between the phylogenetic relationship interpreted from nuclear data sets and chloroplast data sets (except for some minor differences of branch positions), which suggests that all the plants studied here originated from single individual ancestor.
- From our study of combined data set obtained from all the four gene sequences in combination, it can be established that the plants (both cultivated and wild) available in Meghalaya can be grouped into three monophyletic clades – Clade I consisting of members of the subgenus *Potatoe*, clade II includes members of the subgenus *Solanum*, and clade III

comprising of members of the subgenus *Leptostemonum*. A very close evolutionary relationship has been noticed between -

1. *Solanum nigrum* and *Solanum clavatum*, 2. *Solanum kurzii* and *Solanum gilo* and 3. *Solanum sisymbriifolium* and *Solanum khasianum* indicating their recent diversifications in the evolutionary pathway. The results also suggest that *Solanum torvum* is sister to the highly supported clade consisting of *Solanum kurzii* and *Solanum gilo*.

- The PCR-RFLP result also suggests the close evolutionary relationship between -

  1. *Solanum nigrum* and *Solanum clavatum*

  2. *Solanum kurzii, Solanum gilo* and *Solanum kurzii*

- Total protein profiling revealed phylogenetic tree which is also almost congruent to phylogenetic tree constructed by using molecular data.

- The qualitative screening of secondary metabolites revealed that *S. khasianum* is the richest among all the species studied here with regard to the secondary metabolites screened.

- The fruits of *S. khasianum* contains significantly higher amount of solasodine followed by *S. sisymbriifolium*. Among the species studied here, *S. kurzii* and *S. gilo* contains lowest amount of solasodine.

REFERENCES

REFERENCES

Aliero A.A., Asekun O.T., Grierson D.S., Afolayan A.J. (2007). Volatile Components from the Roots of *Solanum pseudocapsicum*, Journal of Medicinal Food, 10(3): 557-558.

Ames M. and Spooner D. M. (2010). Phylogeny of *Solanum* series *Piurana* and related species in *Solanum* section *Petota* based on five conserved ortholog sequences. Taxon. 59 (4): 1091–1101

Ando T., Kokubun H., Watanabe H., Tanaka N., Yukawa T., Hashimoto G., Marchesi E., Suárez E. and Basualdo I.L. (2005). Phylogenetic analysis of *Petunia* sensu Jussieu (Solanaceae) using chloroplast DNA RFLP. Annals of Botany. 96: 289-297

Anonymous (2004). The Wealth of India, NISCAIR, CSIR, New Delhi, pp. 103.

Aoki S. and Ito M. (2000). Molecular phylogeny of *Nicotiana* (Solanaceae) based on the nucleotide sequence of the *matK* gene. Plant Biology. 2: 316-324

Asl M.N. and Hossein H. (2008) Review of pharmacological effects of *Glycyrrhiza* sp. and its bioactive compounds. Phytotherapy Research. 22 (6): 709–24

Bagchi M., Milnes M., Williams C., Balmoori J., Ye X., Stohs S. and Bagchi D. (1999) Acute and chronic stress-induced oxidative gastrointestinal injury in rats and the protective ability of a novel grape seed proanthocyanidin extract. Nutrition Research.19: 1189–1199

Bakker F.T., Olsen J.L., and Stam W.T. (1995). Evolution of nuclear rDNA ITS sequences in the Cladophora albidalsericea clade. Journal of Molecular Evolution. 40(6):640-651

Baldwin B.G., Sanderson M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J. (1995). The ITS region of nuclear ribosomal DNA-A valuable source of evidence on Angiosperm phylogeny. Annals of the Missouri Botanical Garden. 82: 247–277.

Bharathan G. and Zimmer, E.A.(1995). Early branching events in monocotyledons: partial 18S ribosomal DNA sequence analysis, *In* Monocotyledons: systematics and evolution [P. J. Rundall, P. J. Cribb, D. F. Cutler and C. J. Humphries (eds)], Royal Botanic Gardens, Kew, pp. 81-107.

Bhattacharya J. (1984). Isolation of solasodine from the fruits of *Solanum asperum* and *Solanum paludosum*. Journal of Natural Products. 47:1059-1060.

Bhattacharya D. and Medlin L. (1998). Algal phylogeny and the origin of land plants. Plant Physiol. 116: 9–15

Bohs L. (1995). Transfer of *Cyphomandra* (Solanaceae) and its species to *Solanum*, Taxon, 44:583–587.

Bohs L. and Olmstead R.G. (1997). Phylogenetic relationships in *Solanum* (Solanaceae) based on *ndh*F sequences, Systematic Botany. 22: 5–17.

Bohs L. and Olmstead R.G. (1999). *Solanum* phylogeny inferred from chloroplast DNA sequence data, *In* Solanaceae IV: Advances in Biology and Utilization [M. Nee, D. E. Symon, R. N. Lester & J. P. Jessop (eds)], Royal Botanic Gardens, Kew, pp. 97–110.

Bohs L. and Olmstead R.G (2001). A reassessment of *Normania* and *Triguera* (Solanaceae), Plant Systematics and Evolution. 228: 33–48.

Bohs L. (2005). Major clades in *Solanum* based on *ndh*F sequences, *In* A festschrift for William G. D'Arcy: the legacy of a taxonomist [R. C. Keating, V. C. Hollowell, & T. B. Croat (eds.)], Monographs in Systematic Botany from the Missouri Botanical Garden, 104:27-49.

Boulter D. and Gilroy J.S. (1992). Partial sequences of 18S ribosomal RNA of two genera from each of six flowering plant families, Phytochemistry, 31: 1243-1246.

Bradford M.M. (1976). A rapid and sensitive for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72: 248-254..

Bradshaw J.E., Bryan G.J. and Ramsey G. (2006). Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilization in potato breeding. Potato Research. 49: 49–65

Campos M.A., Ribeiro S.G., Rigden D.J., Monte D.C. and Grossi de Sa M.F. (2002). Putative pathogenesis-related genes within *Solanum nigrum* L. var. *americanum* genome: isolation of two gnes soding for PR5-like proteins, phylogenetic and sequence analysis. Physiology and Molecular Plant Pathology. 61(4): 205–216

Chase M.W., Soltis D.E., Olmstead R.G., Morgan D., Les D. H., Mishler B.D.,Duvall M.R., Price R.A., Hills H.G., Qiu Y.L., Kron K.A., Rettig J.H., Conti E., Palmer J. D., Manhart J.R., Sytsma K.J., Michaels H.J.,

Kress W.J., Karol K.G., Clark W.D., Hedroen M.,Gaut B.S.,Jansen R.K.,Kim K.J.,Wimpee C.F.,Smith J.F., Furnier G. R., Strauss S. H., Xiang Q.Y., Plunkett G. M., Soltis P. S., Swensen S., Williams S. E., Gadek P. A., Quinn C. J., Eguiarte L. E., Golenberg E., Learn G. H. Jr., Graham S.W, Barrett S.C.,Dayanandan S., and Albert V.A. (1993). DNA sequence phylogenetics of seed plants: an analysis of the plastid gene *rbcL*. Annals of the Missouri Botanical Garden 80: 528-580.

Cheng L., Tsai T., Wang J., Lin C. and Kuo K. (1998). The rhamnose moiety of solamargine plays a crucial role in triggering cell death by apoptosis. Biochemical and Biophysical Research Communications. 242:21-25.

Child A. and R. N. Lester. (2001). Synopsis of the genus *Solanum* L. and its infrageneric taxa. *In* Solanaceae V: Advances in Taxonomy and Utilization [R. G. van den Berg, G. W. M. Barendse, G. M. van der Weerden & C. Mariani (eds)], Nijmegen Univ. Press, The Netherlands. Pp. 39–52

Chinedu, S. N., Olasumbo, A. C., Eboji, O. K., Emiloju, O. C., Arinola, O. K. & Dania, D. I. (2011). Proximate and phytochemical analysis of *Solanum aethiopicum* L and *Solanum Macrocarpon* L. fruits. Research Journal of Chemical Sciences. 1: 63-71

Christopher T., Martine, Gregory J. Anderson and Donald H. (2009). Gender-bending aubergines: molecular phylogeneticsof cryptically dioecious *Solanum* in Australia. Australian Systematic Botany, 22:107 – 120

Clarkson J. J., Knapp S., Aoki S., Garcia V.F., Olmstead R.G. and Chase M.W. (2004). Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. Molecular Phylogeny and Evolution. 33: 75-90

Coleman A.W., Suarez A., and Goff L.J. (1994). Molecular delineation of species and syngens in Volvocacean green algae (Chlorophyta). Journal of Molecular Evolution, 30:80-90 .

Conner A.J. (1987). Differential solasodine Accumulation in photoautotrophic and heterotrophic tissue cultures of *Solanum lanciniatum* . Phytochemistry 26:2749-2750.

Conti E., Fischbach A., and Sytsma K.J. (1993). Tribal relationships in Onagraceae: implications from *rbc*L sequence data. Annals of the Missouri Botanical Garden. 80: 672–685.

Crabbe P.G. and Fryer C. (1982). Evaluation of chemical analysis for the determination of in *Solanum laciniatum*. Journal of Pharmaceutical Sciences. 71:1356-1362.

Cushnie T.P.T. and Lamb A.J. (2005) Antimicrobial activity of flavonoids. International Journal of Antimicrobial Agents. 26 (5): 343–356

D'Arcy W.G. (1972). Solanaceae studies II: Typification of subdivisions of *Solanum*, Annals of Missouri Botanic Garden. 59: 262–278.

D'Arcy W.G. (1979). Classification of the Solanaceae. *In* The Biology and Taxonomy of Solanaceae [J.Ge. Hawkes, R. N. Lester, and A. D.Skelding (eds.)], Academic Press, London; pp. 3-47.

D'Arcy W.G. (1991). The Solanaceae since 1976, with a review of its biogeography. *In* Solanaceae III: Taxonomy, Chemistry, Evolution [J. G. Hawkes, R. N. Lester, M. Nee & N. Estrada (eds)], Royal Botanic Gardens, Kew, pp 75–137.

Daunay M.C. and Lester R.N. (1988). The usefulness of taxonomy for Solanaceae breeders, with special reference to the genus *Solanum* and to *Solanum melongena* L. (eggplant), Capsicum Newslatter, 7:70-79.

de Sousa R.R., Queiroz K.C., Souza A.C., Gurgueira S.A., Augusto A.C., Miranda., Peppelenbosch M.P., Ferreira C.V., and Aoyama H. (2007). Phosphoprotein levels, MAPK activities and NFkappaB expression are affected by fisetin. Journal of Enzyme Inhibition and Medicinal Chemistry. 22 (4): 439-444.

Dobzhansky T. (1973). Nothing in biology makes sense except in the light of evolution. The American Biology Teacher. 35:125–129.

Doyle J.J., Doyle J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin 19: 11–15.

Drummond A.J., Ashton B., Buxton S., Cheung M., Cooper A., Heled J., Kearse M., Moir R., Stones-Havas S., Sturrock S., Thierer T., Wilson A. (2010). Geneious v5.3, Available from http://www.geneious.com

Dunal, M.F. (1852). Solanaceae. *In* Prodromus Systematis Naturalis Regni Vegetabilis [A.P. de Candolle (ed.)], Victoris Masson, Paris, Vol.13(1). pp. 1–690.

Eltayeb E.A., Al-Ansary A.S. and Roddick, J.G. (1997). Changes in the steroidal alkaloid solasodine during development of *Solanum nigrum* and *Solanum incanum*. Phytochemistry. 46: 489-494.

Embley M.T., Hirt R.P. and Williams D.M. (1994). Biodiversity at the molecular level: the domains, kingdoms, and phyla of life, Philosophical transactions of the Royal Society, B345 (1311): 21-31.

Fay M.F., Olmstead R.G., Richardson J.E., Santiago E., Prance G.T. and Chase M.W. (1998). Molecular data support the inclusion of Duckeodendron cestroides in Solanaceae, Kew Bulletin, 53: 203-212.

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39: 783-791.

Flier W.G., van der Bosch G.M.B. and Turkensteen L.J. (2003). Epidemiological importance of *Solanum sisymbriifolium*, *S nigrum* and *S. dulcamara* as alternative hosts for *Phytophthora infestans*. Plant Pathology. 52:595–603.

Garcia V.F. and Olmstead R.G. (2003). Phylogenetics of Tribe Anthocercideae (Solanaceae) based on *ndhF* and *trnL/F* sequence data. Systematic Botany. 28: 609-615.

Gbile Z. O. and Adesina S. K. (1988) Nigerian *Solanum* Species of economic importance. Annals of Missouri Botanic Garden. 75: 862-865.

Gelder W.M.J. (2006). A new hydrolysis technique for steroid glycoalkaloids with unstable aglycones from *Solanum spp.*, Journal of Science of Food and Agriculture, 35(5): 487-494.

Guerrero M. 1976. Process for recovering and isolating solasodine and glycoside thereof. U.S. Patent 3,960,839.

Hamby R.K. and Zimmer E.A. (1988). Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae), Plant Systematics and Evolution, 160: 29-37.

Hamby R.K., and Zimmer E.A. (1992). Ribosomal RNA as a to be presented, genetic tool in plant systematic, *In* Molecular Systematics of Plants (P. S. Soltis, D. E. Soltis, and J. J. Doyle, Eds.), pp. 50– 91.

Harborne J.B. (1973). Phytochemical methods (London: Chapman and Hall Ltd). Pp-49-188.

Haridasan K. and Rao R.R. (1987). Solanaceae. *In* Forest Flora of Meghalaya [Bishen Singh, Mahendra Pal Singh(eds)], Dehradun, India, pp 646-647.

Hasebe M., Omori T., Nakazawa M., Sano T., Kato M., and Iwatsuki K. (1994). *rbcL* gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. Proceedings of the National Academy of Science USA 91: 5730-5734.

Hillis D.M. and Dixon M.T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference, Quarterly Review of Biology, 66: 411-453.

Hosaka K., Ogihara Y., Matsubayashi M. and Tsunewaki K. (1984). Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. Japanese journal of Genetics. 59: 349-369.

Hunziker A.T. (1979). South American Solanaceae: a synoptic survey. *In* The biology and taxonomy of the Solanaceae [J. G. Hawkes, R. N. Lester, and A. D. Skelding (eds.)], Academic Press, London, pp 120-134.

Huss V.A.R. and Kranz H.D. (1997). Charophyte evolution and the origin of land plants. Plant Systematics and Evolution (Suppl.). 11: 103–114.

Jaccard P. (1908). Nouvelles researches sur la distribution florate. Bulletin de la Société vaudoise des sciences naturelles. 44: 223-270.

Jaeger P.M.L. (1985). Systematic studies in the genus *Solanum* in Africa, Ph.D. dissertation, University of Birmingham, Birmingham, UK.

Jansky S.H. (2009). Breeding, genetics, and cultivar development. *In*, Advances in potato chemistry and technology [Singh J. and Kaur L. (eds.)]. Elsevier, San Diego, California, pp. 27–61.

Jeong J.B. , Jeong H.J., Park J.H. , Lee S.H., Lee J.R. , Lee H.K., Chung G.Y., Choi J.D.,Ben D.L. (2007). Cancer-Preventive Peptide Lunasin from *Solanum nigrum* L. Inhibits Acetylation of Core Histones H3 and H4 and Phosphorylation of Retinoblastoma Protein (Rb), Journal of agricultural and food chemistry. 55: 10707-10713.

Johnson L.A. and Soltis D.E. (1995). Phylogenetic inference using matK sequences. Annals of the Missouri Botanical Garden. 82: 149-175.

Johanson L. A., Schultz J. L., Soltis D. E., and Soltis P. S. (1996). Monophyly and generic relationships of Polemoniaceae based on *mat*K sequences. American Journal of Botany. 83: 1207–1224.

Källersjö M., Farris J.S., Chase M.W., Bremer B., Fay M.F., Humphries C.J., Petersen G., Seberg O.L. and Bremer K. (1998). Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. Plant Systematics and Evolution. 213: 259–287.

Kamoun S., Huitema E. and Vleeshouwers V.G.A.A. (1999). Resitance to Oomycetes: a general role of the hypersensitive response? Trends in Plant Science. 4:196–200.

Kharkongor P. and Joseph J. (1981). Folklore medicobotany of rural Khasi and Jaintia tribes in Meghalaya. *In* Glimpses of Indian Ethnobotany, [S. K. Jain (ed)], Oxford and IBH, New Delhi, pp. 135.

Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Kranz H.D. and Huss V.A.R. (1996). Molecular evolution of ferns and allies, and their relationship to seed plants: evidence from complete 18S rRNA gene sequences. Plant Systematics and Evolution. 202: 1–11.

Kranz H.D., Miks D., Siegler M.L., Capesius I., Sensen C.H.W., Huss V.A. (1995). The origin of land plants: phylogenetic relationships between Charophytes, Bryophytes, and vascular plants inferred from complete small subunit ribosomal RNA gene sequences. Journal of Molecular Evolution. 41: 74–84.

Kritikar K.R., and Basu B.D. (1935) Indian medicinal plants. M/s Bishen Singh Mahendrapal, New Delhi, India. 267-268.

Kron K. A. (1996). Phylogenetic relationships of Empetraceae, Epacridaceae, and Ericaceae: evidence from nuclear ribosomal 18S sequence data. Annals of Botany. 77:293–303.

Kulcheski F.R., Muschner V.C., Lorenz-Lemke A.P., Stehmann J.R., Bonatto S.L., Salzano F.M. and Freitas L.B. (2006). Molecular phylogenetic analysis of *Petunia* Juss. (Solanaceae). Genetica. 126: 3-14.

Lancaster J. and Mann J. D. (1975). Changes in solasodine content during the development of Solanum laciniatum Ait. New Zealand Journal of Agricultural Research, 18(2): 139-144.

Lebecka R. (2008). Host–pathogen interaction between *Phytophthora infestans* and *Solanum nigrum*, *S. villosum*, and *S. scabrum*. European Journal of Plant Pathology. 120: 233–240.

Levin R.A. and Miller J.S. (2005). Relationships within tribe Lycieae (Solanaceae): Paraphyly of *Lycium* and multiple origins of gender dimorphism. American Journal of Botany. 92: 2044-2053.

Levin R.A., Myers N.R. and Bohs L. 2006. Phylogenetic relationships among the "spiny solanums" (*Solanum* subgenus *Leptostemonum*, Solanaceae). American Journal of Botany. 93: 157-169.

Levin R.A., Watson K., Bohs, L. (2005).A four-gene study of evolutionary relationships in *Solanum* section *Acanthophora*. American Journal of Botany. 92(4):603–612.

Lin C., Lu C., Cheng M. and Gan K. (1990). The cytotoxic principles of *Solanum incanum*. Journal of Natural Products. 53:513-516.

Linnaeus, C., 1753. Species Plantarum, first ed. The Ray Society, London.

Livingstone K.D., Lackney V.K., Blauth J.R., Wijk R. Van and Jahn M.K. (1999). Genome mapping in capsicum and the evolution of genome structure in the solanaceae. Genetics.152(3): 1183–1202.

Mace E.S, Lester R.N. and Gebhardt C.C. (1999b). AFLP analysis of genetic relationships among the cultivated eggplant, *Solanum melongena* L., and wild relatives (Solanaceae). Theoretical and Applied Genetics. 99: 626-633.

Mace E.S., Gebhardt C.G. and Lester, R.N. (1999a). AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). Theoretical and Applied Genetics. 99: 634-641.

Manhart J.R. (1994). Phylogenetic analysis of green plant *rbcL* sequences. Molecular Phylogeny and Evolution. 3: 114–127.

Mann JD (1978) Production of solasodine for the pharmaceutical industry. Advances in Agronomy. 30:207–245.

Martine C.T., Vanderpool D., Anderson G.J. and Les D.H. (2006). Phylogenetic relationships of andromonoecious and dioecious Australian species of *Solanum* subgenus *Leptostemonum* section *Melongena*: Inferences from ITS sequence data. Systematic Botany. 31: 410-420.

Martins T.R. and Barkman T.J. (2005). Reconstruction of Solanaceae phylogeny using the nuclear gene SAMT. Systematic Botany. 30: 435-447.

McCourt R., Karol M. K.G,. Kaplan S., and Hoshaw R.W. (1995). Using *rbcL* sequences to test hypotheses of chloroplast and thallus evolution in

conjugating green algae (Zygnematales, Charophyceae). Journal of Phycology. 31: 989-995.

Miller J.S. (2002). Phylogenetic relationships and the evolution of gender dimorphism in *Lycium* (Solanaceaae). Systematic Botany. 27: 416-328.

Miz R., Mentz L. and Chies T. (2007). Overview of the phylogenetic relationships of some southern Brazilian species from section Torva and related sections of "spiny *Solanum*" (*Solanum* subgenus *Leptostemonum*, Solanaceae), Genetica, 3: 132-140.

Morgan D.R., Soltis D.E., and Robertson K.R. (1994). Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. American Journal of Botany 81: 890–903.

Müller K. (2005). SeqState: primer design and sequence statistics for phylogenetic DNA datasets. Applied Bioinformatics 4: 65–69.

Nee M. (1999). Synopsis of *Solanum* in the New World. *In* Solanaceae IV: advances in biology and utilization [M. Nee, D. E. Symon, R. N. Lester, and J. P. Jessop (eds.)], Royal Botanic Gardens, Kew, Richmond, Surrey, UK, pp 285–333.

Nickrent D.L. and Franchina C.R. (1990). Phylogenetic relationships of the Santalales and relatives, Journal of Molecular Evolution. 31: 294-301.

Noda Y., Kneyuki T., Igarashi K. and Packer M.L. (2000). Antioxidant activity of nasunin, an anthocyanin in eggplant peels. Toxicology. 148:119-123.

Olmstead R.G. and Bohs, L. (2007). A Summary of molecular systematic research in *Solanaceae*: 1982-2006, Acta Horticulturae. (ISHS). 745:255-268.

Olmstead R.G. and Palmer J.D. (1992). A chloroplast DNA phylogeny of the Solanaceae: Subfamilial relationships and character evolution, Annals of Missouri Botanic Garden. 79: 346–360.

Olmstead R.G. and Palmer J.D. (1997). Implications for the phylogeny, classification, and biogeography of *Solanum* from cpDNA restriction site variation, Systematic Botany. 22: 19–29.

Olmstead R.G. and Sweere J.A. (1994). Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae, Systematic Biology. 43: 467-481.

Olmstead R.G., Sweere J.A., Spangler R.E., Bohs L. and Palmer J.D. (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *In* Solanaceae IV: Advances in Biology and Utilization [M. Nee, D. E. Symon, R. N. Lester & J. P. Jessop, (eds)], Royal Botanic Gardens, Kew, pp 111–137.

Ortiz R. (2001). The state of the use of potato genetic diversity. *In* Broadening the genetic base of crop production [Cooper, H.D., Spillane, C., Hodgkin, T. (eds.)]Wallingford, U.K. CABI Publishing, pp. 181– 200.

Palmer J.D. and Zamir D. (1982). Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proceedings of National Academy of Science. USA. 79: 5006-5010.

Perez F., Menendez A., Dehal P. and Quiros C.F. (1999). Genomic structural differentiation in *Solanum*: comparative mapping of the A and E genomes. Theoritac and Applied Genetics98: 1183–1193.

Perez-Amador M. C., Ocotero V.M., Castaneda J.M.G. and Esquinca A. R. G. (2007). Alkaloids in *Solanum torvum* Sw (Solanaceae). International journal of Experimental Botany. 76: 39-45.

Plunkett G.M., Soltis D.E. and Soltis P.S. (1997). Clarification of the relationship between Apiaceae and Araliaceae based on matK and rbcL sequence data. American Journal of Botany 84: 565-580.

Posada D. (2008). jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256.

Qiu Y.Q., Cho Y., Cox J.C. and Palmer J.D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature. 394: 671–674.

Rambaut A. (2009). FigTree v1.3.1 2006-2009 [Readme file]. Accessed on October 12, 2011 Available with the program package at http://tree.bio.ed.ac.uk/software/figtree/.

Rodriguez J., Segovia R., Guerreiro E., Derretti F., Zamarbide S.G. and Ertola R. (1979). Production of solasodine by microbial hydrolysis of glycoalkaloids of *Solanum eleagnifolium* Cav. and its application to the synthesis of 16-dehydropregnenolone acetate. Jornal of Chemical Technology and Biotechnology. 29:525-530.

Roecklein J.C. and Leung P.S. (1987). Medicinal- *S.khasianum.* *In* A profile of Economic plants, Transaction publishers [P.S. Leung, J.C. Roecklein, A. L. Jekon (eds)] pp. 295.

Ronquist F, Huelsenbeck J.P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed mo.dels. Bioinformatics 19:1572–1574.

Roorigiuez A. and Spooner D.M. (2002). Subspecies boundaries of the wild potatoes *Solanum bulbocastanum* and *Solanum cardiophyllum* based on morphological and nuclear RFLP data, Acta Botanica Mexicana, 61:9-25.

Saitou N., Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Sarich A.C. and Wilson V.M. (1969). A molecular time scale for human evolution. Proceedings of National Academy of Science, USA. 63(4):1088-1093.

Savolainen V., Chase M.W., Hoot S.B., Morton C.M., Soltis D.E., Bayer C., Fay M.F., de Bruijn A.Y., Sullivan S. and Qiu Y.L. (2000) Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. Systematic Biology. 49(2):306-62.

Sczkowski C.P., Kalinowska M. and Wojciechowski Z. (1988). The 3-O glucosylation of steroidal saponins and alkaloids in eggplant (Solanum melongena); evidence for two separate glycosyltransferences. Phytochemistry. 48: 1151-1159.

Seithe, A. 1962. Die Haararten der Gattung *Solanum* L. und ihre taxonomische Verwertung. Botanische Jahrbücher fur Systematik. 81(3), 261–336.

Smith S.D. and Baum D.A. (2006). Phylogenetics of the florally-diverse Andean clade Iochrominae (Solanaceae). American Journal of Botany. 93:1140-1153.

Sofowara A. (1993). Medicinal plants and traditional medicine in Africa (Ibadan, Nigeria: Spectrum Books Ltd)

Soltis D. E., Morgan D. R., Grable A., Soltis, P. S., and Kuzoff, R. (1993). Molecular systematics of Saxifragaceae sensu stricto. American Journal of Botany. 80: 1056–1081.

Soltis P.S. and Soltis D.E. (1995). Molecular evolution of 18S rDNA in angiosperms: implication for character weighting in phylogenetic analysis. *In* Molecular systematics of plnts II: DNA sequencing [D. E. Soltis, P. S. Soltis and J. J. Doyle (eds)], Kluwer Academic Publishers, Boston, pp 188-210.

Soltis D.E., Hibsch-Jetter C., Soltis P.S., Chase M.W. and Farris J.S. (1997). Molecular phylogenetic relationships among angiosperms: an overview based on *rbcL* and 18S rDNA sequences. *In* Evolution and Diversification of Land Plants [K. Iwatsuki, P.H. Raven (eds.)], Springer-Verlag, New York, pp. 157–178.

Soltis D.E., Kuzoff R.K., Conti E., Gornall R., and Ferguson K. (1996). *matK* and *rbcL* gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. American Journal of Botany. 83: 371–382.

Soltis D.E., Soltis P.S., Mort M.E., Chase M.W., Savolainen V., Hoot S.B. and Morton C.M. (1998). Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Systematic Biology. 47: 32–42.

Spooner D.M. and Sytsma K.J. (1992). Reexamination of series relationships of Mexican and Central American wild potatoes (*Solanum* sect. *Petota*): evidence from chloroplast DNA restriction site variation. Systematic Botany. 17: 432–448.

Spooner D.S., Anderson G.J. and Jansen R.K. (1993). Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae), American Journal of Botany. 80: 676–688.

Spooner D.M., McLean K., Ramsay G., Waugh R. and Bryan G.J. (2005a). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proceedings of National Academy of Science. USA. 102: 14694-14699.

Spooner D.M., Peralta I. and Knapp S. (2005b). Comparison of AFLPs with other markers from phylogenetic inference in wild tomatoes [*Solanum* L. Section *Lycopersicon* (Mill.) Wettst.]. Taxon. 54: 43-61.

Sree A., Rao Y.R. and Mahapatra S.N. (1982). A process for the conversion of solasodine 110 to 16-dehydropregnenolone acetate. Research and Industry 27:326-328.

Stern, S. and Bohs L. (2012). An explosive innovation: Phylogenetic relationships of *Solanum* section *Gonatotrichum* (Solanaceae), PhytoKeys. 8: 83–98.

Sudheesh S., Presannakumar G., Vijaykumar S. and Vijayalashmi N.R. (1997). Hypolipidemic effect of flavonoids from *Solanum melongena.* Plant floods for Human Nutrition. 51:321-30.

Symon D.E. (1981). A revision of the genus *Solanum* in Australia, Journal Adelaide Botanic Garden. 4:1–367.

Symon D.E. (1994). Kangaroo apples: *Solanum* sect. *Archaesolanum.* Published by the author, Keswick, South Australia, pp 234-276.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28(10): 2731-2739.

Tanksley S.D., Bernatzky R., Lapitan N.L. and Prince J.P. (1988). Conservation of gene repertoire but not gene order in pepper and tomato. Proceedings of National Academy of Science. USA. 84: 6419–6423.

Terri L., Weese and Lynn B. (2007) A Three-Gene Phylogeny of the Genus *Solanum* (Solanaceae). Systematic Botany. 32(2): 445–463.

Thompson J.D., Gibson T.J., Plewniak F., Mongin F.J., Higgins D.G. (1997). Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882.

Trease G.E. and Evan W.C. (1989). Pharmacognsy (Brailliar Tiridel Canada: Macmillian publishers) 11th edn., Pp-290.

Valentin Santiago E. and Olmstead R.G. (2003). Phylogenetics of the Antillean Goetzeoideae (Solanaceae) and their relationships within the Solanaceae based on chloroplast and ITS DNA sequence data. Systematic Botany. 28: 452-460.

Vinson J.A., Hao Y., Su X. and Zubik L. (1998) Phenol antioxidant quantity and quality in foods: Vegetable. Journal of Agricultur and Food Chemistry. 48: 3630-3634.

Vohora S.B., Kumar I. and Khan M.S.Y. (1984) Effect of alkaloids of *Solanum melongena* on the central nervous system. *Journal of Ethnopharmacol.* 11: 331-336.

Weese T.L. and Bohs L. (2007). A three gene phylogeny of the genus *Solanum* (Solanaceae). Systematic Botany. 32:445-463.

Walsh B.M. and Hoot S.B. (2001). Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast *atp*B-*rbc*L spacer region and nuclear *waxy* introns. International Jounal of Plant Science. 162: 1409-1418.

Whalen M.D. (1984). Conspectus of species groups in *Solanum* subgenus *Leptostemonum*. Gentes Herbarum, 12: 179–282.

Whitson M. and Manos P.S. (2005). Untangling *Physalis* (Solanaceae) from the physaloids: A two-gene phylogeny of the Physalinae. Systematic Botany. 30: 216-230.

Wolfe K.H. (1991). Protein-coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries, and rates of molecular evolution. *In* The photosynthetic apparatus: molecular biology and operation, Cell culture and somatic cell genetics of plants [L. Bogorad and I. K. Vasil (eds.)], Academic Press, San Diego, CA. 7B: 467–482.

Xiang Q.Y., Crawford D. J.,Wolfe A. D., Tang Y.C., and De Pamphilis C. (1998). Origin and biogeography of *Aesculus* L. (Hippocastanaceae): A molecular phylogenetic perspective. [submitted]

Xiang Q.Y., Soltis D. E., Morgan D. R., and Soltis, P. S. (1993.) Phylogenetic relationships of *Cornus* L. sensu lato and putative relatives inferred from *rbc*L sequence data. Annals of the Missouri Botanical Garden. 80: 723–734.

Xu R., Zhao W., Xu J., Shao B., and Qin G. (1996). Studies on bioactive saponins from Chinese medicinal plants. Advances in Experimental Medicine and Biology. 404: 371-82.

Yamamoto Y. and Gaynor R.B. (2001) Therapeutic potential of inhibition of the NF-kB pathway in the treatment of inflammation and cancer. The Journal of Clinical Investigaion. 107(2):135–142.

Yanthan M., Biate D. and Misra A. K. (2011). Taxonomic resolution of actinorhizal *Myrica* species from Meghalaya (India) through nuclear rDNA sequence analyses. Functional Plant Biology. 38(9):738-746.

Yuan Y., Zhang Z.Y., and Olmstead R.G. (2006). A retroposon insertion to the waxy gene: Defining monophyly of the tribe Hyoscyameae (Solanaceae), and

revealing the allopolyploid origin of *Atropa belladonna*. Molecular Biology and Evolution. 23:2263-2267.

Zimmer E.A., Hamby R.K., Arnold M.L., Leblance D.A. and Theriot E.C. (1989). Ribosomal RNA phylogenies and flowering plant evolution. *In* The hierarchy of life [B. Fernhilm, K. Bremer and J. Jornvall (eds.)], Elsevier Science Publishers, Amsterdam, pp- 205-214.

APPENDICES

APPENDIX 1

```
CLUSTAL 2.1 multiple sequence alignment of 18S rDNA

S.khasianum             CATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCGATGATTCATGATAAC 60
S.sisymbriifolium       CAAGGATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCAGCGATGATTCATGATAAC 60
S.torvum                -----ATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCGATGATTCATGATAAC 55
S.kurzii                CATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCAATGATTCATGATAAC 60
S.gilo                  CATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCGATGATTCATGATAAC 60
S.nigrum                CATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCAGCGATGATTCATGATAAC 60
S.clavatum              CATTTATTAGATAAAAGGTCGACGCGGGCTCTGCCCGTTGCTGCGATGATTCATGATAAC 60
                             ********************************** ** ***************

S.khasianum             TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
S.sisymbriifolium       TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
S.torvum                TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 115
S.kurzii                TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
S.gilo                  TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
S.nigrum                TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
S.clavatum              TCGACGGATCGCACGGCCATCGTGCCGGCGACGCATCATTCAAATTTCTGCCCTATCAAC 120
                        ************************************************************

S.khasianum             TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
S.sisymbriifolium       TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
S.torvum                TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 175
S.kurzii                TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
S.gilo                  TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
S.nigrum                TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
S.clavatum              TTTCGATGGTAGGATAGTGGCCTACCATGGTGGTGACGGGTGACGGAGAATTAGGGTTCG 180
                        ************************************************************

S.khasianum             ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
S.sisymbriifolium       ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
S.torvum                ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 235
S.kurzii                ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
S.gilo                  ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
S.nigrum                ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
S.clavatum              ATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGGCGCGCAA 240
                        ************************************************************

S.khasianum             ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
S.sisymbriifolium       ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
S.torvum                ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 295
S.kurzii                ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
S.gilo                  ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
S.nigrum                ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
S.clavatum              ATTACCCAATCCTGACACGGGGAGGTAGTGACAATAAATAACAATACCGGGCTCTATGAG 300
                        ************************************************************

S.khasianum             TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGA 360
S.sisymbriifolium       TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 360
S.torvum                TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 355
S.kurzii                TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 360
S.gilo                  TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 360
S.nigrum                TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 360
S.clavatum              TCTGGTAATTGGAATGAGTACAATCTAAATCCCTTAACGAGGATCCATTGGAGGGCAAGT 360
                        ************************************************************

S.khasianum             CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
S.sisymbriifolium       CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
S.torvum                CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 415
S.kurzii                CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
S.gilo                  CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
S.nigrum                CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
S.clavatum              CTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTTAAGTTGTTGCAGTT 420
                        ************************************************************
```

```
S.khasianum           AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
S.sisymbriifolium     AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
S.torvum              AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 475
S.kurzii              AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
S.gilo                AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
S.nigrum              AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
S.clavatum            AAAAAGCTCGTAGTTGGACTTTGGGATGGGCCGGCCGGTCCGCCCTAGGTGTGCACCGGT 480
                      ************************************************************

S.khasianum           CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
S.sisymbriifolium     CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
S.torvum              CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 535
S.kurzii              CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
S.gilo                CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
S.nigrum              CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
S.clavatum            CGTCTCGTCCCTTCTGTCGGCGATGCGCTCCTGGCCTTAATTGGCCGGGTCGTGCCTCCG 540
                      ************************************************************

S.khasianum           GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
S.sisymbriifolium     GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
S.torvum              GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 595
S.kurzii              GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
S.gilo                GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
S.nigrum              GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
S.clavatum            GCGCTGTTACTTTGAAGAAATTAGAGTGCTCAAAGCAAGCCTACGCTCTGTATACATTAG 600
                      ************************************************************

S.khasianum           CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
S.sisymbriifolium     CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
S.torvum              CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 654
S.kurzii              CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
S.gilo                CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
S.nigrum              CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
S.clavatum            CATGGGATAACATTATAGGATTTCGGTCCTATT-ACGTTGGCCTTCGGGATCGGAGTAAT 659
                      ********************************* ************************

S.khasianum           GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
S.sisymbriifolium     GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
S.torvum              GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 714
S.kurzii              GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
S.gilo                GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
S.nigrum              GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
S.clavatum            GATTAACAGGGACAGTCGGGGGCATTCGTATTTCATAGTCAGAGGTGAAATTCTTGGATT 719
                      ************************************************************

S.khasianum           TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
S.sisymbriifolium     TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
S.torvum              TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 774
S.kurzii              TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
S.gilo                TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
S.nigrum              TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
S.clavatum            TATGAAAGACGAACAACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGA 779
                      ************************************************************

S.khasianum           AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
S.sisymbriifolium     AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
S.torvum              AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 834
S.kurzii              AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
S.gilo                AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
S.nigrum              AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
S.clavatum            AAGTTGGGGGCTCGAAGACGATCAGATACCGTCCTAGTCTCAACCATAAACGATGCCGAC 839
                      ************************************************************

S.khasianum           CAGGGATCGGCGGATGTTGCTTTTAGGACTCCGCCGGCACCTTATGAAAAATCAAATTTT 899
S.sisymbriifolium     CAGGGATCGGCGGATGTTGCTTTTAGGACTCCGCCGGCACCTTATGAAAAATCAAATTTT 899
S.torvum              CAGGGATCGGCGGATGTTGCTTTTAGGACTCCGCCGGCACCTTATGAGAAATCAAATTTT 894
S.kurzii              CAGGGATCGGCGGATGTTGCTTTTAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTTT 899
S.gilo                CAGGGATCGGCGGATGTTGCTTTTAGGACTCCGCCGGCACCTTATGAGAAATCAAAGTTT 899
S.nigrum              CAGGGATCGGCGGATGTTCTTTTTAGGACTCCCCCGGCACCTTATGAAAAATCAAATTTT 899
S.clavatum            CAGGGATCGGCGGATGTTCTTTTTAGGACTCCGCCGGCACCTTATGAGAAATCAAATTTT 899
                      ****************** ************ ************** ******** *** 
```

```
S.khasianum          TTGGGTCCCGGGGGGAGTATGGTCGCAGGGCTGAAACTTAAAGGGATTGACGGAAGGGCA 959
S.sisymbriifolium    TTGGGTCCCGGGGGGAGTATGGTCGCAGGGCTGAAACTTAAAGGGATTGACGGAAGGGCA 959
S.torvum             TTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCA 954
S.kurzii             TTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCA 959
S.gilo               TTGGGTTCCGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCA 959
S.nigrum             TTGGGTTCCAGGGGGAGTATGGTCGCAAAGCTGACACTTAGAAAAGTTGACGGAAGGGCA 959
S.clavatum           TTGGGTTCCGGGGGGAGTATGGCGGCAAAGCTGGAATTTAAAGGAATTGACGGAAGGGCA 959
                     ****** ** ***********  ***  ****  * *** *    **************

S.khasianum          CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
S.sisymbriifolium    CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
S.torvum             CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1014
S.kurzii             CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
S.gilo               CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
S.nigrum             CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
S.clavatum           CCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCA 1019
                     ************************************************************

S.khasianum          GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
S.sisymbriifolium    GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
S.torvum             GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1074
S.kurzii             GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
S.gilo               GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
S.nigrum             GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
S.clavatum           GACATAGTAAGGATTGACAGACTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATG 1079
                     ************************************************************

S.khasianum          GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
S.sisymbriifolium    GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
S.torvum             GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1134
S.kurzii             GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
S.gilo               GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
S.nigrum             GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
S.clavatum           GCCGTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAG 1139
                     ************************************************************

S.khasianum          CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
S.sisymbriifolium    CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
S.torvum             CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1194
S.kurzii             CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
S.gilo               CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
S.nigrum             CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
S.clavatum           CCTGCTAACTAGCTATGCGAAGGTATCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGC 1199
                     ************************************************************

S.khasianum          CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
S.sisymbriifolium    CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
S.torvum             CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1254
S.kurzii             CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
S.gilo               CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
S.nigrum             CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
S.clavatum           CTTTTAGGCCGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGG 1259
                     ************************************************************

S.khasianum          CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
S.sisymbriifolium    CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
S.torvum             CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1314
S.kurzii             CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
S.gilo               CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
S.nigrum             CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
S.clavatum           CCGCACGCGCGCTACACTGATGTATTCAACGAGCTTATAGCCTTGGCCGACAGGCCCGGG 1319
                     ************************************************************

S.khasianum          TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
S.sisymbriifolium    TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
S.torvum             TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1374
S.kurzii             TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
S.gilo               TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
S.nigrum             TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
S.clavatum           TAATCTTTGAAATTTCATCGTGATGGGGATAGATCATTGCAATTGTTGGTCTTCAACGAG 1379
                     ************************************************************
```

```
S.khasianum           GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
S.sisymbriifolium     GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
S.torvum              GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1434
S.kurzii              GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
S.gilo                GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
S.nigrum              GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
S.clavatum            GAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTTGACTACGTCCCTGCCCTTTGTACAC 1439
                      ************************************************************

S.khasianum           ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
S.sisymbriifolium     ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
S.torvum              ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1494
S.kurzii              ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
S.gilo                ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
S.nigrum              ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
S.clavatum            ACCGCCCGTCGCTCCTACCGATTGAATGATCCGGTGAAATGTTCGGATCGCGGCGACGTG 1499
                      ************************************************************

S.khasianum           GGCGGTTCGCTGCCCGCGACGTCGCGGGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
S.sisymbriifolium     GGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
S.torvum              GGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1554
S.kurzii              GGCGGTTCGCTGCCCGCGACGTCGCGGGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
S.gilo                GGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
S.nigrum              GGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
S.clavatum            GGCGGTTCGCTGCCCGCGACGTCGCGAGAAGTCCATTGAACCTTATCATTTAGAGGAAGG 1559
                      ************************* ********************************

S.khasianum           AGAAGTCGTAAC 1571
S.sisymbriifolium     AGAAGTCGTAAC 1571
S.torvum              AGAAGTCGTAAC 1566
S.kurzii              AGAAGTCGTAAC 1571
S.gilo                AGAAGTCGTAAC 1571
S.nigrum              AGAAGTCGTAAC 1571
S.clavatum            AGAAGTCGTAAC 1571
                      ************
```

APPENDIX 2

CLUSTAL 2.1 multiple sequence alignment of *rrn*ITS

```
S.sisymbriifolium      --------GTC-ACTGCAGCGCACGACCCGCGAACGCGTTCAAACACGGGGGGCGC-GCG 50
S.khasianum            --------CTCGACTGCAGCG-ACGACCCGCGAACACGTTCAAACACTGGGGGGGCCGCG 51
S.clavatum             ----------TGACTGCAGCAGACGACCCGCGAACACGTTTAAACACCGGGGGAGCAGCG 50
S.nigrum               ----------GAACTGCAAGCGACGACCCGCGAACACGTTCAAACACCGGGGGAGCAGCG 50
S.gilo                 ----------CTGCCACAGCAGGTGACCTGCGAACACATTCAAACAATGGGGGACCCACG 50
S.torvum               ----------CTGCCACAGCAGGTGACCCGCGAAAGCATTCAAACACCAGGGGAGCTATG 50
S.kurzii               TGTAATAAACCTGCCACAGCAGGTGACCCGCGAAAGCATTCAAACACCAGGGGAGCTATG 60
                                 *  **      **** *****  * ** *****   ****  *   *

S.sisymbriifolium      --------------------CGGCG-----GGGGTGCTCCGGGCCCTCGCCTCGCCC-CT 84
S.khasianum            --------------------CGGCG-----GGGGTGCTTCGGCGCTGCCCCGCGCGT-CT 85
S.clavatum             --------------------CGGCG-----CGGGTGCTTCGGCGTCCCTCCGCGCGCGTT 85
S.nigrum               --------------------CGGCG-----CGGGTGCTTCGGCGTCCCTCCGCGCGCGTT 85
S.gilo                 --------------------CAGTGT----CGGGCATTTCGGTGCGGCCCCGCGCGT-AT 85
S.torvum               GCCTAAAAACACACGGGGAGCCGCACAGCGGGGGCCCTCAAGCTTCACCCCTCGTGC--T 108
S.kurzii               --------------------CGGCATA----GGGCACTTTAGC---ACCCCACGCGT-AT 92
                                          * *        ***   *   *     * ** **     *

S.sisymbriifolium      CCCTCCCGTCCATGAACGCGCGCTTGCGCGCTCGTTTTTTGGGCGGCTAA--CGAACCCC 142
S.khasianum            CCCCTCCGCCCCCGA-TGCGTGCCTCTGTGCTCGTTCTTGGGG--GCCAAAACGAACCCC 142
S.clavatum             CCCCCTCGTCGCCG---------------GCTCGTTC--GGGGCGACTAA--CGAACCGC 126
S.nigrum               CCCCCTCGTCGCCG---------------GCTCGTTC--GGGGCGACTAA--CGAACCGC 126
S.gilo                 CCCTCTCAC--------------------CTTTTTTTCTGGG--GCCAAA-CAAACCCC 121
S.torvum               CACTCCTGC--------------------CCC-CTTTGGGAG--GAAAAA-CGAACCTC 143
S.kurzii               CCCTCGAGC--------------------CCCATTTTCGGGG--CCCAAA-TGAACCCC 128
                       * *                          *    *    * *     **    **** *

S.sisymbriifolium      GGCGCGGAAAGCGCCAAGGAATACTTAA-CTGGCAGCCCGCCCCTCGCGCCCCCGTCCGC 201
S.khasianum            GGCGCGGAAAGCGCCAAGGAATACTAAA-TTGAGAGCCCTCCCCCCCGCGCCCC-GTCCGC 200
S.clavatum             GGCGCGAAAAGCGCCAAGGAATACTTAAACTGAGAGCCCTCCCCTCGCGCCCC-GTCCGC 185
S.nigrum               GGCGCGAAAAGCGCCAAGGAATACTTAAACTGAGAGCCCTCCCCTCGCGCCCC-GTCCGC 185
S.gilo                 GATGCGTAAAGTACCAAGGAATACTCAAAC-AAGAACCCTCCACCCGCGCCCC-ATCCAC 179
S.torvum               AACGCGAAAGGTGCCAAGGAATACTCGAAC-GAGAGTAGTCCGCTCGCGCCCC-ATATGC 201
S.kurzii               AGCGTGAAAAGTGCCAAGGAATACTCAAAT-GAGAGCCCTCCACCCGCGCCCC-ATCCAT 186
                          * * ** *  ************  *      *     ** * ********  *

S.sisymbriifolium      GGAGCGCGCCGGG--GATGTGTGCTTCTGTCCGGACCAAAACGACTCTC-GGCAACGGAT 258
S.khasianum            GGAGCGTGCGGGGTGGATGTGTGCTTCTTTCTGAACCAAAACGACTCTC-GGCAACGGAT 259
S.clavatum             GGAGTGTGCGGGG-GGATGCGCGCTTCTTTTGAAACCAAAACGACTCTCCGG-AACGGAT 243
S.nigrum               GGAGTGTGCGGGG-GGATGCGCGCTTCTTTTGAAACCAAAACGACTCTCCGGCAACGGAT 244
S.gilo                 GGAGCGTGC-GGGTGGGGGCATGCTTCTTTCGAAACCAAAATGACTCTC-GGCAATGGAT 237
S.torvum               GTAGCGCTC-GAGCAGATGCGTGCTTCTTGAGAAACCAAAATGACTCTC-GGCAATGGAT 259
S.kurzii               GGAGCATAC-AGATGGATGCGTGCTTTTTTCGAAACCAAAACAACTCTC-GGCAATTGAT 244
                       * **    *      *  *   **** *      *******  ****** ** **  ***

S.sisymbriifolium      ATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAG 318
S.khasianum            ATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAG 319
S.clavatum             ATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAG 303
S.nigrum               ATCTCGGCTCTCGCATCGATGAAGAACGTAGCGAAATGCGATACTTGGTGTGAATTGCAG 304
S.gilo                 ATCTCAGCTCTTGCATCGATGAAGAACGTAGTAAAATGTGATACTTGGTGTGAATTGCAG 297
S.torvum               ATCTCAACTC--GCTTCGATGAAGAACGTAAAGAAATGCGATACTTGGTGTGAATTGCAG 317
S.kurzii               ATCTCGACTCTTGCATCGATGAAGAACATAGAAAAATGTGATACTTGGTATGAATTGTAG 304
                       *****  ***  ** ************ **   ***** ********** ******* **

S.sisymbriifolium      AATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGG 378
S.khasianum            AATCCCGTGAACCATCGAGTTTTTGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGG 379
S.clavatum             AATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGG 363
S.nigrum               AATCCCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAAGCCATTAGGCCGAGGG 364
S.gilo                 AATCTCGTGAACCATCGAGTCTTTCAATCCAAGTTGCGCTCGAAGCCGTCATACTGAGGG 357
S.torvum               AATCCCGTGAACCATCGAGTCTTTGAACCCAAGTTGCGCCCAAAGCCGTCAGGCCAAGGG 377
S.kurzii               AATCCCATGAACCATCGAGTCTCTGAACATAAGTTGCACCCAAAGCCATCAGGCCAAGGG 364
                       **** * ************* * * **   ******* * * ***** * *  *  ****
```

```
S.sisymbriifolium       CACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCCGCACGCCGCACGGCGTCGCGGTG 438
S.khasianum             CACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCTGCACGCCGCACGGCGTCGCGGGG 439
S.clavatum              CACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCCGAACGCCGCAAGGCGTCGTGG-G 422
S.nigrum                CACGTCTGCCTGGGCGTCACGCATCGCGTCGCCCCCCGAACGCCGCAAGGCGTCGTGG-G 423
S.gilo                  TACATCTGCCTGGGCGTCAGGCATCGTGTCACCCCC-AAATGTCGCTCGACGTCGTGGGG 416
S.torvum                CACGTCTGCCTAGGCATCACGCATCACGTCTCCCCACTCCGGTCGCTCGGTATCGCGGGG 437
S.kurzii                TATGTCTGCCTCCGGATCACGCATCACGTCATCCCCCACACATTGCTCAGTATCATGGGG 424
                         *  *******  *  *** *****  ***  ***         **      **  ** *

S.sisymbriifolium       GCGGAGACTGGCCTCCCGTGCGCCTCGT-GCTCGCGGCTGGCCCAAATGCGAGTCCACGT 497
S.khasianum             GCGGATACTGGTCTCCCGTGCGCCTCGA-GCCCGCGGCTGGCCTAAATGCGAGTCCACGT 498
S.clavatum              GCGGATACTGGCCTCCCGTGCGCCTCGA-GCTCGTGGCTGGCCTAAATGCGAGTCCACGT 481
S.nigrum                GCGGATACTGGCCTCCCGTGCGCCTCGA-GCTCGTGGCTGGCCTAAATGCGAGTCCACGT 482
S.gilo                  GCGGATACTGGCCTCCCATGCACCTCACCGGCCGCGGCCAGCCTAAATGCGAGTCTACAT 476
S.torvum                GAGGATCCTGGCCTCCCGTGCACCTCGC-ACCCACGGCCGACCTAAATGTCCGGAAACGT 496
S.kurzii                CCAAATACTGGCCTCCCGTCCGG-TCGT-GCCCGTGGCCAGTCTAAATGTGAGTCCACAT 482
                            *  **** ***** * *   **      *  ***    * *****   *   ** *

S.sisymbriifolium       CGACGGACGTCCCGCCATGTGGTGGTTGTAACTCAACTCTCTTGGTGCCGCGGCCACAGC 557
S.khasianum             CGACGGACGTCGCGGCATGTGGTGGTTGTAACTCAACTCTCTTGGTGCCGCGGCCACAGC 558
S.clavatum              CGACGGACGTCGCGGCAAGTGGTGGTTGAAACTCAACTCTCTTTGTGTCGAGGCTACAGC 541
S.nigrum                CGACGGACGTCGCGGCAAGTGGTGGTTGAAACTCAACTCTCTTTGTGTCGCGGCTACAGC 542
S.gilo                  CGATGAACATCACGGCAAGTGGTGGTTTTAACTCAACTCTCTTGGTG-CGCGGCTACAGC 535
S.torvum                CGACGTACATAGCGGAAAGTGGTGGTTGTGACTCAACTCTCTTGGCGCCGCGACTATGGC 556
S.kurzii                CTACGGACATCGCGGTAAGTAGTAGTTGTAACTCAACTCTCTTAGTCCCATGACTCCAGC 542
                        * * * ** *  **  * ** ** ***   ************ *   *  * *    **

S.sisymbriifolium       CCGTCGCGCGTGCGGACTCCCCGACCCTCCCAGCGCTCAGGCGCTCCGACCGCGACCCCA 617
S.khasianum             CCGTCGCGCGTGCGGACTCCCCGACCCTCCCAGCGCTCAGGCGCTCCGACCGCGACCCCA 618
S.clavatum              CCGTCGCGCGTCCCGACTCCA-GACCCTCTAAGCGCTTAGGCGCTCCGACCGCGACCCCA 600
S.nigrum                CCGTCGCGCGTCCCGACTCCA-GACCCTCTAAGCGCTTAGGCGCTCCGACCGCGACCCCA 601
S.gilo                  CCGACACCCG-GCGCA---------TCACTCATCAATGAGATGCAGAGACG--------- 576
S.torvum                CCGTCGTGCGTGCGTGCTCCCCGACCCTTCCGGCACCATCGCGCTCCGACCGCGACCAGC 616
S.kurzii                CCGAAA-------------CCCGGTC------GCA-----ACGCCTATGCAGAGACG--- 575
                        ***                              *        **     *

S.sisymbriifolium       A---- 618
S.khasianum             A---- 619
S.clavatum              AGATT 605
S.nigrum                AGATT 606
S.gilo                  -----
S.torvum                -----
S.kurzii                -----
```

APPENDIX 3

CLUSTAL 2.1 multiple sequence alignment of *rbcL*

```
S.kurzii              ----------GATCAG-CTGGTGTTAAAGAGTACAAATCGACTTATTATACTCCT-GAGT 48
S.gilo                ----------GATCAG-CTGGTGTTAAAGAGTACAAATCGACTTATTATACTCCT-GAGT 48
S.sisymbriifolium     ------------------GGTGTCAA-GAGTACAAATTGACTTATTATACTCCT-GAGT 39
S.khasianum           ----------AGTCGT-CTAGTGTTAA-GAGT-CAAATTGACTTATTATACTCCT-GAGT 46
S.nigrum              -----TTGGAGATCGGTCTGGTGTCAA-GAGTACAAATTGACTTATTATACTCCTGCAGT 54
S.clavatum            -----TTGGAGATCGGTCTGGTGTCAA-GAGTACAAATTGACTTATTATACTCCTGCAGT 54
S.torvum              TCCGTCTTGAATTCAG-CTGGTGTTAA-GAGTACAAATCGACTTATTATACCCTCTGAGT 58
                                  **** ** **** ***** *********** *    ***

S.kurzii              ACCAAACCAAGGATACTAATATATTGGCAGCATTCCGAGAAACTCCTCAACCTGGAGTTC 108
S.gilo                ACCAAACCAAGGATACTGATATATTGGCAGCATTCCGAGAAACTCCTCAACCTGGAGTTC 108
S.sisymbriifolium     ACCAAACCAAAGATACTGATATATTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTC 99
S.khasianum           ACCAAACCAAGGATACTGATATATTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTC 106
S.nigrum              ACCAAACCAAGGATACCGATATATTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTC 114
S.clavatum            ACCAAACCAAGGATACCGATATATTGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTC 114
S.torvum              ACCAAACCAAGGATACTGAAAC--TGGCAGCATTCCGAGTAACTCCTCAACCTGGAGTTC 116
                      ********** *****  * *   *************** ********************

S.kurzii              CACCTGAAGAAAAAGGGGACGCGGCAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 168
S.gilo                CACCTGAAGAAGCAGGGGCCGCGGCAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 168
S.sisymbriifolium     CACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 159
S.khasianum           CACCTGAAGAAGCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 166
S.nigrum              CACCCGAAGAAGCGGGGACCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 174
S.clavatum            CACCCGAAGAAGCGGGGACCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 174
S.torvum              CACCCGAAGAAGCAGGGGCCGCGGTAGCTGCCGAATCTTCTACTGGTACATGGACAACTG 176
                      **** ******   ***  ***** ************************************

S.kurzii              TATGGACCGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 228
S.gilo                TATGGACCGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 228
S.sisymbriifolium     TATGGACCGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 219
S.khasianum           TATGGACCGATGGACTTACCAGTCTTGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 226
S.nigrum              TATGGACCGATGGACTTACCAGTCTCGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 234
S.clavatum            TATGGACCGATGGACTTACCAGTCTCGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 234
S.torvum              TATGGACCGATGGACTTACCAGTCTCGATCGTTACAAAGGGCGATGCTACCGCATCGAGC 236
                      ************************ ************************************

S.kurzii              GTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGCAGCTTACCCTTCAGACCTTTTTG 288
S.gilo                GTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGCAGCTTACCCTTCAAACCTTTTTG 288
S.sisymbriifolium     GTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGCTTACCCCTTAAACCTTTTTG 279
S.khasianum           GTGTTGTTGGAGAAAAAGATCAATATATTGCTTATGTAGCTTACCCTTTAGACCTTTTTG 286
S.nigrum              GTGTTATTGGAGAACAAGATCAATATATTGCTTATGTAGCTTACCCTTTAGACCTTTTCG 294
S.clavatum            GTGTTATTGGAGAACAAGATCAATATATTGCTTATGTAGCTTACCCTTTAGACCTTTTCG 294
S.torvum              GTGTTATTGGAGAAAAAGATCAATATATTGCTTATGTAGCTTACCCTTTAGACCTTTTCG 296
                      ***** ******** ********************* ********* * * ******* *

S.kurzii              AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGCAGGTAATGTATTTGGGTTCAAAG 348
S.gilo                AAAAAGGTTCCGTTACCAATATGTTTACTTCCATTGCAGGTAATGTATTTGGGTTCAAAG 348
S.sisymbriifolium     AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTCAAAG 339
S.khasianum           AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGTAGGTAATGTATTTGGGTTCAAAG 346
S.nigrum              AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGTAGGCAATGTATTTGGGTTCAAAG 354
S.clavatum            AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGTAGGCAATGTATTTGGGTTCAAAG 354
S.torvum              AAGAAGGTTCCGTTACCAATATGTTTACTTCCATTGTAGGCAATGTATTTGGGT--AAGG 354
                      ** ********************************* *** *************  ** *

S.kurzii              CCC--TGCGCGCTCTACGTCTGGAAGATCTGCGAATCCCTCCTGCTTATGTTAAAACTTT 406
S.gilo                CCC--TGCGCGCTCTACGTCTGGAAGATCTGCGAATCCCTCCTGCTTATGTTAAAACTTT 406
S.sisymbriifolium     CCC--TGCGCGCTCTACGTCTGGAAGATCTGCAAATCCCTCCTGCTTATGTTAAAACTTT 397
S.khasianum           CCC--TGCGCGCTCTACGTCTGGAAGATCTGCGAATCCCTCCTGCTTATGTTAAAACTTT 404
S.nigrum              CCC--TGCGCGCTCTACGTCTGGAAGATCTGCGAATCCCTCCTGCTTATACCAAAACTTT 412
S.clavatum            CCC--TGCGCGCTCTACGTCTGGAAGATCTGCGAATCCCTCCTGCTTATACCAAAACTTT 412
S.torvum              CCCCTTGCGCGCTCTAAGTCTGGAAGATCCGCGAATTCCCTTAGTATATGTCAAAACTTT 414
                      ***  *********** ************ ** *** **    *  ***   ********
```

```
S.kurzii              CCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAATCGAACAAGTATGGTCGTCC 466
S.gilo·               CCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAATCGAACAAGTATGGTCGTCC 466
S.sisymbriifolium     CCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAATTGAACAAGTATGGTCGTCC 457
S.khasianum           CCAAGGTCCGCCTCATGGGATCCAAGTTGAAAGAGATAAATTGAACAAGTATGGTCGTCC 464
S.nigrum              CCAAGGTCCGCCTCATGGGATCCAAGTCGAAAGAGACAAATTGAACAAGTATGGTCGTCC 472
S.clavatum            CCAAGGTCCGCCTCATGGGATCCAAGTCGAAAGAGACAAATTGAACAAGTATGGTCGTCC 472
S.torvum              CCAAGGTCCGCCTTAAGGGATCCAAGTCGAAAGAGACAAATTGAACAAGTATTGTCGTCC 474
                      ************ * *********** ******** **** ********** *******

S.kurzii              CCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCCGCTAAAAACTACGGTAGAGC 526
S.gilo                CCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCTAAAAACTACGGTAGAGC 526
S.sisymbriifolium     CCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCTAAAAACTACGGTAGAGC 517
S.khasianum           CCTGTTGGGATGTACTATTAAACCTAAATTGGGGTTATCTGCTAAAAACTACGGTAGAGC 524
S.nigrum              CCTGTTGGGATGTACTATCAAACCCAAATTGGGGTTATCTGCCAAAAACTACGGCAGAGC 532
S.clavatum            CCTGTTGGGATGTACTATCAAACCCAAATTGGGGTTATCTGCCAAAAACTACGGCAGAGC 532
S.torvum              CCTGTTGGGATGTACTATCAAACCCAAATTGGGGTTATCTGCCAAAAACTACGGCAGAGC 534
                      ****************** ***** ************** ** ********** *****

S.kurzii              TGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCAAAGATGATGAGAACGCGAACTC 586
S.gilo                TGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCAAAGATGATGAGAACGCGAACTC 586
S.sisymbriifolium     TGTTTATGAATGTCTTCCCGGTGGACTTGATTTTACCAAAGATGATGAGAACGTGAACTC 577
S.khasianum           TGTTTATGAATGTCTTCGCGGTGGACTTGATTTTACCAAAGATGATGAGAACGTGAACTC 584
S.nigrum              TGTTTACGAATGTCTTCGCGGTGGACTCGATTTTACCAAAGACGACGAGAACGTAAACTC 592
S.clavatum            TGTTTACGAATGTCTTCGCGGTGGACTCGATTTTACCAAAGACGACGAGAACGTAAACTC 592
S.torvum              TGTTTACGAATGTCTTCGCGGTGGGTTCGATTTTACCAAAGACGACGAGAACGTGAACTC 594
                      ****** ********** ******  * ************** ** *******  *****

S.kurzii              ACAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCCCTTTTTAAAGC 646
S.gilo                ACAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCCCTTTTTAAAGC 646
S.sisymbriifolium     ACAACCATTTATGCGTTGGAGAGATCGTTTCTTATTTTGTGCCGAAGCACTTTTTAAAGC 637
S.khasianum           ACAACCATTTATGCGTTGGAGAGATCGTTTCTTATTTTGTGCCGAAGCACTTTTTAAAGC 644
S.nigrum              CCAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCACTTTTCAAAGC 652
S.clavatum            CCAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCACTTTTCAAAGC 652
S.torvum              ACAACCATTTATGCGTTGGAGAGATCGTTTCTGCTTTTGTGCCGAAGCACTTTTCAAAGC 654
                       ******************************  ************** ***** *****

S.kurzii              ACAGGCTGAAACAGGTGAAATCAAAGGACATTACTCGAATGCTACTGCAGGTACATGCGA 706
S.gilo                ACAGGCTGAAACAGGTGAAATCAAAGGACATTACTCGAATGCTACTGCAGGTACATGCGA 706
S.sisymbriifolium     ACAGACTGAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA 697
S.khasianum           ACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA 704
S.nigrum              ACAGGCCGAAACAGGCGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA 712
S.clavatum            ACAGGCCGAAACAGGCGAAATCAAAGGGCATTACTTGAATGCTACTGCAGGTACATGCGA 712
S.torvum              ACAGGCCGAAACAGGCGAAATCAAAGGACATTACTTGAATGCTACTGCAGGTACATGCGA 714
                      **** * ******** ********** ******* ************************

S.kurzii              AGAAACGATCAAAAGAGCTGTATTTGCTAGAGAATTGGGCGTTCCGATCGCAATGCATGA 766
S.gilo                AGAAACGATCACAAGAGCTGTATTTGCTAGAGAATTGGGCGTTCCGATCGCAATGCATGA 766
S.sisymbriifolium     AGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGA 757
S.khasianum           AGAAATGATCAAAAGAGCTGTATTTGCTAGAGAATTGGGCGTTCCGATCGTAATGCATGA 764
S.nigrum              AGAAATGATCAAAAGAGCTGTATTTGCCAGAGAATTGGGCACTCCGATCGTAATGCACGA 772
S.clavatum            AGAAATGATCAAAAGAGCTGTATTTGCCAGAGAATTGGGCACTCCGATCGTAATGCACGA 772
S.torvum              AGAAATGATGAAAAGAGCTATATTTGCCAGAGAATTGGGCGTTCCGATCGTAATGCACGA 774
                      ***** *** * ******* ******* ************  ******** ****** **

S.kurzii              CTACTCAACGGGGGG-ATTCACCGCAAATACTACCTTGGCTCATTATTGCCGAGATAATG 825
S.gilo                CTACTCAACGGGGGG-ATTCACCGCAAATACTACCTTGGCTCATTATTGCCGAGATAATG 825
S.sisymbriifolium     CTACTTAACAGGCGG-ATTCACCGCAAATACTAGCTTGGCTCATTATTGCCGAGATAATG 816
S.khasianum           CTACTTAACGGGGGGGATTCACCGCAAATACCAGCTTGGCTCATTATTGCCGAGACGATG 824
S.nigrum              CTACTTAACGGGGGG-ATTCACCGCAAATACCAGCTTGGCTCATTATTGCCGAGACAATG 831
S.clavatum            CTACTTAACGGGGGG-ATTCACCGCAAATACCAGCTTGGCTCATTATTGCCGAGACAATG 831
S.torvum              CTACTTAACGGGGGG-ATTCACCGCAAATACCAGCTTGGCTCATTATTGCCGAGACAATG 833
                      ***** *** ** ** *************** * *********************  ***

S.kurzii              GTCTACTTTCTTCACATCCACCG-GCAATGCATGCGGTTACCGATTCACCGCAAATTCAT 884
S.gilo                GTCTACTTTCTTCACATCCACCG-GCAATGCATGCGGTTATTGATAGACAGAAGAATCAT 884
S.sisymbriifolium     GTCTACTT-CTTCACATCCACCG-GAAATGCATGCGGTTATCGATTCACCGCAAATTCAT 874
S.khasianum           GTCTGTTT-ATGAATATCCTCCG-CAA-TGCATGCGGTTAACGATTCCCCGCAAATTCAT 881
S.nigrum              GTCTACTT-CTTCACATCCACCG-GCAATGCATGCGGTTATTGATAGACAGAAGAAACAT 889
S.clavatum            GTCTACTT-CTTCACATCCACCG-GTAATGCATGCGGTTACCGATTCACCGCAAATTCAT 889
S.torvum              GTCTACTT-CTTCACATCCACCGTGCAATGCATGCGGTTATCGACAGACAGAAGAATCAT 892
                      ****  **  *  * **** ***   * *************  **    * * * *  ***
```

```
S.kurzii              GGTATCCACTTCCGGGTATCAGCAAAAGCGTTACGTATGTCTGGTGGAGATCATATTCAC 944
S.gilo                GGTATCCACTTCCGGGTATCAGCAAAAGCGTTACGTATGTCTGGTGGAGATCATATTCAC 944
S.sisymbriifolium     GGTATCCACTTCCGGGTATCAGCAAAACCGTCAAGTATGTTTGGTGGAGCTCATGTTCAT 934
S.khasianum           GGTATCCACTTCCGGGTATCAGCAAAACCGTCAAGTATGTCTGGTGGAGCTCATGTTCAC 941
S.nigrum              GGTATCCACTTCCGGGTATTAGCAAAAGCGTTACGTATGTCTGGTGGAGATCATATTCAC 949
S.clavatum            GGTATCCACTTCCGGGTATTAGCAAAAGCGTTAAGTATGTCTGGTGGAGATCATGTTCAC 949
S.torvum              GGTATCCACTTCCGGGTATTAGCAAAAGCGTTACGTATGTCTGGTGGAGATCATATTCAC 952
                      ****************** ******* *** * ****** ******** **** ****

S.kurzii              GCTGGTACCGCAGCAGGTAAACTTGAAGGTGAAAGAGACACAACTTTGGGCTTTGTTGAT 1004
S.gilo                GCTGGTACCGCAGCAGGTAAACTTGAAGGTGAAAGAGACACAACTTTGGGCTTTGTTGAT 1004
S.sisymbriifolium     TGCGGTTCCGTCGCAGGTAAACTTGAAGGTGAAAGAGACACAACGTTGGGCTTTGTTGAT 994
S.khasianum           TCCGGTTCCGTCGCAGGTAAACTTGAAGGTGAAAGAGACACAACGTTGGGCTTTGTTGAT 1001
S.nigrum              TCTGGTACCGTAGTAGGTAAACTTGAAGGTGAAAGAGACATAACTTTGGGCTTTGTTGAT 1009
S.clavatum            TCTGGTACTGTAGTAGGTAAACTTGAAGGTGAAAGAGAAATAACTTTGGGCTTTGTTGAT 1009
S.torvum              GCTGGTACCGTAGTAGGCAAACTCGAAGGCGAAAGAGACATAACTTTGGGCTTTGTCGAT 1012
                         *** * *  * *** ***** ***** ******** * *** ********** ***

S.kurzii              TTACTGCGTGATGATTTTGTTGAACAAGATAGAAGTCGCGGTATTTATTTCACTCAAGAT 1064
S.gilo                TTACTGCGTGATGATTTTGTTGAACAAGATAGAAGTCGCGGTATTTATTTCACTCAAGAT 1064
S.sisymbriifolium     TCAATGCGTGATGATTTTGTTGAACCAGTTTGAAGTAGCGGTATTTATTTCAGTCACGAT 1054
S.khasianum           TTACTGCGTGATGATTTTGTTGAACAAGATTGAAGTCGCGGTATTTATTTCACTCAAGAT 1061
S.nigrum              TTACTGCGTGATGATTTTATTGAACAAGATAGAAGTCGCGGTATTTATTTCACTCAAGAT 1069
S.clavatum            TTACTGCGTGATGATTTTGTTGAACAAGATAGAAGTCGCGGTATTTATTTCACTCAAGAT 1069
S.torvum              TTACTGCGCGACGATTTTATCGAAGAAGACAGAAGTCGCGGTATTTATTTCACTCAAGAT 1072
                      * * **** ** ****** * **** **   ***** *************** *** ***

S.kurzii              TGGGTCTCTTTACCAGGTGTTATACCTGTGGCTTCAGGAGGTATTCACGTTTGGCATATG 1124
S.gilo                TGGGTCTCTTTACCAGGTGTTATACCTGTGGCTTCAGGAGGTATTCACGTTTGGCATATG 1124
S.sisymbriifolium     TGCGTCTCTTTACCAGGTGTTTCCCCCGTCGCTTCAGGAGGTATTCCCTTTCCGCATATG 1114
S.khasianum           TGCGTCTCTTTACCAGGTGTTCTACCTGTGGCTTCAGGAGGTATTCCCTTTTGGCATATG 1121
S.nigrum              TGGGTCTCTTTACCAGGTGTTCTACCTGTGGCTTCAGGAGGTATTCACGTTTGGCATATG 1129
S.clavatum            TGGGTCTCTTTACCCGGTGTTTTACCTGTGGCTTCAGGAGGTATTCACGTTTGGCATATG 1129
S.torvum              TGGGTCTCTTTACCAGGTGTTCTACCTGTGGCTTCAGGAGGTATTCACGTTTGGCATATG 1132
                      ** ********** ******   ** ** **************** * **  *******

S.kurzii              CCTGCTCCGACCGAGATCTTTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTCA 1184
S.gilo                CCTGCTCCGACCGAGATCTTTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTCA 1184
S.sisymbriifolium     CCTGTTACGACCGAGATCTCTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTCA 1174
S.khasianum           CCTGCTCCGACCGAGATCTTTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTCA 1181
S.nigrum              CCTGCTCTGACCGAGATCTTTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTTA 1189
S.clavatum            CCTGCTCTGACCGAGATCTTTGGGGATGATTCCGTACTACAGTTCGGTGGAGGAACTTTA 1189
S.torvum              CCTGCTCTGACCGAGATCTTTGGGGACGATTCCGTACTACAGTTCGGTGGAGGAACTTTA 1192
                      **** *  ********** ****** ******************************** *

S.kurzii              GGACATCCTTGGGGTAATGCGCCAGGTGCCGCAGCTAATCGAGCAGCTCCAGAAGCATGT 1244
S.gilo                GGACATCCTTGGGGTAATGCGCCAGGTGCCGCAGCTAATCGAGCAGCTCCAGAAGCATGT 1244
S.sisymbriifolium     GGACATCCTTGGGGTAACGCGCCAGGTCCGGCAGATAATTGAGCAGCTGCAGAAGCATGA 1234
S.khasianum           GGACATCCTTGGGGTAATGCGCCAGGTCCCGCAGCTAATCGAGCAGCTCCAGAAGCATGT 1241
S.nigrum              GGACATCCTTGGGGTAATGCGCCAGGTGCCGTAGCTAATCGAGTAGCTCTAGAAGCATGT 1249
S.clavatum            GGACATCCTTGGGGTCACCCTCCTGGTGCCGTAGCTAATCGAGTAGCTCTAGAAGCATGT 1249
S.torvum              GGACATCCTTGGGGCAATGCGCCAGGTGCCGTAGCCAATCGAGTAGCTCTAGAAGCATGT 1252
                      ************** *  * ** *** * * **  *** *** ****  *********

S.kurzii              GCAAAACCTCGTAATGA-AGGACGTGATCTTTGCTCAGGAAGGTAATGAGATTATTCGCG 1303
S.gilo                GCAAAAGCTCGTAATGA-AGGACGTGATCTTTGCTCAGGAAGGTAATGAGATTATTCGCG 1303
S.sisymbriifolium     GCAAAAGCT-ACAACGA-AGGACGCGATCTT-GCTCGGGAGGGCAACGAGAT-GTCCGCG 1290
S.khasianum           GCAAAAGCTCGTAATGCTAGGACGTGATCTT-GCTCGGGAAGGTAATGAGATTATCCGCG 1300
S.nigrum              GTAAAAGCTCGTAATGA-AGGACGTGATCTT-GCTCGGGAAGGTAATGAGATTATTCGCG 1307
S.clavatum            GTAAGGTCTTATAATGA-AGGACGTGATCTT-GCTCGTGGAGGTAATGAGATTATTCGCG 1307
S.torvum              GTAAAAGCTCGTA-TGA-AGGACGTGATCTT-GCTCAGGAAGGTAATGAGATTATTCGCG 1309
                      * **   **   *  *  ****** ****** ****  *  ** ** *****  * ****

S.kurzii              AGGCTCCCAACTGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1363
S.gilo                AGGCTTCCAAATGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1363
S.sisymbriifolium     AGGCTCCAAACTGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1350
S.khasianum           AGGCTCCAAACTGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1360
S.nigrum              AGGCTGCCAACTGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1367
S.clavatum            AGGCTGCCAACTGGAGCCCGGAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1367
S.torvum              AGGCTTCCAAATGGAGCCCGAAACTAGCTGCTGCTTGTGAGGTATGGAAAGAGATCGTAT 1369
                      ***** * ** ********* ************************************
```

```
S.kurzii              TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1398
S.gilo                TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1398
S.sisymbriifolium     TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1385
S.khasianum           TTAATTTTGCAGCAGTAGACGTTTTGGATAAGTAA 1395
S.nigrum              TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1402
S.clavatum            TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1402
S.torvum              TTAATTTTGCAGCAGTGGACGTTTTGGATAAGTAA 1404
                      ********** ********** **********
```

APPENDIX 4

CLUSTAL 2.1 multiple sequence alignment of *matK*

```
S.kurzii              GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAC 60
S.gilo                GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAC 60
S.sisymbriifolium     GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAC 60
S.torvum              GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAC 60
S.khasianum           GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAG 60
S.clavatum            GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTA---TCCCAAG-GTGAAAC 56
S.nigrum              GGTCGATTTTGTTGGAAAATTCAGGTTATAACAATAAATTTAGTTTCCTAATTGTGAAAC 60
                      ****************************************   *** **  ******

S.kurzii              GTTTAATTACTCGAATGTATCAACAGAATCATTTTCTTATTTCTACTAATGATTCTAACA 120
S.gilo                GTTTAATTACTCGAATGTATCAACAGAATCATTTTCTTATTTCTACCAACGATTCCAACA 120
S.sisymbriifolium     GTTTAATTACTCGAATGTATCAACAGAATCATTTTCTTATTTCTACTAATGATTCTAACA 120
S.torvum              GTTTAATTACTCGAATGTATCAACAGAATCATTTTCTTATTTCTACTAATGATTCTAACA 120
S.khasianum           GTTTAATTACTCGAATGTATCAACAGAATCATTTTCTTATTTCTATCAACGATTCTACCA 120
S.clavatum            GTTTAATTAT--GAAAGAATCAACAGAATCATTTTATTATTTCTACTAATGATTCTAACA 114
S.nigrum              GTCGAATTACGCGAATGTGG-GTCAGAATCATTTTCTTATTTCTACTAATGATTCTAACA 119
                      **  *****   *** *     ************ *********  ** ***** * **

S.kurzii              AAAATCCATTTTTGGGGGGGAACAAGAGTTTGTATTCTCAAATGATATCAGAGGGATTTT 180
S.gilo                AAAATCCATTTTTGGGGTGCAACAAAAGTTTGTATTCTCAAATGATATCAGAGGGATTTT 180
S.sisymbriifolium     AAAATCCATTTTTGGGGTGCAACAAGAGTTTGTATTCTCAAATGATATCAGAGGGATTTG 180
S.torvum              AAAATCCATTTTTGGGGTGCAACAAGAGTTTGTATTCTCAAATGATATCAGAGGGATTTT 180
S.khasianum           AAAATCCATTTTTGGGGTGCAACAAGAGTTTGTATTCTCAAATGATATCAGAGGGATTTG 180
S.clavatum            AAAATCCTTTTTTGGGGTGCAACAAGAGTTTGTATTCTCAAACGATTTCAGAGGGATTTG 174
S.nigrum              AAAATCCATTTTTGGATTGCAACTCGAGTTTGTATTCTCAAACGATATCAGAGGGATTTG 179
                      ******* *******   * ***   ***************** *** ************

S.kurzii              CATGTATTGTGGAAATTCCATTTTCTCTACGATTAATATCTTCTTTATCTTCTTTCGAA- 239
S.gilo                CATGTATTGTGGAAATTCCATTTTCTCTACGATTAATATCT---------TCTTTCGAA- 230
S.sisymbriifolium     CATGTATTGTGGAAATTCCGTTTTCTCTACGATTAATATCTTCTTTATCTTCTTTCGAA- 239
S.torvum              CATGTATTGTGGAAATTCCGTTTTCTCTACGATTAATATCTTCTTTATCTTCTTTCGAAA 240
S.khasianum           CATGTATTGTGGAAATTCCGTTTTCTCTACGATTAATATCTTCTTTATCTTCTTTCGAA- 239
S.clavatum            CATGTATTGTGGAAATTCCGTTTTCTCTACGATCAATATCT---------TCTTCAGAA- 224
S.nigrum              CATGTATTGTGGAAATTCCGTTTTCTCTACGATCAATATCTTCTTTATCTTCTTTCGAA- 238
                      ****************** ************ *******         ****  ***

S.kurzii              GAAAAAAAGATTTTCAAATCTCATAATTTACGATCAATTCATTCAACATTTCCTTTTTTA 299
S.gilo                GAAAAAAAGATTTTCAAATCTCACAATTTACGATCAATTCATTCAACATTTCCTTTTTTA 290
S.sisymbriifolium     GGCAAAAAGATTTTCAAATCTCATAATTTACGATCAATTCATTCAACATTTCCTTTTTTA 299
S.torvum              GGAAAAAAGATTTTCAAATCTCATAATTTACGATCAATTCATTCAACATTTCCTTTTTTA 300
S.khasianum           GGCAAAAAGATTTTCCAATCTCACAATTTACGATCAATTCATTCAACATTTCCTTTTTTA 299
S.clavatum            GGCAAAAAGATTTTCAAATCTCATAATTTACGATCAATTCATTCAACATTTCCTTTTTCA 284
S.nigrum              GGCAAAAAGATTTTCAAATCTCATAATTTACGATCAATTCATTCAACATTTCCTTTTTCA 298
                      *  ************ ******* ********************************** *

S.kurzii              GAGGA-CAATTTGTC-ACATCTAAATTATGTATTA-GATATACTAATACCCTA-CCCCC- 354
S.gilo                GAGGA-CAATTTGTC-ACATCTAAATTATGTATTA-GATATACTAATACCCTA-CCCC-- 344
S.sisymbriifolium     GAGGGACAATTTTTTCCACATCCAAATTATGTATCA-GATATACCAATACCCTA-CCCC-G 356
S.torvum              GAGGACCAATTTGTC-ACATCTAAATTATGTATTAAGATATACCAATACCCTA-CCCCCG 358
S.khasianum           GAGGA-CAATTTTTC-ACATCTAAATTATGTATTA-GATATACTAATACCCTAACCCCCG 356
S.clavatum            GAGGA-CAATTTTTC-CTATTCAAATTATGTATCA-GATATACCAATACCTTTACCCC-- 339
S.nigrum              GAGGA-CAATTTTTC-ACATCCAAATTATGTATCA-GATATACCAATACCCTA-CCCC-- 352
                      ****  ****** **   **  *********** * ******* ****** *  ****

S.kurzii              GTTCATCT--GGAAATCTTGGTTCAAACTCTTCGTTATTGGGTAAAA-GATGCCTCTTCT 411
S.gilo                GTTCATCT--GGAAATCTTGGTTCAAACTCTTCGTTATTGGGTAAAA-GATGCCTCTTCT 401
S.sisymbriifolium     GTTCATCT--GGAAATCTTGGTTCAAACTCTTCGCTATTGGGTAAAA-GATGCCTCTTCT 413
S.torvum              GTTCATCT--GGAAATCTTGGTTCAAACTCTTCGCTATTGGGTAAAAAGATGCCTCTTCT 416
S.khasianum           TTTCATCCTGGAAAAATCTTGGTTCAAACTTTTCGCTATTGGGTAAAACGATGCCTCTTCT 416
S.clavatum            GTTCATCT--GGAAATCTTGGTTCAAAATATTCGCTATTGGGTAAAA-GATGCTTCTTCT 396
S.nigrum              GTTCATCT--GGAAATCTTGGTTCAAACTCTTCGCTATTGGGCAAAA-GATGCCTCTTCT 409
                       ******   * *************** * **** ******* **** ***** ******
```

```
S.kurzii              TTACATTTATTACGATTCTTT-CTCCACGAAATATTGTAATTCGAATAGTCTTATTACTT 470
S.gilo                TTACATTTATTACGATTCTTT-CTCCACGAA-TATTGCAATTTGAACAGTCTTATTACTT 459
S.sisymbriifolium     TTACATTTATTACGATTCTTT-CTCCACGAA-TATTGTAATTTGAATAGTCTTATTACTT 471
S.torvum              TTACATTTATTACGATTCTTT-CTCCACGAA-TATTGTAATTCGAATAGTCTTATTACTT 474
S.khasianum           TTACATTTAGTACGATTTTTTTCTCCACGAA-TATTGTAATTGGAATAGTCTTATTACTT 475
S.clavatum            TTACATTTATTACGATTCTTT-CTCCACGAA-TATTGTAATTTGAATAGTCTTATTACTT 454
S.nigrum              TTACATTTATTACGATTCTTT-CTCCACGA-----TACAGT---GGCAGTCTTATTTCTT 460
                      ********* ******* *** ********     *  * *      ********* ***

S.kurzii              CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAA------ATCCAAA-A-TTCTTCTT 520
S.gilo                CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAATCCAAAATTCAAA-AATTCTTCTT 516
S.sisymbriifolium     CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAA--------TCAAA-GATTCTTCTT 520
S.torvum              CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAA--------TCAAA-GATTCTTCTT 523
S.khasianum           CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAA--------TCACA-GATTCTTCTT 524
S.clavatum            CAAAGAAGCC-CGGTTACTCT-TTTTCAAAAAAAAA--------TCAAA-GATTCTTCTT 503
S.nigrum              CAAAGAAGCCACGGTTACTCCGTTTTCAACAAAAAC--------TCAAAAGATTCTTCTT 512
                      ********** *********  ******* *****          ** *   ********

S.kurzii              CTTCTTATAT-AATTCTT-ATGTATATGAA-TGCGAATCTACTTTCGTCTTTCTACGGAA 577
S.gilo                CTTCTTATAT-AATTCTTTATGTATAGGGA-TGCGAATCTACTTTTGTCTTTCTACGGAA 574
S.sisymbriifolium     CTTCTTATAT-AATTCTT-ATATATATGAA-TGCGAATCTACTTTCGTCTTTCTACGGAA 577
S.torvum              CTTCTTATAT-AATTCTT-ATGTATATGAA-TGCGAATCTACTTTCGTCTTTCTACGGAA 580
S.khasianum           CTTCTTATAT-AATTCTT-ATGTATATGAA-TGCGAATCTACTTTCGTCTTTCTACGGAA 581
S.clavatum            CTTCTTATAT-AATTCTT-ATGTATATGAA-TACGAATCCACTTTCGTCTTTCTACGGAA 560
S.nigrum              CTTCTTATATCAATTCTT-ATATATAAGAAATGCGAATCCACTTTCGTCTTTCTACGGAA 571
                      ********** ******* ** **** * * * ****** ***** **************

S.kurzii              ACAATCTTTT-CATTT-ACGATCAACATCTTTTGGAGCCCTTCTTGAACGAATATATTTC 635
S.gilo                ACAATCTTTTTCATTTTACGATCAACATCTTTTGGAGCCCTTCTTGAACGAATATATTTC 634
S.sisymbriifolium     ACAATCTTTT-CATTT-ACGATCAACATCTTTTGGAGCCCTTCTTGAACGAATATATTTC 635
S.torvum              ACAATCTTTT-CATTT-ACGATCAACATCTTTTGGAGCCCTTCTTGAACGAATATATTTC 638
S.khasianum           ACAATCTTTT-CATTT-ACGATCAACATCTTTTGGATCCCTTCTGGAACGAATATATTTC 639
S.clavatum            ACAATCTTCT-CATTT-ACGATCAACATCTTTTGGAGCCCTTCTTGAACGAATATATTTC 618
S.nigrum              CCAATCTTCT-CATTT-ACGATCAACATCTTGGGGAGCCCTTCTTGAACGAATATATTTC 629
                       ******* * ***** **************  *** ******* ***************

S.kurzii              TATGGAAAAATAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTATGG 695
S.gilo                TATGGAAAAATAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTATGG 694
S.sisymbriifolium     TATGGAAAAATAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTATGG 695
S.torvum              TATGGAAAAATGGAACGTCTTGCAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTATGG 698
S.khasianum           TATGGAAAAACAGAACGTCTTGCAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTATGG 699
S.clavatum            TATGGAAAAATAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTTCAGCTTACCCTATGG 678
S.nigrum              TATGGAAAAATAGAACGTCTTGTAGAAGTCTTTGCTAAGGATTTTCAGGTTACCCTCTGG 689
                      **********  ********** ************************* ******* ***

S.kurzii              TTATTCAAGGATCCTTTGATGCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 755
S.gilo                TTATTCAAGGATCCTTTGATGCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 754
S.sisymbriifolium     TTATTCAAGGATCCTTTGATGCATTATGTTAGGTATGAAGGAAAAGCAATTCTGGCTTCA 755
S.torvum              TTATTCAAAGATCCTTCGATGCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 758
S.khasianum           TTATTCAAGGATCCTTCGATGCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 759
S.clavatum            TTATTCAAGGATCCTTTGATGCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 738
S.nigrum              TTATTCAAGGACCCTTTCATCCATTATGTTAGGTATGAAGGAAAATCAATTCTGGCTTCA 749
                      ******** ** ****  ** ************************* **************

S.kurzii              AAAGGGACGTTTCTTTTGATGAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 815
S.gilo                AAAGGGACGTTTCTTTTGATGAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 814
S.sisymbriifolium     AAAGGGACGTTTCTTTTGATGAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 815
S.torvum              AAAGGGACGTTTCTTTCGACGAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 818
S.khasianum           AAAGGGACGTTTCTTTCGACGAATAAATGGAAATTTTRCCTTGTCAATTTTTGGCAATGT 819
S.clavatum            AAAGGGACGTTTCTTTTGATTAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 798
S.nigrum              AAAGGGACGTTTCTTTTGATGAATAAATGGAAATTTTACCTTGTCAATTTTTGGCAATGT 809
                      **************** **  *************** ***********************

S.kurzii              CATTTTTCTATGTACTTTCACACAGGAA-GGATCCATATAAA-CCAATTATCC-AACCAT 872
S.gilo                CATTTTTCTATGTACTTTCACACAGGAA-GGATCCATATAAA-CCAATTATCC-AACCAT 871
S.sisymbriifolium     CATTTTTCTATGTACTTTCACACAGGAA-GGATCCATATAAA-CCAATTATCC-AACCAT 872
S.torvum              CATTTTTCTATGTACTTTCACACAGGAAAGGATCCATATAAAACCAATTATCCCAACCAT 878
S.khasianum           CATTTTTCTATGTACTTTCACCCAGGAA-GGATCCATACAAA-CCAATTATCC-AACCAT 876
S.clavatum            CATTTTTCTATGTACTTTCACACAGGAA-GGATCCATATAAA-CCAATTATCC-AACCAT 855
S.nigrum              CATTTTTCTATGTACTTTCACACAGGAA-GGATCCATATAAA-CCAATTATCC-AACCAT 866
                      ********************** ****** ********* *** ********** ******
```

```
S.kurzii              TCCC-GTGACTTTA-TGGGCTATCTTTCAAGTGTGCGACTAAA-TCATTCAATGG-TACG 928
S.gilo                TCCC-GTGACTTTA-TGGGCTATCTTTCAAGTGTGCGACTAAA-TCATTCAATGG-TACG 927
S.sisymbriifolium     TCCC-GTGACTTTA-TGGGCTATCTTTCAAGTGTGCGACTAAA-TCATTCAATGG-TACG 928
S.torvum              TCCCCGTGACTTTAATGGGCTATCTTTCAAGTGTGCGACTAAAATCATTCGATGGGTACG 938
S.khasianum           TCCC-GTGWCTTTA-TGGGCTATCTTTCAAGTGTGCGACCAAA-TCATTCAATGG-TACG 932
S.clavatum            TCCC-GTGACTTTA-TGGGCTATCTTTCAAGTGTGCGACTAAA-TCATTCAATGG-TCCG 911
S.nigrum              TCCC-GTGACTTTA-TGGGCTATCTTTCAAGTGTGCGACTAAAATCATTCAATGGGTACG 924
                      **** *** ***** ********************** *** ****** **** * **

S.kurzii              TAGTCAAATGTT-CGAAAATT-CATTT-CTAATCAATAA-TCCAATT--AAGAAATTC-G 981
S.gilo                TAGTCAAATGTT-CGAAAATT-CATTT-CTAATCAATAA-TCCAATT--AAGAAATTC-G 980
S.sisymbriifolium     TAGTCAAATGTT-CGAAAATG-CATTT-CTAATCAATAA-TCCAATTCAAGAAAATTTCG 984
S.torvum              TAGTCAAATGTTTCGAAAATTTCATTTTCTAATCAATAA-TCCAATT--AAGAAATTCGA 995
S.khasianum           TAGTCAAATGTT-CGAAAATG-CATTTCCAAATCAATAAATCCAAATT-AAGAAATTC-G 988
S.clavatum            TAGTCAAATGTT-AGAAAATG-CATTT-CTAATCAATAA-TCCAATT--AAGAAGTTC-G 964
S.nigrum              TAGTCAAATGTTTAGCAAATT-CATTT-CTAATCAATAA-TCCAATT--AAGAAGTTC-G 978
                      ************  * ****  ***** * ********* ***** *  *  ** **

S.kurzii              ATACCCTT-GTT-CCAATT-ATTCG--TTTG--ATT-GGAT-CATT-AGCTAAA-GCACA 1030
S.gilo                ATACCCTT-GTT-CCAATT-ATTCC--TTTG--ATT-GGAT-CATT-AGCTAAA-GCACA 1029
S.sisymbriifolium     ATACCCTT-GTTTCCAATTTATTCCCTTTTG--ATTTGGAT-CATC-AGCTAAAAGCACA 1039
S.torvum              ATACCCTT-GTT-CCAATT-ATTCC--TTTG--ATT-GGAT-CATT-AGCCAAA-GCACA 1044
S.khasianum           ATACCCTTTGTTCCAAATTTATTTCCCTTTGGAATTGGAATTCATTTAGCTAAA-GCACA 1047
S.clavatum            ATACCCTT-GTT-CCAATT-ATTCC--TTTG--ATT-GGAT-CATT-AGCTAAA-GCACA 1013
S.nigrum              ATACCCTT-GTT-CCAATT-ATTCC--TTTG--ATT-GGAT-CATT-AGCTAAA-GCACA 1027
                      ******** *** * **** ***    ****  *** * ** ***  *** *** *****

S.kurzii              -CTTTTGTACCGTATTAGGGCATCCCATTAGTAAACCGGTTTGGTCCGATTTATCAGATT 1089
S.gilo                -CTTTTGTACCGTATTAGGGCATCCCATTAGTAAACCGGTTTGGTCCGATTTATCAGATT 1088
S.sisymbriifolium     ACTTTTGTACCGTATTAGGGCATCCCATCAGCAAACCGGTTTGGTCCGATTTATCAGATT 1099
S.torvum              -CTTTTGTACCGTATTAGGGCATCCCATCAGCAAACCGGTTTGGTCCGATTTATCAGATT 1103
S.khasianum           -CTTTTGTACCGTATTAGGGCATCCCATTAGTAAACCGGTTTGGTTCGATTTATCAGATT 1106
S.clavatum            -CTTTTGTACCGTATTAGGGCATCCCATTAGTAAACCGGTTTGGTCCGATTTATCAGATT 1072
S.nigrum              -CTTTTGTACCGTATTAGGGCATCCCATTAGTAAACCGGTTTGGTCCGATTTATCAGATT 1086
                       *************************** ** ************ **************

S.kurzii              CTGATATTATTGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTATAGCGGAT 1149
S.gilo                CTGATATTATTGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTATAGCGGAT 1148
S.sisymbriifolium     CCGATATTATCGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTACAGCGGAT 1159
S.torvum              CCGATATTATCGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTACAGCGGAT 1163
S.khasianum           CTGATATTATTGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTATAGCGGAT 1166
S.clavatum            CTGATATTATTGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTATAGCGGAT 1132
S.nigrum              CTGATATTATTGACCGATTTGGGCGTATATGCAGAAATCTTTTTCATTATTATAGCGGAT 1146
                      * ******** ****************************************** *******

S.kurzii              CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGACTTTCTTGTGCTAGAA 1209
S.gilo                CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCTACTTTCTTGTGCTAGAA 1208
S.sisymbriifolium     CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGACTTTCTTGTGCCAGAA 1219
S.torvum              CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGACTTTCTTGTGCCAGAA 1223
S.khasianum           CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGACTTTCTTGTGCTAGAA 1226
S.clavatum            CTTCCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGAGTTTCTTGTGCTAGAA 1192
S.nigrum              CTTGCAAAAAAAAGACTTTATATCGAATAAAGTATATACTTCGACTTTCTTGTGCTAGAA 1206
                      *** ************************************ * ********** ****

S.kurzii              CTTTAGCTCGGAAACACAAAAGTGCTGTACGCACTTTTTTGAAAAGATCGGGCTCGGTTA 1269
S.gilo                CTTTAGCTCGGAAACACAAAAG-GCT-TACGCACTTTTTTGAAAAGATCGGACTCGG--A 1264
S.sisymbriifolium     CTTTAGCTCGGAAACACAAAAGTACTGTACGCACTTTTTTGAAAAGATCGGGCTCGG--A 1277
S.torvum              CTTTAGCTCGGAAACACAAAAGTACTGTACGCACTTTTTTGAAAAGATCGGGCTCGG--A 1281
S.khasianum           CTTTAGCTCGGAAACACAAAAGTACTGTACGCACTTTTTTGAAAAGATCAGGCTCAG--A 1284
S.clavatum            CTTTAGCTCGGAAACACAAATGTTCTGTACGTACTTTTC-GAAAAGATCA-ACTCTCACA 1250
S.nigrum              CTTTAGCTCGGAAACACAAAAGTACTGTACGCACTTTTTTGAAAAGATCGGGCTCGGTTA 1266
                      ******************* *  ** **** ******  *********   ***    *

S.kurzii              TTCATTGGTAGAGTCCTT 1287
S.gilo                ATTATTGGAAGAGTCCTT 1282
S.sisymbriifolium     ATTATTGGAAGAGTCCTT 1295
S.torvum              ATTATTGGAAGAGTCCTT 1299
S.khasianum           ATTATTGGAAGAGTCCTT 1302
S.clavatum            TGCACTTGCAGAGTCCTT 1268
S.nigrum              TGCACTTGCAGAGTCCTT 1284
                        * * * *********
```

APPENDIX 5

Rf	*S.khasianum*	*S.sisymbrifolium*	*S.kurzii*	*S.gilo*	*S.nigrum*	*S.clavatum*	*S.torvum*
0.034	+	-	+	+	+	-	+
0.171	+	+	+	-	+	+	+
0.197	+	+	+	+	+	-	+
0.231	+	+	+	+	+	+	-
0.256	-	+	-	-	+	-	-
0.299	+	+	+	+	+	-	+
0.325	-	+	+	+	-	-	+
0.342	+	+	+	-	-	+	-
0.368	+	+	+	+	-	-	-
0.393	+	+	+	+	-	-	+
0.419	+	-	+	+	+	+	+
0.444	+	+	+	-	-	+	-
0.47	+	+	+	-	-	+	-
0.487	-	-	+	+	-	+	-
0.513	+	+	+	+	-	+	+
0.53	+	-	+	+	-	+	+
0.599	-	+	-	-	-	-	+
0.624	+	+	-	-	-	-	-
0.641	-	+	-	-	-	-	-
0.658	-	-	+	+	+	-	+
0.684	-	-	+	-	-	+	-
0.726	+	-	-	-	-	+	+
0.744	-	-	+	-	-	+	-
0.786	+	-	+	-	-	+	+
0.812	+	+	-	+	-	+	+
0.889	-	+	+	+	+	+	-
0.915	+	+	-	+	+	+	-
0.966	-	-	+	+	-	+	+

Table: Polymorphic bands visible in total protein gel.

Curriculum Vitae

INSAN ARA RAHMAN

insanara.rahman13@gmail.com

PERMANENT ADDRESS

C/O Saiful Islam
Ganeshnagar, Bamunimaidan
Guwahati, Assam
India
Pin: 781021

TEMPORARY ADDRESS

New kenches Trace, Bishnupur
Shillong, Meghalaya
India
Pin: 793004

EDUCATION

2008-present	Ph.D, North Eastern Hill University
	Thesis: **Biochemical, molecular and phylogenetic studies of selected wild *Solanum* species of Meghalaya**
	Supervisor: Prof. Arvind K. Misra
2006-2007	M.Phil, Gauhati University
	Dissertation: **Studies on the cytotoxic effects of leaf extracts of *Aristolochia littoralis* on onion root tip cells**
	Supervisor: (Late) Prof. S. P. Borah
2005-2006	B. Ed, Gauhati University
2003-2005	M.Sc., Gauhati University
	Botany, Spl- Cytology, Genetics and Plant Breeding
2000-2003	B.Sc., Gauhati University
	Botany (Major)

RESEARCH INTEREST

Molecular Genetics

Molecular phylogeny

TECHNICAL SKILLS

Extraction of DNA from plant tissue, Handling Agarose Gel Electrophoresis, Polymerase Chain Reaction(PCR), PCR-RFLP, DNA Sequencing, Extraction of Total Protein, SDS-PAGE, Extraction of Secondary Metabolites, UV-VIS spectrophotometer.

BIOINFORMATICS SKILLS

Primer designing by using various software like DS-Gene, Chromous; DNA sequence analysis by BLAST, CLUSTAL W, CLUSTAL X; Phylogenetic tree construction by using MEGA5, PHYLIP, MR-BAYES etc; PCR-RFLP interpretation by NTSys etc.

CONFERENCE ATTENDED

2011 Paper presented entitled "**Molecular phylogeny of *Solanum* spp based on *18SrRNA* and *rbcL* genes**" in **'International Conference on New Horizons in Biotechnology'** (NHBT, 2011) in Trivandrum (Kerala) organized by BRSI and National institute of Interdisciplinary Science and Technology, CSIR.

2011 Presented paper entitled '**Phylogenetic studies of the genus *Solanum* based on Nuclear and Chloroplast gene'** in "**National Symposium on Biodiversity Assessment, Conservation and Value Addition**" organized by School of Life Sciences at NEHU.

2010 Presented paper entitled '**Molecular phylogeny of *Solanum* found in Meghalaya'** in the seminar on "**Biodiversity of North-East India- Issues and Concerns**" organized by Regional Centre, NAEB, NEHU.

WORKSHOP ATTENDED

2011 DBT-Sponsored training course on "**Bioinformatics: A Practical Approach in Genomics and Proteomics**" conducted by Bioinformatics centre, NEHU.

LANGUAGES

English, Hindi, Assamese