Evolutionary genetics and conservation of *Citrus* genetic resources in home gardens in Northeast India

Atiqur Rahman Barbhuiya

A thesis

in the

Department of Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy (Biology) at

Concordia University

Montreal, Québec, Canada

September 2015

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Atiqur Rahman Barbhuiya**

Entitled: **Evolutionary genetics and conservation of *Citrus* genetic resources in homegardens in Northeast India**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Brandon Findlay Chair

Dr. Edivani Franceschinelli External Examiner

Dr. Andreas Bergdahl External to Program

Dr. Patrick Gulick Examiner

Dr. Dylan Fraser Examiner

Dr. Selvadurai Dayanandan Thesis Supervisor

Approved by Dr. Selvadurai Dayanandan
Chair of Department or Graduate Program Director

September' 2015 _____
Dean of Faculty

ABSTRACT

Evolutionary genetics and conservation of *Citrus* genetic resources in homegardens in Northeast India

Atiqur R. Barbhuiya
Concordia University, 2015.

The genus *Citrus* L. is a major source of commercial fruits, cultivated in tropical and subtropical regions of the world. Many *Citrus* species and cultivars are commonly found in home gardens and play an important role in supporting the livelihood of local inhabitants in northeast India. This study includes the phylogenetic relationship among *Citrus* species, population genetics of a medicinally important and native *Citrus* species (*C. medica*) and plant diversity in the home gardens in northeast India. The phylogenetic relationships of 24 species of *Citrus* based on nucleotide sequences of three chloroplasts (trnL-trnF, trnS-trnG and rps16) and one nuclear (ITS2) DNA regions were inferred through three major phylogeny reconstruction methods. The analyses grouped morphologically distinct 24 *Citrus* species into five phylogenetically defined groups with presence of a true species (*C. medica, C. reticulata* and *C. grandis*) and their probable hybrids in three groups. Furthermore, this study revealed two additional groups with two wild, endemic and endangered species (*C. indica* and *C. assamensis)*. The species of acid and *Papeda* groups are polyphyletic.

The genetic diversity and structure of 219 *Citrus medica* individuals collected from 8 domestic and 4 wild populations were assessed using 5 polymorphic microsatellite markers. In total 67 alleles were detected with an average of 13.4 alleles per locus. The mean observed and expected heterozygosity values ranged between 0.220 - 0.540 and 0.438 - 0.733 respectively among the wild and domestic populations. Domestic populations showed close genetic relationships as compared to wild populations and pairwise Nei's genetic distance ranged from 0.062 to 2.091. Analysis of molecular variance (AMOVA) results showed higher genetic diversity among- than within-populations. The analysis of population structure revealed five groups, partly corresponding to geographical location of populations. The admixture of individuals among wild and domestic populations revealed their introgression in populations by

natural or farmer mediated agricultural practices. *Citrus medica* populations in the region are genetically diverse.

The eastern Himalayan region of northeast India is well known for its traditional home gardens, which play important role in the maintenance of livelihoods of indigenous communities and conservation of biological diversity. This study determined the plant diversity and their importance in conservation of plant genetic resources. This study was conducted in 90 home gardens located in 6 villages in two different districts in Mizoram and data collected through direct observations and thorough discussions with the farmers. The size of home gardens ranged between 0.10 – 0.60 ha and showed significant ($P<0.001$) positive correlation between the garden size and plant species diversity. A total of 333 plant species (133 trees, 92 shrubs and 108 herbs) belonging to 122 families with an average of 78 species per home garden were recorded. The species diversity indices for trees, shrubs and herbs were 4.76, 4.39 and 4.58 respectively. The species similarity within each life-form was high with 50% for trees, 38% for shrubs and 49% for herbs. Plant species in the home gardens could be grouped into 11 major use categories and majority of plants were of medicinal or multiple use categories. These home gardens are reservoirs of plant genetic resources and play a vital role in sustaining the livelihood of local inhabitants.

ACKNOWLEDGMENTS

I express my deep sense of gratitude and sincere appreciation to my supervisor Dr. Selvadurai Dayanandan for his valuable suggestions and guidance throughout the course of this study. I am also grateful for his tremendous support, encouragement and motivation.

I would also like to thank my committee members Dr. Patrick Gulick and Dr. Dylan Fraser for their valuable comments and suggestions during my research work and thesis preparation. I would like to thank Dr. M.L. Khan, Dr. H.S.G. Central University, Sagar, India for support and encouragement. I am especially thankful to Dr. U. K. Sahoo, Mizoram University, India for his support through providing an opportunity for me work in a CSIR, Govt. of India research project funded to him.

I would like to thank the School of Graduate Studies, Faculty of Arts and Science, and Department of Biology at Concordia University and Natural Sciences and Engineering Research Council of Canada for financial supports.

I thank all the former and current lab mates of Dayanandan's Lab for their kind cooperation and constructive criticism. Dr. Baharul Choudhury, thanks a lot for your help and encouragement. Thanks to Edith Pounden for her great help with editing.

I am thankful to the home gardeners of northeast India for providing *Citrus* species leaf materials. Special thanks to the home garden owners of Aizawl and Serchhip districts of Mizoram for permission to work in their gardens. The help received from the Principal Scientist, Krishi Vigyan Kendra, Citrus Research Station, Tinsukia, Assam in identifying species and providing leaf materials is gratefully acknowledged. Thanks are also due to the Dr. H. Lalramnghinglova, Dept. of Environmental Science, Mizoram University, Mizoram, India; Director and staffs of Botanical Survey of India, northeast Circle, Shillong, Meghalaya, India for help in species identification and permission to consult the herbarium.

I am deeply indebted to my parents, relatives and well wishers who encouraged me throughout the course of this research. I am thankful to my wife Zian Imdad for her support and encouragement. Thanks to my loving daughters Aneesha and Ambereen for always cheering me up and understanding.

Last but not the least I thank the Almighty, who blessed me with will power and strength to complete the present research.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF ABBREVIATIONS

A- Adenine

AFLP- Amplified Fragment Length Polymorphism

AIC- Akaike Information Criterion

AMOVA- Analysis of Molecular Variance

ANOVA- Analyses of Variance

BI- Bayesian Inference

BIC- Bayesian Information Criterion

BLAST- Basic Local Alignment Search Tool

bp- Base pair

BS- Bootstrap

C- Cytosine

CI- Consistency Index

cpDNA- Chloroplast DNA

CTAB- Cetyl Trimethyl Ammonium Bromide

DMSO- Dimethyl sulfoxide

DNA- Deoxyribonucleic acid

E- East

EDTA- Ethylenediaminetetraacetic acid

G- Guanine

GTR- General Time Reversible

ha- Hectare

HI- Homoplasy Index

HWE- Hardy Weinberg Equilibrium

ILD- Incongruence Length Difference

IRD- Infrared Fluorescent Dyes

ISSR- Inter Simple Sequence Repeat

ITS- Internal Transcribed Spacer

IUCN- International Union for Conservation of Nature

IUPAC- International Union of Pure and Applied Chemistry

IVI- Importance Value Index

km- Kilometer

LD- Linkage Disequilibrium

m- Meter

MCMCMC- Metropolis Coupled Markov Chain Monte Carlo

min- Minute

ML- Maximum Likelihood

mM- Milimolar

mm- Milimeter

MP- Maximum Parsimony

N- North

NCBI- National Center for Biotechnology Information

ng- Nanogram

nrDNA- Nuclear DNA

PAUP- Phylogenetic Analysis Using Parsimony

PCR- Polymerase Chain Reaction

PIC- Polymorphic Information Content

pmol- Picomole

PP- Posterior Probability

RAPD- Randomly Amplified Polymorphic DNA

RFLP- Restriction Fragment Length Polymorphism

RI- Retention Index

RNA- Ribonucleic Acid

rps- Ribosomal Protein Subunit

SCAR- Sequence Characterized Amplified Regions

SE - Standard Error

sec- Second

SPSS- Statistical Package for the Social Sciences

SSR- Simple Sequence Repeats

TBE- Tris-Borate-EDTA Buffer

TBR- Tree Bisection and Reconnection

TL- Tree Length

trn- Transfer RNA

T- Tyrosine

UPGMA- Unweighted Pair-Group Method with Aithmetic Averages

µl- Microliter

GLOSSARY

Abundance - Abundance is defined as the relative representation of a species in a particular ecosystem. The number of organisms in a population, combining intensity (density within inhabited areas) and prevalence (number and size of inhabited areas).

Accessions - A collection of plant material from a particular location and time, a new member to a plant collection process.

Adaptation - Inherited characteristic of an organism that enhances its survival and reproduction in a specific environment.

Admixture - The formation of novel genetic combinations through hybridization of genetically distinct groups.

Allele - Any of the alternative versions of a gene that may produce distinguishable phenotypic effects.

Allelic Diversity - A measure of genetic diversity based on the average number of alleles per locus present in a population.

Allelic Richness (A_R) - A measure of the number of alleles per locus; allows comparison between samples of different sizes by using various statistical techniques.

Analyses of Variance (ANOVA) - ANOVA is a collection of statistical methods used to analyze the differences between group means and variance among and between groups.

Analysis of Molecular Variance (AMOVA) - AMOVA is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. It is used to summarize the population structure with the marker data from different genotypes.

Annual Plant - A plant species with a life cycle of approximately 12 months or rather less to complete, whose life cycle is therefore directly related to the annual cycle of weather, and whose generations are therefore discrete.

Apomixis - Apomixis produces progeny that are an exact genetic replica of their mother plant that results from changes in the female reproductive pathway such that female gametes develop without meiosis and embryos develop without fertilization.

Biennial Plant - Applied to a plant that lives for two years. During the first season food may be stored for use during flower and seed production in the second year.

Biodiversity Hotspot- A relatively small area with numerous endemic species and a large number of endangered and threatened species.

Biological Species - A species as a group of populations whose members have the potential to interbreed in nature and produce viable, fertile offspring, but do not produce viable, fertile offspring with members of other groups.

Bootstrapping - Bootstrapping is a resampling method that measures the accuracy of samples by using random resampling methods.

Bottleneck- A population bottleneck is a sharp reduction in the size of a population due to environmental stochastic events (such as earthquakes, floods, fires, or droughts) or human activities. Such events can reduce the variation in the gene pool of a population drastically and lead to population extinction.

Chromatograms - A chromatogram is the visual representation of a DNA nucleotides detected by a sequencing machine. In a chromatogram file, the signal intensities are presented in a graph with the four nucleotides bases (Adenine, Guanine, Cytosine and Thymine) and each identified by different colors.

Clades - A group of species that includes an ancestral species and all of its descendants.

Cladistics - An approach to systematics in which organisms are placed into groups called clades based primarily on common descent relationships.

Clone - An individual that is genetically identical to another individual.

Cluster - Grouping of taxa in phylogenetic tree based on similarities of characters.

Codominance - The situation in which the phenotypes of both alleles are exhibited in the heterozygote because both alleles affect the phenotype in separate, distinguishable ways.

Codon - A three-nucleotide sequence of DNA or mRNA that specifies a particular amino acid or termination signal; the basic unit of the genetic code.

Community - The organisms that inhabit a particular area; an assemblage of populations of different species living close enough together for potential interaction.

Congruence - Congruence is the state of agreement. This is broadly applied in evolutionary biology to justify multigene phylogeny or phylogenomics. Congruence also applies in studies of coevolution, lateral gene transfer, and as evidence for common descent.

Crossing Over - Crossing over is the process of exchanging genetic materials during meiosis.

Cultivars - A distinct breed or subset of a species that behave uniformly and predictably when grown in an environment to which it is adapted. Also known as variety or release.

Density - The number of individuals per unit area or volume.

Diploid - The condition of having two sets of chromosomes (and therefore two sets of the genes carried on them) is called diploidy.

Disturbance - A natural or human-caused event that changes a biological community and usually removes organisms from it.

Diversity - An index of community diversity through the measurement of species richness and the relative abundance of species.

Diversity Index - A mathematical index of species diversity in a community.

Domestic Populations - Domestication is the process of adapting wild plants or animals for human use. Domestic species are raised for food, work, clothing, medicine, and many other uses and a number of species form domestic populations. Domesticated populations are raised and cared by humans.

Dominance - The species having the most influence on community composition and form. The largest and most abundant species in the community.

Dominant Species - Species which make up a large proportion of community biomass or numbers.

Effective Number of Alleles (Ne) - It is the number of alleles that can be present in a population. The measure tells about the number of alleles that would be expected in a locus in each population.

Effective Population Size - The number of individuals in a population that can actively contribute to the gene pool of the next generation.

Electrophoresis - Polarized acetate, agarose or acrylamide gel through which one runs proteins or DNA materials. The material then separated by weight or polarity depending upon their molecular weight.

Endangered Species - A technical definition used for classification referring to a species that is in danger of extinction throughout all or a significant portion of its range. IUCN defines species as endangered if the factors causing their vulnerability or decline continue to operate.

Endemic Species - Referring to a species that is confined to a specific geographic area.

Evolution - Descent with modification; the present species are descendants of ancestral species and may be different in morphology and genetic makeup; also defined more narrowly as the change in the genetic composition of a population from generation to generation.

Exon - A sequence within a primary transcript that remains in the RNA after RNA processing; also refers to the region of DNA from which this sequence was transcribed.

***Ex-situ* Conservation** - A conservation method that entails the removal of germplasm resources (seed, pollen, sperm, individual, organisms etc), from their original habitat or natural environment and growing and maintaining them outside of their original habitats.

Family - In Linnaean classification, the taxonomic category above Genus.

Frequency - Species frequency is the number of times a plant species is present in a given number of quadrats of a particular size or at a given number of sample points.

Gene - A discrete unit of hereditary information consisting of a specific nucleotide sequence in DNA (or RNA, in some viruses).

Gene Flow (Nm) - The transfer of alleles from one population to another, resulting from the movement of fertile individuals or their gametes.

Gene Pool - The aggregate of all copies of every type of allele at all loci in every individual in a population.

Genetic Differentiation (F_{ST}) - F_{ST} is a measure of genetic divergence among subpopulations. This is also known as fixation index, which is the proportion of the total genetic variation that is due to genetic differentiation among local populations. Fixation index is the proportional increase of homozygosity through population subdivisions.

Genetic Diversity - Genetic diversity refers to the variation at the level of individual genes (polymorphism), and provides a mechanism for populations to adapt to their ever changing environment. The more variation the better the chance that at least some of the individuals will have an allelic variant suited for the new environment. Further, the offspring with the variant will reproduce and continue the population into subsequent generations.

Genetic Drift - Drift is one of the major forces of evolutionary change that reduces heterozygosity by the random loss of alleles.

Genotype - The genetic makeup, or set of alleles, of an organism.

Genus - A taxonomic category above the species level, designated by the first word of a species in binomial nomenclature system.

Germplasm - The genetic material that forms the physical basis of heredity and is transmitted from one generation to the next by means of the germ cells. Often synonymous with genetic material. When applied to plants it is the name given to seed or other material from which plants are propagated.

Haploid - Organisms having only one set of chromosomes.

Hardy Weinberg Principle – This states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences.

Herb - A small, non woody, seed bearing plants in which all the aerial parts die back at the end of each growing season.

Herbarium - A herbarium (plural- herbaria) is a collection of preserved plant specimens with its identity and other primary information. These specimens may be whole plants or plant parts, usually in dried form mounted on a sheet or may also be kept in alcohol or other preservatives. They are often used as reference material in identification and describing taxa.

Heterozygosity - A measure of genetic variation that estimates either the observed or expected proportion of individuals in a population that are heterozygotes.

Heterozygous - Having two different alleles for a given gene.

Homoplasy - A similar (analogous) structure or molecular sequence that has evolved independently in two species but not present in their common ancestor.

Homozygous - Having two identical alleles for a given gene.

Hybridization - The interbreeding of distinct species.

Inbreeding Depression - The reduced fitness of species or populations due to increased homozygosity (therefore expression of recessive deleterious alleles) from inbreeding.

Incomplete Dominance - The situation in which the phenotype of heterozygotes is intermediate between the phenotypes of individuals homozygous for either allele.

Ingroup - In a cladistic analysis, the set of taxa which are hypothesized to be more closely related to each other than any are to the outgroup.

***In-situ* Conservation** - A conservation method that attempts to preserve the genetic resources in their original habitat or natural environment.

Introgression - Movement of genes (or traits) between species or between well-differentiated populations.

Intron - A noncoding, intervening sequence within a primary transcript that is removed from the transcript during RNA processing; also refers to the region of DNA from which this sequence was transcribed.

Isozymes - Isozymes also known as isoenzymes or more generally as multiple forms of enzymes that differ in amino acid sequence but catalyze the same chemical reaction.

ISSR/ SSR - Inter Simple Sequence Repeat is a type of microsatellite marker, are highly repeating sequences of 2-5 base pairs of DNA. They are highly polymorphic and used in populations, phylogenetic and other related studies. SSR are also called VNTRs (variable number of tandem repeats) and consist of tandem repeats units. More than one microsatellite locus can be PCR amplified from a single tube and then identified separately on a sequencing gel using different colors of fluorescent dyes for each locus. Inter-simple sequence repeat (ISSR) markers generate a large number of DNA fragments from a single PCR. ISSR primers are based upon the simple sequence repeats found in microsatellites. Bands are generated by a single-primer PCR reaction, where the primer is a repetition of a di-, tri- or tetranucleotide and the amplified region is a portion of genome between two identical microsatellite primers.

Landraces - A landrace is a domesticated, regional ecotype; a locally adapted, traditional variety of a domesticated species that has developed over time, through adaptation to its natural and cultural environment of agriculture.

Linkage - An association in inheritance between traits, such that the parental trait combinations appear among the progeny more often than the non-parental. The proximity of two or more genetic markers on a chromosome; the closer together the markers are, the lower the probability that they will be separated during DNA repair or replication processes and hence the greater the probability that they will be inherited together.

Linkage Disequilibrium (LD) – LD is the non-random association of alleles at two or more loci that may or may not be on the same chromosome.

Linkage Equilibrium - Populations where combinations of alleles or genotypes can be found in the expected proportions are said to be in linkage equilibrium.

Loci /Locus - A specific place along the length of a chromosome where a given gene is located.

Markov Chain - A mathematical system that undergoes transitions from one state to another, as a random process in which the next state depends only on the current state.

Markov Chain Monte Carlo (MCMC) - A tool or algorithm for sampling from probability distributions based on constructing a Markov chain. The state of the chain after many steps is then used as a sample from the desired distribution.

Mean Number of Alleles (MNA) - The MNA is the average numbers of alleles observed in a population and are obtained by direct counting.

Monophyletic - Pertaining to a group of taxa that consists of a common ancestor and all of its descendants.

Nei's Genetic Distance (D_S) - Genetic distance is a measure of the genetic divergence between species or between populations within a species. This states that if the rate of genetic change (amino acid substitution) is constant per year or generation then Nei's standard genetic distance increases in proportion to divergence time.

Nei's Unbiased Genetic Distance (D_A) - This distance assumes that genetic differences arise due to mutation and genetic drift.

Outgroup - Taxon phylogenetically outside the clade of interest (the ingroup). It is used in phylogenetic inference for determining polarity (direction of character change/whether a character is or isn't ancestral).

Paraphyletic Group - Artificial assemblage of taxa that includes a common ancestor and some but not all of its descendants.

Parsimony - The principle that the preferred phylogeny of an organism is the one that requires the fewest evolutionary changes; the simplest explanation.

Perennial Plant - A plant that normally lives for more than two seasons and which after an initial period, produces flowers annually.

Phylogeny - The evolutionary history of a species or group of related species.

Phylum - In Linnaean classification, the taxonomic category above class.

Polymorphic Information Content (PIC) - The polymorphic information content is a measure of polymorphism for a marker locus used in linkage analysis.

Polymorphism- The presence of more than one allele at a locus. The existence within a species or population of different forms of individuals, beyond those that are the result simply of recurrent mutation.

Polyphyletic Group - Artificial assemblage of taxa derived from two or more common ancestors.

Polyploidy - Polyploidy refers to numerical change in the whole set of chromosomes; a chromosomal alteration in which the organism possesses more than two complete chromosome sets.

Populations - A group of individuals of the same species that live in the same area and interbreed, producing fertile offspring.

Primer - A small oligonucleotide (typically 18–22 base pairs long) that anneals to a specific single stranded DNA sequence to serve as a starting point for DNA replication.

Private Allele - An allele present in only one of many populations sampled.

Quadrat - A basic sampling unit of vegetation surveys.

RAPD - Random Amplification of Polymorphic DNA- a type of PCR reaction where segments of DNA amplified randomly. This method does not require any specific knowledge of the DNA sequence of the target organism, and hence popular for comparing DNA of biological systems.

Rare Species - A rare species is one that is at risk because of its small population size and usually confined to small geographic areas or habitats, or scattered thinly over a larger area.

RFLP - A single nucleotide polymorphism (SNP) that exists in the restriction site for a particular enzyme, thus making the site unrecognizable by that enzyme and changing the lengths of the restriction fragments formed by digestion with that enzyme.

SCAR - Sequence Characterized Amplified Regions- a type of molecular marker developed with a pair of longer primers usually the extended sequence of a RAPD primer that has a specific sequence of approximately 20 bases. Compared with universal primers they are very specific and more reproducible.

Shrub - A woody plant which branches below or near ground level into several main stems, so has no clear trunk.

Sister Group - The two clades resulting from the splitting of a single lineage.

Species - A group of organisms with a high degree of physical and genetic similarity, that naturally interbreed among themselves and can be differentiated from members of related groups of organisms.

Species Diversity - The number and relative abundance of species in a biological community.

Species Richness - The number of species in a biological community.

Subspecies - Subdivisions of a species, with clear morphological distinctions and/or limited interbreeding between them.

Sustainability - Where an activity can be continued or repeated for the foreseeable future.

Sympatric and Allopatric Populations - Speciation is the formation of new species and is also referred to as macroevolution, an increase in biodiversity, or as taxonomic multiplication. Sympatric populations are those where many varieties in one range becomes species through adaptation to different aspects of the range. However, in allopatric each variety in its own range becomes species due to drift and local adaptation.

Taxon - A named taxonomic unit at any given level of classification.

Taxonomy - The study of the rules, principles and practice of classifying living organisms.

Threatened Species - A technical classification referring to a species that is likely to become endangered within the foreseeable future, throughout all or a significant portion of its range.

Topology - The shape of phylogenetic trees. Two trees have the same topology if rotating branches shows that the patterns of relationships among the operational taxonomic units are identical.

Trait - An attribute or character of an individual within a species for which heritable differences can be defined.

Tree - A woody plant with a single main stem (the trunk) that is unbranched near the ground, some trees have multi-trunked forms.

Unweighted Pair-Group Method with Arithmetic Averages (UPGMA) - A method of arithmetic averages, tree-building technique for phylogenetic analysis. Data required are distances (genetic distance or other distance measure) between taxa, arranged in a matrix form. This method constructs a tree by identifying the shortest distance in the matrix, clustering those two taxa into a single operational taxonomic unit for use in all subsequent calculations, and then repeating these steps.

Variation - Differences between members of the same species.

Varieties - In botanical nomenclature, variety is a taxonomic rank below that of species and subspecies. A plant variety is a plant grouping within a single botanical taxon of the lowest known rank, which can be defined by the expression of the characteristics resulting from a given genotype or a combination of genotypes distinguished from any other plant grouping, by the expression of at least one of the said characteristics, and considered as units with regard to its suitability for being propagated without change.

Wild Species/Populations - Organisms captive or living in the wild that have not been subject to breeding to alter them from their native state. Occurring, growing or living in a state of nature without cultivation or the care of human and existed in any area for many years.

This glossary is based on the following sources:

Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, Robert B. Jackson 2011. Campbell Biology, 9th edition. Pearson Education, Inc., USA.

Fred W. Allendorf, Gordon Luikart and Sally N. Aitken. 2013. Conservation and the Genetics of Populations, 2nd Edition, John Wiley & Sons, Ltd, UK.

Lawrence B. 2011. Henderson's Dictionary of Biology, 15th Edition, Pearson Education Limited, England, UK.

http://www.biology-online.org/dictionary

GENERAL INTRODUCTION

The loss of biological diversity due to habitat destruction and accompanied loss of wild relatives and germplasm of commercially important domesticated crop plants remains as one of the greatest concerns of our time. The wild and semi domesticated plants may harbor genetic diversity that could be harnessed to improve domesticated plants to sustain food and other material production to meet increasing global demands under changing climatic conditions. Thus, programs targeted to conserve wild and semi-domesticated plants with commercial interests are urgently needed, and the effective conservation of genetic resources will largely depend on the detailed understanding of the target plant group. In particular, information on evolutionary relationships among species (phylogeny) and genetic structure of domesticated and semi-domesticated populations are needed to develop germplasm conservation programs. The plant domestication is a result of the selection of phenotypes of plants by human adapting to various agro-ecological niches. This process has led to selection of specific genotypes desirable phenotypes (Abbo et al. 2014, Larson et al. 2014). Vavilov (1926) and Engelbrecht (Zeven 1973) suggested that the diverse phenotypic variations found in the domesticated plants are likely to have arisen through natural selection in response to abiotic and biotic factors encountered during domestication. Domesticated plants differ from their wild progenitors in numerous characteristics or traits, the rates of phenotypic evolution between wild and domestic species are not similar (Fuller et al. 2014), and the intensity of natural selection on specific traits varies between wild and domestic species (Purugganan and Fuller 2011).

Phylogenetic studies

Robust phylogenetic trees of chosen plant groups as a foundation onto which life history traits as well as morphological and ecological data can be superimposed to elucidate evolutionary patterns are needed for better understanding of the evolutionary ecology of the group and systematic classification of taxa into hierarchical groupings in a phylogenetic context. Delimiting species is important in understanding the historical and ongoing evolutionary mechanisms and processes (Sites and Marshall 2003). Species can be recognized through their differences in morphology and known as the morphological species concept, which may not explain the true biological distinctiveness of the species (Mayr 1996). The controversies and weakness of the

morphological species concept led to the species delimitation based on 'Biological Species Concept' (BSC) defined as "interbreeding natural populations that are reproductively isolated from other such groups", which explains species as cohesive units of genomes and are separated from each other by reproductive barriers (Mayr 1942). Although BSC is widely accepted, its limitations have been pointed in cases of asexual reproduction (Templeton 1989) and hybridization (Whittemore 1993). The development of cladistic methods to recognize monophyletic grouping of taxa (Mallet 2007) leading to 'phylogenetic species concept' (Hennig 1966) gained popularity in delimiting species. A phylogenetic based species is considered as an irreducible (basal) cluster of organisms distinct from other such clusters, and within which there is a parental pattern of ancestry and descent (Cracraft 1989). Nixon and Wheeler (1990) defined phylogenetic species as 'the smallest aggregation of populations (sexual) or lineages (asexual) diagnosable by a unique combination of character states in comparable individuals'. In general, phylogenetic trees are needed to delimit species and phylogenetic trees based on molecular sequences have gained popularity in recognizing monophyletic groups for delimiting species boundaries.

Population genetics

The amount and distribution of genetic variation or the genetic structuring of populations play a crucial role in the adaptability to the environmental changes and long-term survival of populations. The genetic structure of populations is a result of interacting genetic processes of selection, genetic drift and gene flow. Through natural selection, individuals tend to adapt to their local environment and genetic transfer of such variation from parents to the offsprings and finally increasing fitness and survival in the changing environment (Endler 1986). Divergent selection processes for different traits in the wild or semi-domesticated crop plant populations is followed by selection of different genotypes. In contrast, the genetic drift, one of the major forces of evolutionary change, affect populations through random loss of alleles. Infinitely large populations generally may not be affected by genetic drift, whereas small populations may experience major changes through genetic drift. Genetic drift causes loss of genetic variation from generation to generation through random changes in allele frequencies. This further affected through founder effects, where severe reduction in populations size referred to as population bottleneck. The effects can vary depending on both the size to which the population is

reduced and the duration of the bottleneck (the number of generations). Limited population size can lead to a loss of genetic variation and subsequent loss of evolutionary potential of populations (Allendorf et al. 2013). Gene flow, the successful movement of genes among plant populations is an evolutionary force that counteracts the affects of selection and genetic drift (Slatkin 1985, Mayr 1963). Gene flow in plants can be accomplished by cross-fertilization or by the dispersal of whole plants, plant fragments, seeds, and spores (Ellstrand 2003) and this can be a primary source of genetic variation in any population (Mayr 1963). Gene flow plays an important role in spatial genetic structuring of populations. Therefore, the measurement of genetic differentiation among existing populations can serve as a good indicator of the gene flow (Selkoe and Toonen 2006). The existing population genetic structure represents the effects of evolutionary forces over generations and gene flow inferred from population genetic structure thus provides historic information at various levels from species through individuals and populations (Ellstrand 2003, 2014).

Genetic diversity defined as the variation at the level of individual genes, plays a crucial role in adaptation under changing environments. High genetic variations provide better chance for producing genetically variable offspring in subsequent generations. Determination of mean number of alleles (MNA) and heterozygosities at the individual and populations levels serve as good indicators of the genetic diversity. The MNA is the average number of alleles observed in a population. Expected heterozygosity is the probability that an individual will be heterozygous at a locus and the observed heterozygosity is the frequency of heterozygous individual per locus. Expected heterozygosity is calculated as $He = 1 - \sum_{i=1}^{n} pi^2$; where n is the number of allele and where *p*i is the frequency of the *i*th allele. Overall gene diversity is the proportion of polymorphic loci across the genome under Hardy-Weinberg equilibrium (HWE), i.e., the total number of genetic characteristics in the genetic makeup of a species and the expected heterozygosity measures is refered as gene diversity (Nei 1987). Allelic richness (A_R) is the total number of alleles present in populations at different locus. Effective number of allele is the number of equally frequent alleles that would create the same heterozygosity as observed in the population and can be calculated as $Ne = 1/\sum pi^2$; where *pi* is the frequency of the *i*th allele. The Hardy-Weinberg principle states that allele and genotype frequencies in a random and large population will remain constant from generation to generation in the absence of other

evolutionary influences. The factors that affect HWE include non-random mating, mutation, migration or gene flow, selection and random genetic drift. In any populations study, it is important to determine whether the loci and the populations genotyped are in HWE and whether there are any significant deviations from the HWE.

Genetic distance is a measurement of genetic divergence between species or between populations within a species (Nei 1987). This difference between two populations provides a good estimate of their divergence (Avise 1994). The most commonly used genetic distance measurement is Nei's standard genetic distance (Nei 1972) and is defined as $D = -\ln [G_{XY}/\sqrt{G_X G_Y}]$; where G_X, G_Y and G_{XY} are the means of $\sum pi^2$, $\sum qi^2$ and $\sum piqi$ respectively. *pi* and *qi* being the frequencies of the *i*th allele in populations X and Y respectively, and X*i* and *Yi* be the corresponding sample allele frequencies (Nei 1978).

The individuals in populations are subdivided or structured and genetic variation is partitioned within and between local populations. Studies about the genetic structure of a population or differentiation between populations are important in determining the number of alleles exchanged between populations. The commonly used metrics of genetic differentiation are F-statistics (Wright 1978) that describe the distribution of genetic variation within a species through the measurement of different inbreeding coefficients such as F_{IS}, F_{ST}, and F_{IT}. F_{IS} is a measure of departure from Hardy-Weinberg proportions within the local subpopulations. F_{ST} is a measure of allele frequency divergence among subpopulations and F_{IT} is a measure of the overall departure from HW proportions in the overall population. F-statistics are a measure of the deficit of heterozygotes relative to expected HW proportions in the population and can be calculated as F = 1– (Ho/He); Ho and He are the observed and expected heterozygotes. F_{IS} is a measure of departure from HW proportions within local subpopulations and can be expressed as $F_{IS} = 1– (Ho/H_S)$; where Ho is the observed heterozygosity of all subpopulations, and H_S is the expected heterozygosity averaged over all subpopulations. F_{ST} is the measure of genetic divergence among subpopulations and can be calculated as $F_{ST} = 1– (H_S/H_T)$; where H_T is the expected heterozygosity of the allele frequencies averaged over all subpopulations (Allendorf et al. 2013).

The study system

The genus *Citrus* L. of the family Rutaceae is a major source of commercially important fruits, which includes orange, lemon and lime cultivated in tropical and subtropical regions throughout

the world. Although a broad area covering northeast India, China, Japan and Australia is generally considered as the centre of origin of *Citrus* species (Tanaka 1954, Swingle and Reece 1967, Scora 1975, Mabberley 2004), the occurrence of large number of *Citrus* species in natural forests and home gardens under semi domesticated condition in northeast India suggests this region as centre of origin of *Citrus* (Scora 1975) and may contain high level of genetic diversity. Bhattacharya and Dutta (1956) reported 17 *Citrus* species, 52 cultivars and 7 probable natural hybrids from the region. Many *Citrus* species and cultivars are commonly found in home gardens under semi-domesticated conditions and play an important role in supporting the livelihood of local inhabitants. Home gardeners of northeast India maintained *Citrus* species for generations because of their utilitarian value, which resulted in accumulation of large number of species in their home gardens. The taxonomy and phylogeny of *Citrus* remain poorly understood due to sexual compatibility between *Citrus* and related genera leading to intra- and inter-generic hybridization, polyploidy, somatic mutations (Araujo et al. 2003, Mabberley 2004), long history of cultivation in extensive geographic regions. Over the years, numerous classification systems have been formulated, however, controversies still exist in defining species and varieties of *Citrus*. Therefore, the first objective of the present study is to reconstruct the phylogeny of *Citrus* species in northeast India using chloroplast and nuclear DNA markers.

The second objective of the study is to determine the levels of genetic diversity in wild and domesticated populations of *C. medica* L.*,* one of the medicinally important native *Citrus* species for assessing the genetic structure in natural and domesticated population to gain insights into genetic impacts of domestication. Several studies and botanical explorations (Hooker 1875, Bhattacharya and Dutta 1956, Tanaka 1977, Nair and Nayar 1997) reported many wild populations in primary and secondary forests in the foothills of eastern Himalayas in northeast India. However, these populations of citron (*C. medica*) have declined in recent years due to natural and anthropogenic disturbances. Thus, conservation measures are essential to prevent further decline of citron genetic resources, and information on the genetic structure and diversity is crucial for formulating conservation and management strategies. A few previous studies through different molecular methods and markers reported lower heterogeneity among the citron accessions as compared to the other *Citrus* species. Those studies were based on limited number of accessions and studies with populations of natural habitats are unknown. The present study

covers extensive sampling of citron from northeast India represents the first study to assess the genetic variability of *C. medica* in its natural habitat using microsatellite markers.

A large proportion of human population in northeast India are native tribal communities living in the mountains and practice indigenous agricultural practices and their livelihoods are dependent on natural systems. Several tribal communities in this region maintain sustainable livelihoods through adopting integrated farming systems (Liu et al. 2007). Mizos of Mizoram is one such highland tribal community that practices home gardens as sustainable subsistence agriculture. Many species of *Citrus* are commonly cultivated in home gardens, which are considered to play a significant role in conservation of *Citrus* genetic resources. The home gardening system in the region believed to have evolved from slash and burn agriculture locally known as 'jhum'. The 'jhum' is a labour intensive cultivation system that requires minimal capital and nutrient input and often practiced at the village outskirts through slashing and burning the forest. Upon realization of adverse impacts of 'jhum' agriculture, many farmers in the region shifted to home gardening system for the maintenance of crop diversity, household food security, nutrition and subsistence income generation. Since most of the landscapes in the region are steep slopes, such system of land-use is a suitable approach to minimize soil erosion and easily adaptable for ecological rehabilitation and agricultural productivity (Sahoo 2007). Along with the indigenous and local varieties of crops, farmers of the region cultivate large number of improved varieties of annual/biennial crops. These gardens are often enriched by wild germplasm from nearby forests. This complex farming system is dynamic and includes various life forms of plants ranging from herbs, shrubs, trees through climbers. Despite their biological richness and importance, the species composition in these systems remain poorly understood. Thus, the third objective of my thesis is to determine the plant diversity in home gardens to assess the importance of home gardens in conservation of biodiversity, including *Citrus* genetic resources in the region.

Chapter 1: Molecular phylogeny of *Citrus* species in the Eastern Himalayan region based on chloroplast and nuclear DNA sequence data

Abstract

The genus *Citrus* L. (Rutaceae) is a major source of commercial fruits, including orange, lemon and lime cultivated in tropical and subtropical regions of the world. A large number of *Citrus* species are found in the Assam-Burma area of the Indo-Burma Biodiversity Hotspot, suggesting this area as a centre of origin of *Citrus* and a source of rich genetic resources of *Citrus*. The phylogeny of the genus *Citrus* remains poorly known due to its high morphological diversity distributed across a broad geographical range, natural hybridization and a long history of human-mediated selection. The phylogenetic relationships representing 24 species of *Citrus* were reconstructed based on nucleotide sequences of three chloroplasts (trnL-trnF, trnS-trnG and rps16) and one nuclear (ITS2) DNA regions. The phylogenetic relationships were inferred through three major phylogeny reconstructions approaches, namely Maximum parsimony, Maximum likelihood and Bayesian inferences. The analyses grouped morphologically distinct 24 *Citrus* species into five phylogenetically defined groups with presence of a true species (*C. medica, C. reticulata* and *C. grandis*) and their probable hybrids in three groups and two additional groups with two wild, endemic and endangered species (*C. indica* and *C. assamensis)*. The species of acid and *Papeda* groups are polyphyletic.

Keywords: *Citrus*, chloroplast and nuclear DNA, northeast India, phylogeny.

Introduction

The genus *Citrus* L. of the subfamily Aurantioideae (Rutaceae) is a major source of commercial fruits, which include orange, lemon and lime cultivated in tropical and subtropical regions of the world. The taxonomy and phylogeny of the genus *Citrus* remain largely ambiguous due to its morphological diversity across a wide geographical range, natural hybridization and a long history of human-mediated selection. The region of northeast India, China, Japan and Australia is considered to be the centre of origin of *Citrus* species (Tanaka 1958, Swingle and Reece 1967, Scora 1975, Mabberley 2004). A large number of *Citrus* species are found in the Assam-Burma area of the Indo-Burma Biodiversity Hotspot, suggesting this area as a centre of origin of *Citrus*

(Scora 1975). Bhattacharya and Dutta (1956) reported 17 *Citrus* species, 52 cultivars and 7 probable natural hybrids from the region as well as occurrence of numerous *Citrus* populations in natural forests, further supporting this region as a centre of origin of *Citrus*. A botanical field exploration by Sharma et al. (2004) reported 23 species, one subspecies and 68 varieties of *Citrus* in the region. A thorough understanding of evolutionary relationships among these species and cultivars is needed for systematic classification and improved understanding of the evolution of *Citrus* species in the region. Although a large number of *Citrus* species and varieties are found in nature, only limited numbers of species have been commercialized as a source of fruits. Besides their food values, many species are also used for their medicinal properties. Various natural and anthropogenic disturbances in the region lead to the reduction of wild populations and many species are currently under threats. Therefore, *Citrus* genetic resource assessment is crucial for sustainability and proper conservation strategies.

In general both wild and domesticated species of *Citrus* are diploid (2n=18) and a limited number of species (*C. aurantifolia*, *C. medica*, *C. paradisi*) reported to have polyploidy either spontaneously or through crossing (Krug 1943). Recent morphological and cytological studies by Hynniewta et al. (2011, 2014) confirmed that *Citrus* species of northeast India do not have variations in their chromosome numbers and all species have n=9 chromosomes. The number of *Citrus* species recognized based on morphological traits ranges from three or four (Linnaeus 1753, Hooker 1875) through 145 to 162 (Tanaka 1954, 1969, 1977). Two of the commonly used taxonomical treatments by Swingle and Reece (1967) and Tanaka (1977) recognize 16 and 162 species respectively. In a comprehensive phylogenetic study employing 146 morphological and biochemical characters, Barrett and Rhodes (1976) recognized only three true species within cultivated *Citrus*, namely *C. medica* L. (citron), *C. reticulata* Blanco (mandarin) and *C. grandis* (L.) Osbeck (pomelo). The classification of all *Citrus* taxa into only three species has been further supported by taxonomic studies (Scora 1975) and DNA marker based studies (Fang et al. 1998, Federici et al. 1998, Nicolosi et al. 2000, Bayer et al. 2009, Jena et al. 2009, Zhen-Hua et al. 2011, Ollitrault et al. 2012, Garcia-Lor et al. 2013, Penjor et al. 2013).

Evolutionary genetic studies based on isozymes (Herrero et al. 1996), RAPD and PCR-RFLP (Federici et al. 1998, Asadi Abkenar et al. 2004), RAPD, SCAR and PCR-RFLP (Nicolosi et al. 2000), AFLP (Liang et al. 2007, Pang et al. 2007), SSR (Barkley et al. 2006), ISSR (Shahsavar et al. 2007) and analysis of non-coding chloroplast DNA sequences (Chase et al.

1999, Araujo et al. 2003, Morton et al. 2003, Bayer et al. 2009, Jena et al. 2009, Zhen-Hua et al. 2011) have shed light on the evolution of *Citrus* species in the region. Different studies using cpDNA (trnL-trnF, psbH-petB, trnS-trnG, matK) and nrDNA (Zhen-Hua et al. 2011, Kumar et al. 2013, Penjor et al. 2013) supported a monophyletic origin and the three-species concept and diversification of the genus. The aim of this study is to delimit the species boundaries and to determine whether the morphologically diverse *Citrus* species are true biological or genetically distinct species from each other. The specific objectives of the present study are (i) to reconstruct the phylogeny of *Citrus* species in northeast India and (ii) to assess the levels of congruence between the derived phylogeny and existing classification systems. The resulting information is crucial for germplasm characterization and to develop conservation strategies for *Citrus* species in the region.

Materials and methods

Taxon sampling

Leaf samples of *Citrus* representing 24 species were collected from Assam, Meghalaya, Mizoram and Arunachal Pradesh of northeast India. The identification of collected samples was based on the comparison of morphological characters with those of herbarium specimens and following taxonomic monographs on *Citrus* (Swingle 1943, Tanaka 1954, 1977, Bhattacharya and Dutta 1956, Swingle and Reece 1967, Mabberley 2004). Samples collected for the present study included all major *Citrus* species (sweet and sour orange, mandarin, citron, pomelo and grapefruit) and four species of the subgenus *Papeda* (*C. latipes*, *C. macroptera, Poncirus trifoliata* and *C. ichangensis)* (Table 1.1). Most of the collected species and varieties are also available at the Citrus Research Station, Assam Agriculture University, Tinsukia, Assam, India.

The species collected for this study were classified into 12 and 24 species based on the Swingle and Reece (1967) and Tanaka (1969, 1977) classification systems (Table 1.1). Based on the morphological characteristics that were included in the above classification systems, the *Citrus* species in northeast India are categorized into four major groups (i) acid, (ii) orange/mandarin, (iii) pomelo/grapefruit and (iv) *Papeda*. The acid group members include citron, lemon and lime with the distinct characteristics including thorny shrub to small trees and straggling growth; leaves are large, oval to oblong, serrate margin, short, wingless petioles; flower small to large unbranched; fruits small to medium in size, shape long-oval to ellipsoid,

sometime necked, apex blunt, pointed to nipple, color green and yellow; smooth to rough fleshy thick rinds; low to high juice content and highly acidic with varied aroma, numerous seeds with white cotyledons. The distinct characteristics of mandarin/orange are: small to medium size spreading and drooping tree; a few thorns to thornless; leaves large, petiolated, dark green, lanceolate, tapering at the base and apex; flowers single; fruits small to medium, loose skin, oblate to sub-globose, sometimes slightly necked, orange-red colored, easily separable thin and leathery rind, sweet flavor and aroma, sometimes slightly acidic, a few seeds to seedless with greenish cotyledons. The characteristics of the pomelo/grapefruit are: large and spreading trees; less woody thorns; round-pointed, glabrous, petiolated large and broadly winged leaves; the flowers are large and branched, fruits born in single to cluster, fruit size medium to large and very large, shape round, obovate, or pyriform; thick spongy tightly adherent yellow, red, pink and white colored segmented rinds; a few to large number of seeds, the flavors ranged from dull sweet to moderate acid with distinct aroma. *Papeda* are highly thorny deciduous shrub to small trees; simple to compound winged petiolated leaves; flowers are small and unbranched; fruits are small to medium, oblate, obovoid to globose, segmented, yellow color, thick tightly adherent rinds; the flavor bitter to acid; a few to large number of seeds with white cotyledons.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from dry leaf samples following Doyle and Doyle (1987) and Dayanandan et al. (1997). The quality and quantity of extracted DNA were assessed with Nanodrop UV-VIS Spectrophotometer (Thermo Scientific) and gel electrophoresis. The nuclear ribosomal internal transcribed spacer (ITS2) region and three chloroplast non-coding regions (trnL-F, trnS-trnG and rpsl6) were used for the present study (Bayer et al. 2009, Jena et al. 2009, Zhen-Hua et al. 2011, Kumar et al. 2013). The ITS2 region was amplified through polymerase chain reaction (PCR) using primers ITS3F (White et al. 1990) and ITS28ccR (Hillis and Dixon 1991) (ITS3F = 5′-GCATCGATGAAGAACGCAGC-3′ and ITS28ccR = 5′-GCCGTTACTAGGGGAATCCTTGTAAG-3′). The universal cpDNA primers (Taberlet et al. 1991) were used for PCR amplification of the trnL-trnF region (5′ GGTTCAAGTCCCTCTATCCC-3′ and 5′-ATTTGAACTGGTGACACGAG-3′). The rpsl6 intron was amplified using the rpsF (5′-GTGGTAGAAAGCAACGTGCGACTT-3′) and rpsR2 (5′-TCGGGATCGAACATCAATTGCAAC-3′) primers (Oxelman et al. 1997) and trnS-trnG

intergenic spacer between trnS and trnG was amplified using primers trnS (5′-GCCGCTTTAGTCCACTCAGC-3′) and trnG (5′-GAACGAATCACACTTTTACCAC-3′) (Hamilton 1999). The PCR amplifications were performed on a GeneAmp 9700 thermal cycler using 25 µl volume reactions containing 2.5 µl (10 ng) template DNA, 0.4 µl (0.5 U) Taq polymerase, 2.5 µl of 10 X PCR buffer, 2.5 µl of 2.5 mM $MgCl_2$, 2.5 µl of 0.2 mM dNTP, 1.0 µl each of forward and reverse oligonucleotide primers (2.5 pmol) and 12.5 sterile dH_2O. The thermal profile of PCR amplification of ITS2 was: initial denaturation for 1 min at 97°C, 35 cycles 1 min at 97°C, 45 sec annealing at 50 - 55°C, 2 min at 72°C, and 7 min extension at 72°C. The PCR amplification profile of trnL-F was: 4 min of initial template DNA denaturing, 35 cycles consisting of denaturing at 94°C for 45 sec, primer annealing at 52 - 55°C for 45 sec, and primer extension at 72°C for 5 min and final extension of 7 min at 72°C. The amplification conditions for trnS-trnG consisted of: initial 94°C for 2 min, 35 cycles of 94°C 1 min, 47 - 50°C annealing temperatures for 45 sec and 72°C for 2 min, and one cycle of 72°C for 5 min as final extension. The amplification conditions of rps16 were: 2 min of initial template DNA denaturing, 35 cycles consisting of denaturing at 94°C for 1 min, primer annealing at 55°C for 50 sec, and primer extension at 72°C for 2 min and final extension of 1 min at 72°C. The amplified PCR products were visualized under UV light after electrophoresis on 1.0% w/v agarose gels in TBE. Successfully amplified PCR products were sequenced at the Génome Québec Innovation Centre Sequencing Services on an Applied Biosystems 3730xl DNA Analyzers using the same primers used for PCR amplification.

Sequence alignment and analysis

The chromatograms of DNA sequences were assembled, edited and visually assessed using Geneious Pro 4.7.6 (Drummond et al. 2009) and the assembled consensus sequences were aligned using Clustal-W (Thompson et al. 1994) with default parameter settings. The conflicting or ambiguous bases were coded following IUPAC code. The resulting sequences were cross-checked with sequences in the NCBI data base using BLAST. The aligned sequences were exported into a NEXUS format file (Maddison and Maddison 2001) for phylogenetic analyses using PAUP* (version 4.0b8; Swofford 2001).

Phylogenetic analyses

Phylogenetic analyses were conducted using separate and combined data matrices of nuclear (ITS2) and chloroplast loci (rps16, trnL-trnF and trnS-trnG). Phylogenetic trees were

reconstructed based on maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) methods. To assess the congruency between chloroplast and nuclear data sets, a partition homogeneity test (Farris et al. 1994), also known as an incongruence length difference (ILD) test was performed. The ILD test was carried out in PAUP* with the heuristic search set to 1000 replicates and 10 random addition sequence replicates, TBR and 'Mulpars' option on. The pairwise sequence divergence among taxa was calculated using 'Show pair-wise distance' options in PAUP*. The nucleotide sequences of *Aegle marmelos* (L.) Corr. and *Murraya paniculata* from Genbank were used as outgroup (Gene Bank accession numbers: AF025507.1, EF176492.1, AY295268.1, FJ434169.1 KJ641529.1, EF176562.1, AY295254.1 and KM514676.1).

Parsimony analysis

The parsimony searches were performed on a Macintosh computer using PAUP* with all characters treated as unordered, independent and of equal weight; gaps were treated as missing data. During this process, heuristic tree searches were performed with the addition of 1000 random taxon sequence replicates, using the tree bisection reconnection (TBR) branch swapping option, and saving no more than ten trees per replicate. Once these trees were generated, a final heuristic search was conducted on the trees found by this method and all trees were allowed to swap to completion. A strict consensus tree was retrieved from the set of equally parsimonious trees resulting from the heuristic search in PAUP*. Goodness of fit scores of the trees including tree length (TL), consistency index (CI), retention index (RI) and homoplasy index (HI) were recorded. The consistency index measures the amount of homoplasy within a data set (Schuh 2000). Support for clades was evaluated using nonparametric bootstrapping (Felsenstein 1985) with 1000 simple sequence addition replicates, TBR branch swapping, saving no more than 10 trees per replicate and all characters were equally weighted. The clades with bootstrap (BS) values of 50-74% represent weak support, 75-84% moderate support and 85-100% strong support (Richardson et al. 2000).

Maximum likelihood analysis

ML analysis (Felsenstein 1981) was conducted with the RAxML 7.2.6 software (Stamatakis 2006), using RAxMLGUI, a graphical front-end for RAxML (Silvestro and Michalak 2012), by choosing the general time reversible model (GTR) with the GAMMA [I] rate heterogeneity algorithm. To determine the best-fit ML trees, I executed 10-tree searches from distinct random

stepwise addition sequence maximum parsimony starting trees and 1000 non-parametric thorough bootstrap replicates for nodal support. Finally, bootstrap support values were recorded on the strict consensus ML trees and visualized using FigTree v1.4.

Bayesian analysis

The appropriate models of nucleotide substitution for the individual and combined data sets were selected through Akaike and Bayesian information criteria (AIC and BIC) as implemented in jModelTest 2 (Darriba et al. 2012). The general time reversible (GTR+I+G) model was identified as the best-fit evolutionary model for all data sets. Bayesian posterior probability support for the clades was obtained using Metropolis Coupled Monte Carlo Markov chain (MCMCMC) analysis as implemented in MrBayes 3.2 (Huelsenbeck and Ronquist 2001). Two completely independent sets of 4 MCMC chains (three heated and one cold chain) were run for $1.5x10^5$ generations or until the standard deviation of their split frequencies was below 0.01. The standard deviation of the split frequencies and the log-likelihood values were examined graphically using Microsoft Excel and trees generated prior to reaching stationary phase were discarded as burn-in. Trees were sampled every 100^{th} generation and trees were summarized using the MrBayes default settings after discarding the first 25% of samples from the cold chain as burn-in. Multiple runs resulted in satisfactory convergence of the posterior probability distribution of the two tree samples of similar results. Finally, the 50% majority-rule consensus cladogram of Bayesian phylogenetic analyses with all the relevant clade support (posterior probability) values and branch length information was saved and trees were visualized using FigTree v1.4.

Results

The resulting data matrices included three plastid (trnL-trnF, trnS-trnG and rps16) and one nuclear (ITS2) nucleotide sequences of 24 *Citrus* species belonging to two subgenera (*Eucitrus* and *Papeda*) and two outgroup species (*Aegle marmelos* and *Murraya paniculata*). The pairwise sequence divergence in the individual genomic regions was low as compared to the combined sequences. The minimum pairwise sequence divergence observed in trnS-trnG region ranged from 0% (between mandarin, pomelo and acid groups of species) to 0.3% (between *C. nobilis* and *P. trifoliata*) with an average of 0.2%. The rps16 pairwise sequence divergence ranged from 0% (between the species of acid and mandarin groups) to 1.4% (between *C. latipes* and *C. indica*) with an average

of 0.7%. The nuclear ITS2 pairwise sequence divergence ranged from 0% (between the species of acid and mandarin groups) to 2.0% (between *C. grandis* and *C. pseudolimon*) with an average of 2.8% (between ingroup and outgroup species). Among the individual chloroplast genes the maximum pairwise sequence divergence was observed in the trnL-trnF region and the value ranged from 0% (between a large number of species of acid and mandarin groups) to 8.4% (between *C. jambhiri* and *C. medica*) with an average of 2.0%. Furthermore, the combined pairwise sequence divergence ranged between 0% only in two pairs of species (*C. limon* x *C. limettioides* and *C. ichangensis* x *C. rugulosa*) to 1.7% (between *C. jambhiri* and *C. medica*) with an average of 1.1%.

The trnL-trnF intron/intergenic spacer sequences varied in length from 253 bp in *C. rugulosa* to 323 bp in *C. medica*. The aligned sequences resulted in a matrix of 267 characters for 26 individuals and the matrix comprised 223 (83.52%) constant, 4 (1.49%) parsimony-informative and 40 (14.98%) parsimony-uninformative variable characters. In the aligned trnL-trnF sequence, the shortest insertions of 1bp occurred only in *C. assamensis* (coordinate 10 and 190) and single substitutions were recorded in *C. indica* at coordinates 117 (G→A), *C. karna*, *C. volkameriana*, *C. nobilis* at 44 (G→T) and *C. reticulata* at 154 (G→A). Single nucleotide substitutions at coordinates 44 (G→T) and 154 (T→G) were recorded in *C. grandis*, *C. limettioides*, *C. limon*, *C. ichangensis*, *C. rugulosa* and *C. pseudolimon*. Single base pair substitution observed in *C. aurantifolia* in three different coordinates: 138 (C→A), 154 (T→G) and 240 (G→A). Multiple nucleotide substitutions were also recorded in *C. jambhiri* (12:T→G; 27:G→A; 117: G→A; 198:G→A; 227:T→G and 240:G→A) and *C. medica* (4:T→A; 17:G→A; 20:A→C; 81:T→C; 90:T→G; 138:C→A; 154:T→G; 197:A→T and 207:G→A).

The aligned trnS-trnG data set resulted in a matrix of 674 characters, where 651 (96.58%) positions were constant, 15 positions (2.23%) were variable and 8 (1.18%) were potentially parsimony-informative. The individual trnS-trnG sequence length varied between 629 bp in *C. volkameriana* and 685 bp in *C. reshni*. In the trnS-trnG aligned sequences, single nucleotide substitution was recorded at coordinates 10 (A→T) in *C. medica* and *P. trifoliata*; 133 (T→A) in *C. macroptera* and *C. nobilis* and at 291(C→T) in two acid (*C. karna* and *C. limon),* one *Papeda (C. macroptera)* and in all mandarin species. Single nucleotide insertion (at coordinates 34 in *C. limon*, 476 in *C. assamensis*,

C. sinensis and *C. nobilis*) and deletion (at coordinates 475 in *C. pseudolimon*, *C. grandis* and in three *Papeda* species) were recorded.

The rps16 intron region data set consisted of 720 characters, of which 667 (92.63%) were constant, 38 (5.27%) were variable, and 15 (2.08%) were parsimony-informative. Individual sequence length varied from 760 bp in *C. megaloxycarpa* to 800 bp in *C. indica*. In the aligned rps16 gene sequences a single nucleotide substitution at coordinates 46 (T→A) and 74 (T→G) were recorded in all the species of mandarin and at 338 (G→A) and 503 (C→A) in all the members of acid *Citrus*. Single nucleotide substitutions at five different coordinates 46 (T→G), 80 (G→A), 261 (A→T), 375 (T→A), 322 (G→C) were recorded in *C. medica* and *C. indica;* and six nucleotides insertion at coordinates 381-386 and two deletions at 483 and 530 in *C. indica* were recorded. Single nucleotide insertion, deletion and substitutions were also recorded in multiple coordinates in two *Papeda* (*P. trifoliata* and *C. macroptera*) and one acid (*C. megaloxycarpa*) species.

The nuclear ITS2 sequence length of individuals ranged between 330 bp in *C. volkameriana* and 613 bp in *P. trifoliata*. The aligned ITS2 data matrix included 590 characters, of which 485 (82.20%) were constant, 96 (16.27%) were variable and only nine (1.52%) were potentially parsimony-informative characters. In the aligned nuclear ITS2 sequence a single nucleotide insertion (at coordinates 16, 21, 37 in *C. megaloxycarpa*; 59, 225, 287 in *C. reshni*; 21, 59, 94 in *C. pseudolimon*; 303, 307, 320 in *C. assamensis* and *C. grandis*; 320 in *C. indica*; 350, 366 in *C. nobilis*; 350 in *P. trifoliata*; 366 in *C. karna*) and deletions (at 220, 425 in *C. pseudolimon*; 374 in *C. karna* and *C. nobilis*; 256 in *P. trifoliata*) were recorded in multiple species and coordinates. Single nucleotides substitutions were recorded only in *C. indica* (86: C→T; 228: A→T) and *C. limonia* (121: C→T; 160: C→T) at two coordinates and at one coordinate 229 (G→A) in *C. karna, C. nobilis* and *C. pseudolimon* (341: G→T).

Among the three chloroplast regions used in this study, trnS-trnG was the least variable region and trnL-trnF region showed the lowest number of parsimony-informative sites compared to the other chloroplast regions. The chloroplast rps16 region showed the highest percentage (2.08%) of parsimony-informative sites (Table 1.2). In most cases, chloroplast sequences were identical or nearly identical within the examined *Citrus*

species. Maximum parsimony (MP) analysis of the individual chloroplast marker resulted in 2, 1740 and 8 parsimonious trees based on trnL-trnF, trnS-trnG and rps16 respectively. The consistency indices (CI) were 1.000, 1.000 and 0.981 and retention indices (RI) were 1.000, 1.000 and 0.975 respectively for the three chloroplast regions (Table 1.2). The GTRGAMMA [I] model was the best fitted substitution model for maximum likelihood (ML) analysis for the three different chloroplast genes that yielded single tree for each dataset with likelihood scores of -617.23, -1047.74 and -1279.95 respectively (Table 1.3). The observed nucleotide base frequencies and substitutions rates under the same model are mentioned in Table 1.3. In Bayesian (BI) analysis, mean –lnL values ranged between 0.679.97 in trnL-trnF to -1358.80 in rps16 with potential scale reduction factor values of 1.000 and standard deviation values of 0.004 - 0.009 (Table 1.4). Phylogenetic trees reconstructed using MP, ML and BI methods produced congruent topologies. As all tree topologies were identical and thus the MP topology is shown with bootstrap support (BS) for MP and ML analyses; and posterior probability (PP) values of BI analysis are given above the branches (Figures 1.1 -1.4). There was no clear phylogenetic resolution among the 24 *Citrus* species based on individual analysis of trnL-trnF and trnS-trnG chloroplast genomic regions. The topologies based on different analyses using these two chloroplast regions resulted in two lineages with mixture of species from different groups and the statistical support values were low (Figures 1.1 and 1.2). Cluster-1 consisted of 7-9 species from acid, mandarin, pomelo and *Papeda* groups and the cluster-2 comprised of the rest of the species from the different groups. However, the third chloroplast region rps16 showed six lineages and revealed better phylogenetic resolution among the species (Figure 1.3). Clade I comprised two species *C. indica* and *C. medica* supported by higher statistical values (BS = 98 in MP and 99 in ML, PP = 1.00 in BI). Clade II and III comprised of lone species *C. assamensis* and *C. macroptera* with lower support values. Clade IV comprises all the five mandarin species (*C. nobilis, C. reticulata, C. aurantium, C. sinensis* and *C. reshni*) (BS = 74 in MP and 64 in ML, PP = 0.89 in BI). Clade V comprises seven species having 6 acid members (*C.aurantifloia, C. limonia, C. volkameriana, C. limettioides, C. pseudolimon* and *C. limon*) and one wild *Papeda* (*P. trifoliata*) (BS = 62 in MP and 60 in ML, PP = 0.90 in BI). Clade VI comprises eight species having four pomelo (*C. megaloxycarpa, C. grandis, C. rugulosa* and *C. paradisi*),

two wild *Papeda* (*C. ichangensis* and *C. latipes*), and two acid members (*C. karna* and *C. jambhiri*) (BS = 63 in MP and 66 in ML, PP = 0.89 in BI).

The nuclear ITS2 region showed 1.52% of potential parsimony informative characters which is slightly higher than the two chloroplast region (trnL-F and trnS-G). MP analysis of the nuclear ITS2 region resulted in 2000 parsimonious trees having CI and RI values of 1.000 (Table 1.2). The different parameters estimation of the ML and BI analyses under GTRGAMMA [I] and GTR+I+G models are mentioned in the Tables 1.3 and 1.4. The MP tree (Figure 1.4) was identical to the ML and BI analyses. Three different analyses of the nuclear gene grouped these *Citrus* species into 3 different phylogenetic groupings, clade I formed by the lone *C. indica* species, clade II consisted of one acid (*C. assamensis*) and one sweet pomelo (*C. grandis*) and clade III formed by the rest of the *Citrus* species of different acid, mandarin, pomelo and *Papeda* members with lower statistical BS and PP support values (Figure 1.4). The relationships among the species were also not well resolved through the independent nuclear genomic analysis like the other two individual chloroplast (trnL-trnF and trnS-trnG) genomic analyses.

In general, separate and individual analyses of the chloroplast and nuclear data sets resulted in largely unresolved phylogenetic trees. The phylogenetic trees inferred from the concatenated chloroplast and nuclear data matrix showed better resolution. Thus, I will focus on the results based on the concatenated data set. The aligned concatenated data sets of the cpDNA and nrDNA sequences of trnL–trnF, trnS–trnG, rps16 and ITS2 were 267+674+720+590= 2252 bp long and comprised of 36 parsimony-informative sites (Table 1.2). Analyses of the aligned chloroplast and nuclear genome sequences resulted in well resolved phylogenetic trees of *Citrus* species. The incongruence length difference (ILD) test of the concatenated data sets confirmed that they are highly congruent (P = 1.00). The ILD test showed no conflicting phylogenetic signals in the combined data sets, allowing these markers to be combined in a single analysis and similar approach was reported to be useful in resolving conflicting phylogenies (Garcia-Jacas et al. 2001, Pridgeon et al. 2001, Finet et al. 2010). The results of this test for the combined data sets suggest that the phylogenetic signals in the data sets are homogeneous and can be combined. The phylogenetic trees inferred from the combined data sets provided the best estimate of phylogenetic relationships among these

taxa. Phylogenetic trees derived using MP, ML and BI methods revealed that *Citrus* species belonging to two subgenera in northeast India are polyphyletic, consisting of five clades.

Maximum parsimony (MP) analysis of chloroplast and nuclear DNA sequences together using equally weighted character states resulted in 1140 parsimonious trees with a length of 10,0000 steps, consistency index (CI) of 0.906 and retention index (RI) of 0.747. The ML analyses with the combined data set yielded the best ML tree with a log-likelihood score of -4567.61 under GTRGAMMA [I] as the best fitted substitution model. The base frequencies under the same model were A=0.309, C=0.206, G=0.211 and T=0.274 and rate matrix were A-C: 0.738, A-G: 0.988, A-T: 0.359, C-G: 0.595, C-T: 1.179 and G-T: 1.000 (Table 1.3). The BI analyses yielded 50% majority rule consensus tree with a log-likelihood score of -4632.05 under GTR+I+G model and base frequencies were A=0.309, C=0.219, G=0.194 and T=0.276 under the same model of nucleotide substitution. The nucleotide substitutions rates were A-C: 0.161, A-G: 0.214, A-T: 0.115, C-G: 0.071, C-T: 0.241 and G-T: 0.195 (Table 1.4). In general, the combined nuclear and chloroplast dataset showed greater mean log-likelihood score and the posterior probability values in the ML and BI analysis as compared to the individual sequence analysis. The trees produced by MP, ML and BI analyses are identical, comprising five lineages with similar topologies. Therefore, MP topology from the combined dataset is chosen as the primary tree and in the same tree statistical BS values for MP and ML; and PP values for BI analyses are provided above the branches (Figure 1.5). In all the analyses, Clade I comprised of two species (BS = 81 in MP and 70 in ML, PP = 0.83 in BI), *C. indica* and *C. medica*. Clade II comprised of only a single wild and endemic species (*C. assamensis*) (BS = 100 in MP and 72 in ML, PP=0.92 in BI). Clade III comprised of seven species in all the topologies (BS = 100 in MP and 83 in ML, PP = 0.96 in BI), including 6 acid members (*C.aurantifloia, C. limonia, C. volkameriana, C. limettioides, C. pseudolimon* and *C. limon*) and one wild *Papeda* (*P. trifoliata*). Clade IV comprised six species, including all five mandarin species (*C. nobilis, C. reticulata, C. aurantium, C. sinensis* and *C. reshni*) and one endangered and endemic *Papeda* species (*C. macroptera*) (BS = 100 in MP and 82 in ML, PP = 0.82 in BI). Clade V comprised of eight species including four pomelo (*C. megaloxycarpa, C. grandis, C. rugulosa* and *C.*

paradisi), two wild *Papeda* (*C. ichangensis* and *C. latipes*), and two acid members (*C. karna* and *C. jambhiri*) (BS = 100 in MP and 85 in ML, PP = 0.89 in BI) (Figure 1.5).

Discussion

The three methods of phylogenetic analyses using three chloroplasts and one nuclear genomic region data resulted in similar topologies. The combined sequences were most useful in resolving phylogeny, suggesting high information content in the combined data matrix. This also improved the phylogenetic resolution among the members of different groups of *Citrus* species. In general, the individual chloroplast and nuclear sequences have less polymorphism due to their conservative nature and yielded short branch lengths and made them less useful for resolving phylogenetic relationships at lower taxonomic levels of *Citrus*. The phylogenetic relationships obtained by three different analyses also suggest polyphyletic groupings of acid and *Papeda* members with the other *Citrus* species. Different analyses resulted in five phylogenetic clades and relationships among the different *Citrus* species are discussed in detail.

In the three different analyses, *C. medica* consistently grouped with *C. indica*, an endemic and endangered wild species in northeast India. Similar relationships between this two species also reported by other authors (Bhattacharya and Dutta 1956, Kumar et al. 2013, Malik et al. 2013, Hynniewta et al. 2014) (BS = 81 in MP and 70 in ML, PP = 0.83 in BI analyses). Similar relationship between these two species also suggested by previous cpDNA studies (Federici et al. 1998, Nicolosi *et al*. 2000, Jena et al. 2009). Distinct clustering of *C. medica* and *C. indica* are due to similar and multiple single nucleotide base pair substitutions in the aligned chloroplast and nuclear DNA sequences. *C. indica* is thought be a hybrid species between wild (*C. latipes*) and cultivated species (*C. reticulata*) (Mabberley 2004, Swingle and Reece 1967), based on a large number of morphological characters. However, Federici et al. (1998) rejected its hybrid origin through RAPD and RFLP studies. This study didn't find close relationship between these two species (*C. latipes* and *C. reticulata*). The separation of *C. indica* in three different analyses indicates that *C. indica* is not closely related to *C. reticulata* and *C. latipes*. This may be an indication that *C. indica* is a true species. Similar results based on chloroplast and nuclear sequence studies were also reported by Jena et al. (2009) and Kumar et al.

(2013). A recent chromosomal and ITS sequence study by Hynniewta et al. (2014) also concluded that *C. indica* is a true species and an ancestor to many other cultivated species.

In three different analyses, a wild and endemic species (*C. assamensis*) formed an independent clade from the rest of the *Citrus* species (BS = 100 in MP and 72 in ML, PP = 0.92 in BI). Insertions of single nucleotides in the aligned trnL-trnF, trnS-trnG and ITS2 sequences at different positions make it different from the rest of the *Citrus* species. This species is much divergent and did not show any close relationship with the rest of the *Citrus* species. This confirms its unique genetic identity and indicative of a true species. Though, *C. assamensis* shares some morphological similarities with other acid species (e.g., fruit shape and size), however, its strong acidic taste and smells similar to ginger (hence the regional common name 'adajamir') make it different from other *Citrus* species. This endemic species was first reported by Bhattacharya and Dutta (1956) from Assam and Meghalaya and this species is currently considered as a threatened species (Singh and Singh 2003, Malik et al. 2013). Hynniewta et al. (2014) reported its relationships with *C. latipes*, however, they have distinct morphological and cytogenetical differences among them. Further study with more accessions and molecular marker is recommended for its robust phylogenetic relationships and true identity.

Among the eight acid members, six of them (*C. aurantifolia*, *C. limonia*, *C. volkameriana*, *C. limettioides*, *C. pseudolimon* and *C. limon*) grouped with one of the wild *Papeda* (*P. trifoliata*) species (BS = 100 in MP and 88 in ML, PP = 0.96 in BI). Three acid members (*C. aurantifolia, C. limonia* and *C. volkamerina*) showed 100% genetic identity in the chloroplast and nuclear sequences and consistently grouped together with the other acid members in all the analyses (BS = 100 in MP and 70 in ML, PP = 0.95 in BI). *C. volkamerina* is morphologically very similar to lemon (*C. limon*), and this similarity is confirmed by other molecular marker studies (Nicolosi et al. 2000, Shahsavar et al. 2007, Tripolitsiotis et al. 2013). These three species have very similar morphological features in plant and fruits characteristics, and grouped together in all analyses. Fruits of these species are globular in shape and have thick rind, a rough skin surface, and highly acidic pulp and juice.

The mandarin/orange (*C. aurantium, C. nobilis, C. reticulata, C. reshni* and *C. sinensis)* cluster is well resolved in all the analyses (BS = 100 in MP and 77 in ML, PP = 0.82 in BI). Clustering of sour orange (*C. aurantium*) and sweet orange (*C. sinensis*) with the other mandarin species (*C. reticulata, C. reshni* and *C. nobilis*) in all the analyses in agreement with the notion of mandarin as one of their parental species. Similar relationships were also supported by a large number of prior and current molecular studies by Nicolosi et al. (2000), Araujo et al. (2003), Asadi Abkenar et al. (2004), Barkley et al. (2006), Pang et al. (2007), Lu et al. (2011), and Penjor et al. (2013). In all analyses, these sweet oranges showed close relationships with one of the wild *Papeda* (*C. macroptera*) (BS = 100 in MP and 83 in ML, PP = 0.92), further confirming genetic similarities between mandarin and *Papeda*. Federici et al. (1998) and Pessina et al. (2011) hypothesized that *C. macroptera* (as syn, *C. hystrix*) is an ancient member of the *Papeda* subgenus and Nicolosi et al. (2000) also reported *Papeda* affinity with mandarins (*C. reticulata*). In the current study, there was clear differentiation between mandarin and pomelo group of species (Barrett and Rhodes 1976, Nicolosi et al. 2000, Moore 2001, Garcia-Lor et al. 2013, Penjor et al. 2013). In a recent complete genome sequencing study of mandarin, pomelo and oranges, Wu et al (2014) concluded that sour and sweet oranges are pomelo and mandarin admixtures resulted through interbreeding either in domestication or in natural habitats. This study also revealed similar sour and sweet orange/ mandarin relationships but without any close relationships between orange/mandarin and pomelo group of species.

The four pomelo/grapefruit species (*C. grandis, C. megaloxycarpa,* C. *paradisi* and *C. rugulosa*) were grouped with two acid (*C. karna* and *C. jambhiri*) and *Papeda* members (BS = 100 in MP and 85 in ML, PP = 0.89 in BI). Chloroplast and nuclear sequence were 100% identical and similar nucleotide substitutions between acid, pomelo and *Papeda* species reveals genetic similarity among these species. Pomelo or grapefruit might also have played an important role as parents of many *Citrus* varieties such as acid and *Papeda* species, as evident in their close relatedness with two different groups of species in phylogenetic trees. A large number of workers have described pomelo as one of the true *Citrus* species (Barrett and Rhodes 1976, Federici et al. 1998, Nicolosi et al. 2000, Barkley et al. 2006) through morphological and molecular studies. One of the wild

Papeda (*C. inchangensis*) showed very close relationships with two pomelo/grapefruit species (*C. rugulosa* and *C. grandis*). *C. megaloxycarpa*, one of the sour pomelo members, grouped with other acid *Citrus* and *Papeda* in all the analyses and showed 100% chloroplast and nuclear sequence identity. Lushington (1910), Bhattacharya and Dutta (1956) and Tanaka (1977) considered this as a true species. Swingle and Reece (1967) and Nair and Nayar (1997) considered *C. megaloxycarpa* a probable hybrid species between *C. grandis* (as syn. *C. maxima*) and *C. limon.* However, this study showed close genetic relationship of *Papeda* (*C. latipes*) and an acid member (*C. karna*). Two other acid members *C. jambhiri* and *C. karna*, showed very close relationships with the sweet pomelo (*C. grandis*, *C. paradisi* and *C. rugulosa*) and with one of the acid pomelo (*C. megaloxycarpa*) member; these species have identical chloroplast and nuclear sequences and also shares similar morphological traits. Kumar and Nair (2013) reported close relationships between the species of acid and pomelo groups through ITS sequence studies. One of the sour pomelo species (*C. megaloxycarpa*) suspected to be a probable hybrid between *C. grandis* and *C. limon* (Nair and Nayar 1997) grouped with two other acid members and sweet pomelo in all the analyses, with significant BS and PP (100 in MP and 73 in ML, 89 in BI) support values.

Originally considered *Citrus* species *C. ichangensis, C. latipes*, *C. macroptera* and *P. trifoliata* are currently classified under the *Citrus* subgenus *Papeda* (Swingle 1943, Tanaka 1954). The close relationships between the four wild *Papeda* with the other *Citrus* species of acid, mandarin and pomelo groups suggesting that *Papeda* are closely related to *Citrus* at the DNA level. This result contradicts Swingle's classification of *Poncirus* in subgenus *Papeda* (Swingle and Reece 1967), and confirms more recent findings of close relationships between *Citrus* and *Poncirus*. Several other studies have revealed the close relationship between *Poncirus* and *Citrus* using cpDNA and nrDNA analyses (Araujo et al. 2003, Morton et al. 2003, Li et al. 2007, Bayer et al. 2009, Lu et al. 2011, Kumar et al. 2013, Penjor et al. 2013, Hynniewta et al. 2014). However, a few other studies of *Citrus* phylogeny using RFLP (Asadi Abkenar et al. 2004), SSR (Barkley et al. 2006) and RAPD (Nicolosi et al. 2000) found distant relationships of *Poncirus* with *Citrus*. Separation and nesting of *Papeda* species with three phylogenetic groups (acid, mandarin and pomelo) reveal their polyphyletic relationship. The present study also

supports its close relationships with pomelo and acid *Citrus*. The divergence of two other *Papeda* species *C. ichangensis* and *C. latipes* from the other members of *Papeda* and grouping with pomelo also reported in previous cpDNA marker studies (Federici et al. 1998; Nicolosi et al. 2000). Abkenar et al (2004) considered *C. ichangensis* as a hybrid between mandarin and other *Papeda*, however, this study also confirm its hybrid origin but with pomelo (*C. grandis* or *C. rugulosa*) and *Papeda* (*C. latipes*) members. Togetherness of acid, pomelo and *Papeda* members reveal their common genetic ancestry and long history of co-existence in cultivation and wild habitats in the region.

Conclusion

Citrus species of the eastern Himalayan region of northeast India are morphologically variable but have low level of genetic divergence in both chloroplast and nuclear DNA regions. There may not be as many as 24 or more true biological species that were described on the basis of horticultural/morphological characteristics. Chloroplast and nuclear DNA sequences phylogeny in the present study revealed five phylogenetic lineages among the *Citrus* taxa. This study further revealed the polyphletic relationships among the members of *Citrus* and *Papeda* subgenera. Besides the three well recognized true species (*C. grandis, C. medica* and *C. retiuclata*), other two species (*C. indica* and *C. assamensis*) may also be considered as true species that require further study with more accessions and molecular markers. In general, the topologies based on combined data sets showed higher resolution along the internal nodes within the *Citrus* species than previous molecular phylogenetic studies.

Table 1.1. Details of *Citrus* species collected for the present study.

Common name	Swingle and Reece System	Tanaka System	Status	Distribution in NE Indian states
Subgenus: Eucitrus				
1. Citron	*C. medica* L.	*C. medica* L.	W+D	All
2. Lemon	*C. limon* (L) Burm.f.	*C. limon* (L) Burm.f.	D	All
3. Acid lime	*C. aurantifolia* (Christm.) Swingle	*C. aurantifolia* (Christm.) Swingle	D	AS,AP,ML,MZ,MN
4. Sour orange	*C. aurantium* L.	*C. aurantium* L.	D	AS,MZ,ML,MN,TR
5. Sweet orange	*C. sinensis* Osbeck	*C. sinensis* Osbeck	D	AS,AP,ML,MZ,NG
6. Mandarin	*C. reticulata* Blanco	*C. reticulata* Blanco	D	All
7. Pomelo	*C. maxima* (Burm.) Merr.	*C. grandis* Osbeck	D	All
8. Grapefruit	*C. paradisi* Macf.	*C. paradisi* Macf.	D	AS,MZ,ML,NG
9. Indian wild orange	**C. indica* Tanaka	**C. indica* Tanaka	W+D	ML
10. Sweet lime		*C. limettioides* Tanaka	D	AS,ML,NG
11. Rough lemon		*C. jambhiri* Lush.	W+D	All
12. Rough lemon		**C. megaloxycarpa* Lush.	W	AS,AP,ML,MZ
13. Hill lemon		*C. pseudolimon* Tanaka	W	AS,AP,MN
14. Kharna khatta		*C. karna* Raff	W+D	AS,AP,MZ
15. Rangpur lime		*C. limonia* Osbeck	D	AS,TR,ML
16. King mandarin		*C. nobilis* Loureio	W	AS,NG
17. Spice mandarin		*C. reshni* Tanaka	D	AS,AP,MZ,TR
18. Ginger lime		**C. assamensis* Dutta & Bhatt.	W+D	AS
19. Volkamer lemon		*C. volkameriana* Ten et Pasq.	W	AS,ML
20. Attani		**C. rugulosa* Tanaka	W	NG
Subgenus: Papeda				
21. Ichang papeda	**C. ichangensis* Swingle	**C. ichangensis* Swingle	W	NG
22. Khasi papeda	**C. latipes (*Swingle) Tanaka	**C. latipes* (Swingle) Tanaka	W+D	AS,ML
23. Hatkhora	**C. macroptera* Montr.	**C. macroptera* Montr.	W+D	AS,MZ
24. Troyer citrange		*Poncirus trifoliata* (L.) Raf	D	AS

W: Wild, D: Domestic

* Endangered species in the region.

AS: Assam, AP: Arunachal Pradesh, ML: Meghalaya, MZ: Mizoram, MN: Manipur, NG: Nagaland, TR: Tripura.

Table 1.2. Sequence characteristics of the 24 *Citrus* species and comparison of statistics for MP analysis.

Statistics	trnL-trnF	trnS-trnG	rps16	ITS2	Combined (trnL-trnF+trnS-trnG+ rps16+ITS2)
Sequence length (range) (bp)	253-323	629-685	760-800	330-613	-
Aligned sequence length (bp)	267	674	720	590	2252
Number of constant characters	223(83.52)	651(96.58)	667 (92.63)	485 (82.20)	2051(91.07)
Number of variable characters	40 (14.98)	15 (2.23)	38 (5.27)	96 (16.27)	165 (7.32)
Parsimony-informative characters	4 (1.49)	8 (1.18)	15 (2.08)	9 (1.52)	36 (1.59)
Tree length	47	24	55	110	233
Consistency Index	1.000	1.000	0.982	1.000	0.906
Rescaled Consistency Index	1.000	1.000	0.957	1.000	0.677
Retention Index	1.000	1.000	0.975	1.000	0.747
Homoplasy Index	0.000	0.000	0.018	0.000	0.094

Value in parentheses is the percentage of the corresponding values.

Table 1.3. Maximum likelihood parameter estimation under the GTRGAMMA [I] model for the different data set.

Parameters	trnL-trnF	trnS-trnG	rps16	ITS2	Combined
Likelihood score	-617.23	-1047.74	-1279.95	-1322.43	-4567.61
Base frequencies					
A	0.312	0.353	0.354	0.201	0.309
C	0.196	0.715	0.141	0.332	0.206
G	0.186	0.144	0.196	0.315	0.211
T	0.305	0.328	0.309	0.153	0.274
Substitution rates					
A-C	0.632	1.131	0.926	0.460	0.738
A-G	0.948	0.771	0.681	1.646	0.988
A-T	0.296	0.395	0.787	0.821	0.359
C-G	0.812	0.378	0.499	0.325	0.595
C-T	0.478	0.497	0.751	2.335	1.179
G-T	1.000	1.000	1.000	1.000	1.000

Table 1.4. Bayesian estimates (mean tree length, SD=mean standard deviation, -lnL=likelihood score, PSRF= Potential scale reduction factor of 95% credibility interval of the posterior probability distribution, base frequencies and substitution rates) under (GTR+I+G) model for the different data set.

Parameters	trnL-trnF	trnS-trnG	rps16	ITS2	Combined
Mean	3.98	4.13	5.28	0.431	0.168
SD	0.004	0.009	0.007	0.007	0.004
-lnL	-679.97	-1119.96	-1385.80	-1379.74	-4632.05
PSRF	1.000	1.000	1.000	1.000	1.000
Base frequencies					
A	0.287	0.313	0.306	0.216	0.309
C	0.239	0.201	0.161	0.317	0.219
G	0.203	0.215	0.230	0.270	0.194
T	0.269	0.269	0.301	0.195	0.276
Substitution rates					
A-C	0.957	0.303	02.32	0.096	0.161
A-G	0.023	0.369	0.144	0.347	0.214
A-T	0.009	0.081	0.160	0.046	0.115
C-G	0.002	0.011	0.017	0.075	0.071
C-T	0.061	0.056	0.326	0.372	0.241
G-T	0.006	0.177	0.117	0.060	0.195

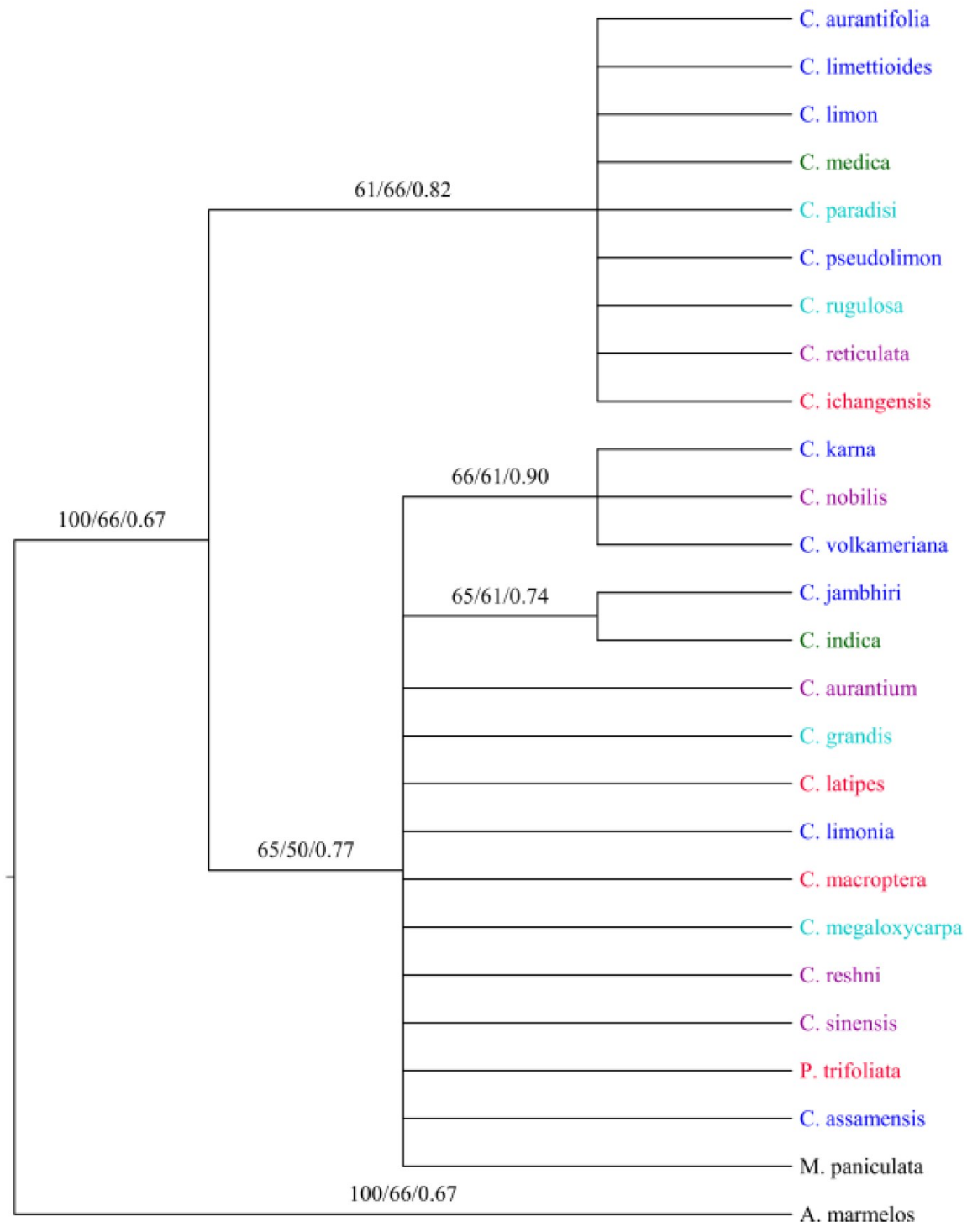


Figure 1.1. Maximum parsimony analysis results. Strict consensus of 2 equally parsimonious trees based on trnL-trnF chloroplast sequences. Bootstrap support values (BS) for MP and ML analyses; and posterior probability (PP) values for BI analysis are shown above the branches (MP/ML/BI).

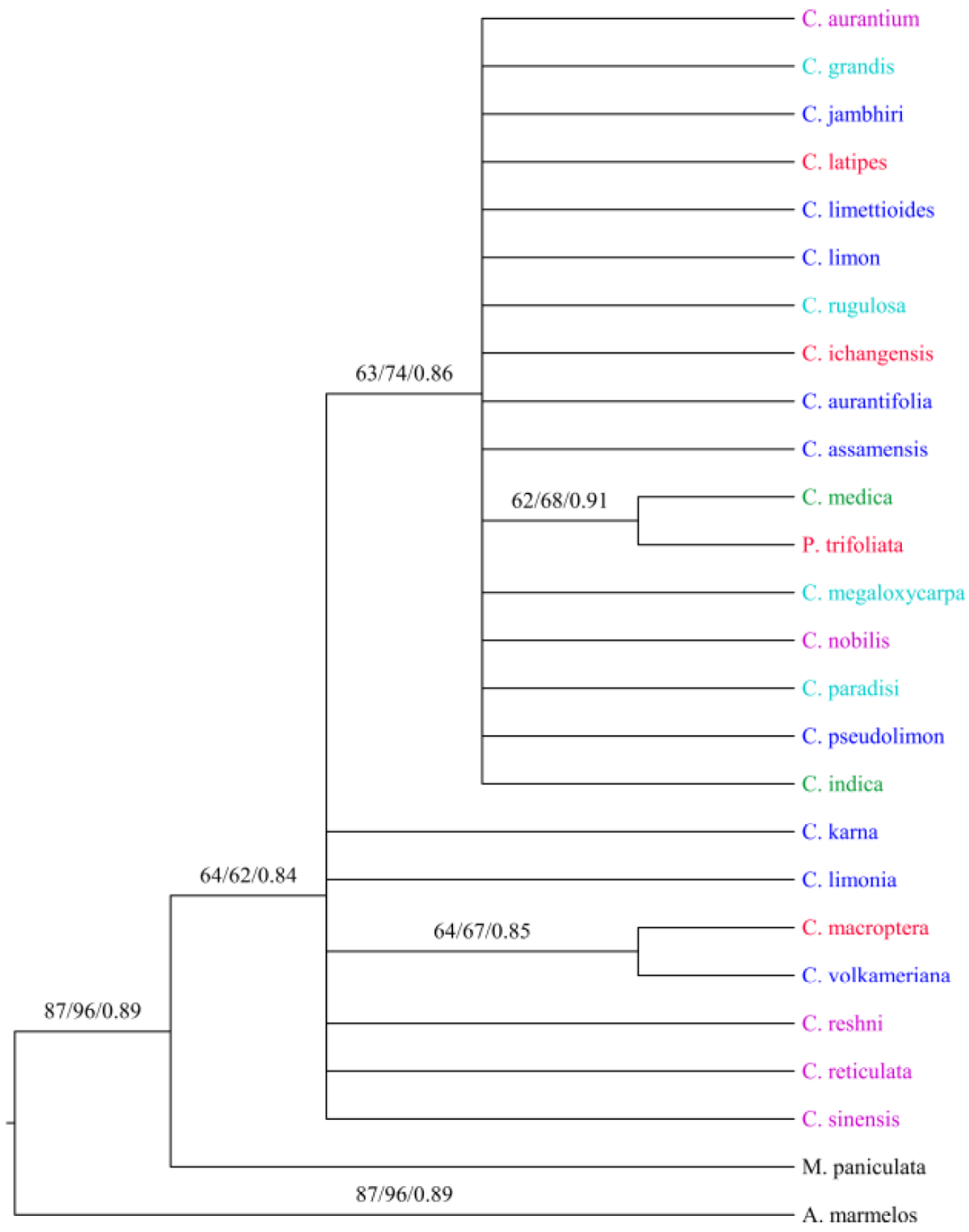


Figure 1.2. Maximum parsimony analysis results. Strict consensus of 1740 equally parsimonious trees based on trnS-trnG chloroplast sequences. Bootstrap support values (BS) for MP and ML analyses; and posterior probability (PP) values for BI analysis are shown above the branches (MP/ML/BI).

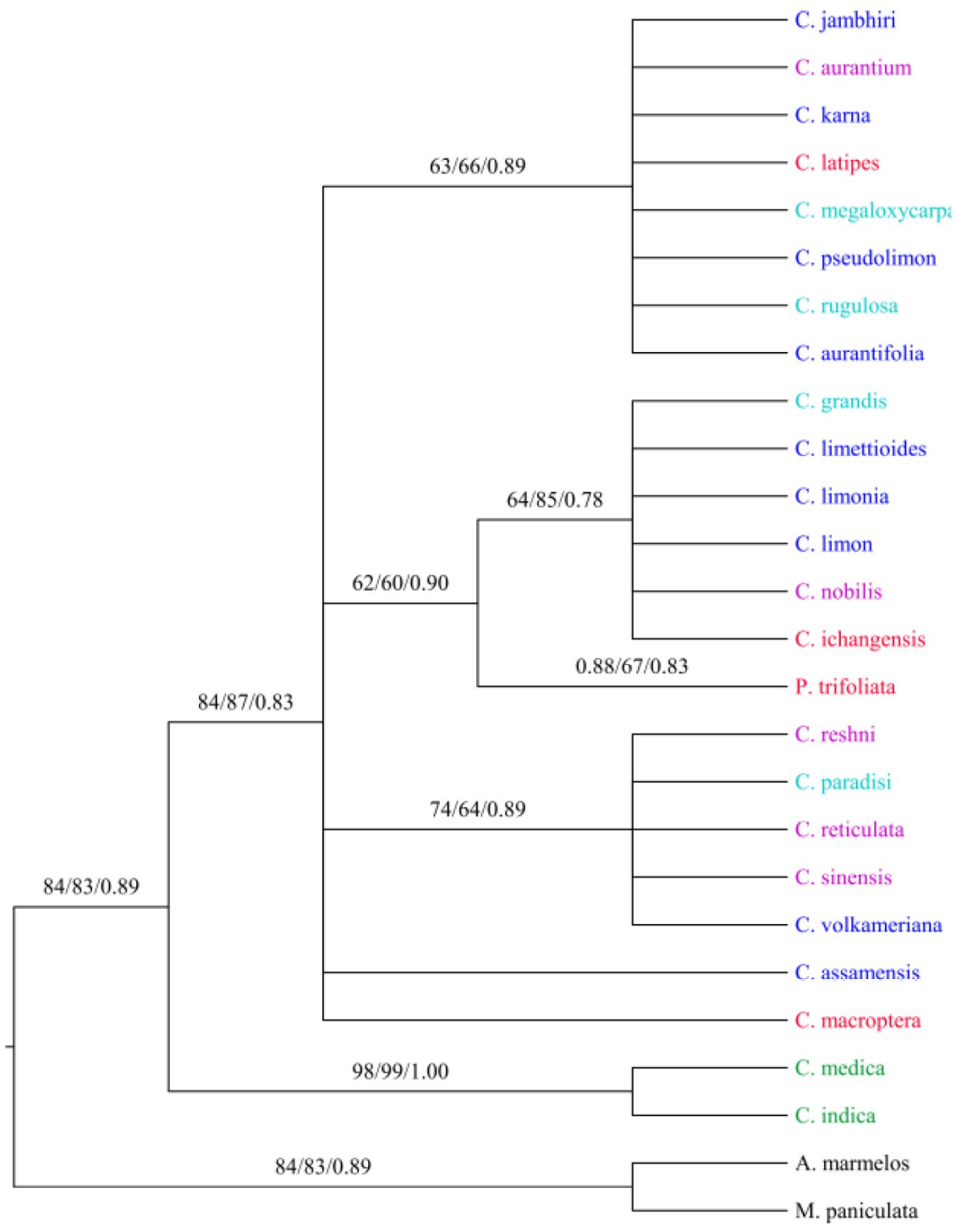


Figure 1.3. Maximum parsimony analysis results. Strict consensus of 8 equally parsimonious trees based on rps16 chloroplast sequences. Bootstrap support values (BS) for MP and ML analyses; and posterior probability (PP) values for BI analysis are shown above the branches (MP/ML/BI).

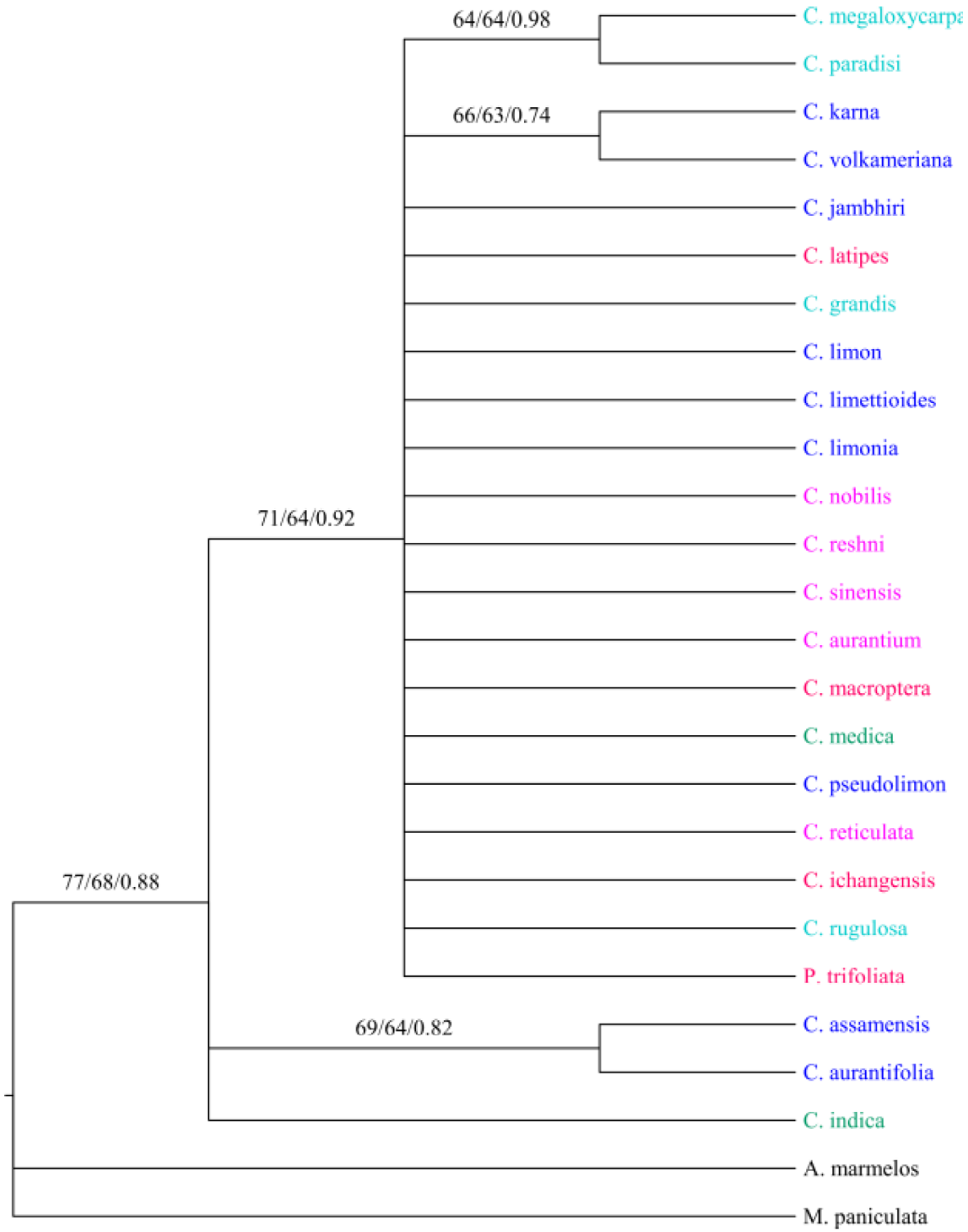


Figure 1.4. Maximum parsimony analysis results. Strict consensus of 2000 equally parsimonious trees based on ITS2 nuclear sequences. Bootstrap support values (BS) for MP and ML analyses; and posterior probability (PP) values for BI analysis are shown above the branches (MP/ML/BI).

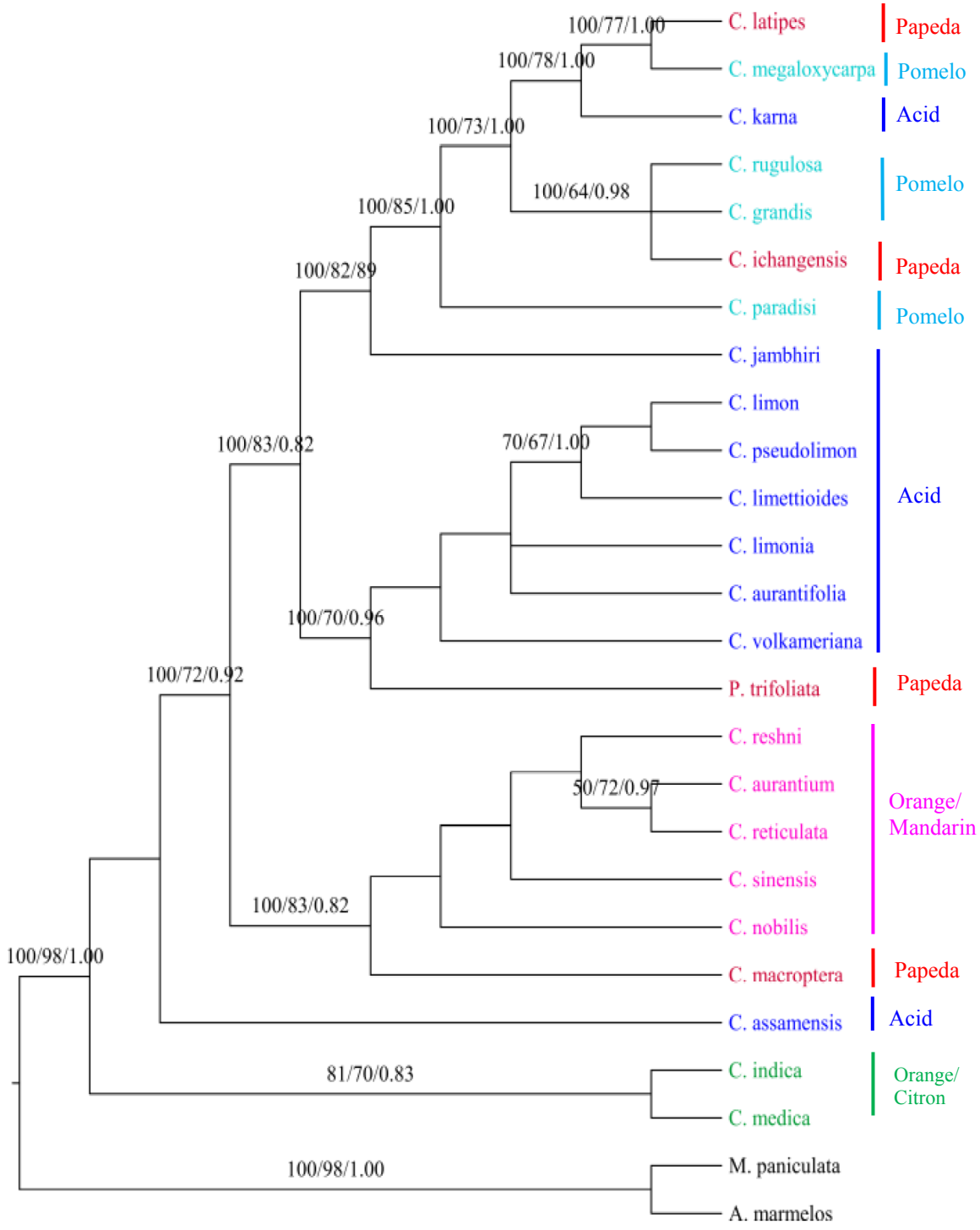


Figure 1.5. Maximum parsimony analysis results. Strict consensus of 1140 equally parsimonious trees based on the combined nuclear and chloroplast data sets. Bootstrap support values (BS) for MP and ML analyses; and posterior probability (PP) values for BI analysis are shown above the branches (MP/ML/BI).

Chapter 2: Genetic structure and diversity of natural and domesticated populations of *Citrus medica* in the Eastern Himalayan region of Northeast India

Abstract

Citron (*Citrus medica* L.) is a medicinally important species of citrus native to India and occurs in natural forests and home gardens in the foothills of the eastern Himalayan region of northeast India. The wild populations of citron in the region have undergone rapid decline due to natural and anthropogenic disturbances and most of the remaining individuals of citron are found in fragmented natural forests and home-gardens in the region. In order to assess the genetic structure and diversity of citron in wild and domesticated populations, I analyzed 219 individual of *C. medica* collected from four wild and eight domesticated populations using microsatellite markers. The genetic analysis based on five polymorphic microsatellite loci revealed an average of 13.40 allele per locus. The mean observed and expected heterozygosity values ranged between 0.220 - 0.540 and 0.438 - 0.733 respectively among the wild and domestic populations. Domestic populations showed close genetic relationships as compared to wild populations and pairwise Nei's genetic distance ranged from 0.062 to 2.091 among wild and domecated populations. Analysis of molecular variance (AMOVA) results showed higher genetic diversity among- than within-populations. The analysis of population structure revealed five groups. Mixed ancestry among a few individuals of different populations revealed their intercrossing through the exchange of genetic materials among farmers in the region. Citron populations in the region show high genetic variation. The knowledge gained through this study is invaluable for devising genetically sound strategies for conservation of citron genetic resources in the region.

Key words: Admixture, citron, domestic, diversity, Himalaya, wild.

Introduction

Citrus medica L., commonly known as citron, is native to India (Scora 1975, Mabberley 2004) and occurs as wild and semi-wild populations in both primary and secondary forests in the foothills of the Himalayas in northeast India (Hooker 1875, Bhattacharya and Dutta 1956, Tanaka 1977, Nair and Nayar 1997). Citron fruits are widely used in local medicinal practices and are a socioeconomically important genetic resource of the region. Citron is considered to

have been a parental contributor to several cultivated *Citrus* accessions, and has mostly acted as the male parent (Nicolosi et al. 2000). In combination with sour orange (*Citrus* × *aurantium*), citron contributed to the origin of lemon (*Citrus limon*), bergamot (*Citrus bergamia*) and key lime (*Citrus aurantifolia*) (Moore 2001, Barkley et al. 2006, Li and Xie 2010, Ollitrault et al. 2010). Natural populations of citron are severely affected by harvesting and deforestation, and most of the remaining individuals are confined to home gardens and agroforestry systems in the region. Thus, conservation measures are urgently needed to prevent further decline of citron genetic resources, and information on its genetic structure and diversity is essential for formulating conservation and management strategies.

A limited number of population genetic studies of citron using RFLP (Federici et al. 1998), RAPD, SCAR and cpDNA (Nicolosi et al. 2000), and SSR and ISSR (Corazza-Nunes et al. 2002, Barkley et al. 2006, Kumar et al. 2010) markers are reported in the literature. Through RFLP analyses, Federici et al. (1998) reported low heterozygosity levels among three *C. medica* accessions in the Citrus Variety Collection (CVC) at the University of California, Riverside. Barkley et al. (2006) studied 29 citron accessions from the CVC using SSR markers and reported lower heterozygosity values among the *C. medica* accessions as compared to the other *Citrus* species. The low genetic diversity observed among citron accessions could be attributable to selfing, as citrons are known to produce vigorous, highly homozygous seedlings through selfing (Barrett and Rhodes 1976). Genetic studies based on ISSR data also revealed a low level of heterozygosity (Ht = 0.160) among the seven accessions of *C. medica* in northeast India (Kumar et al. 2010). However, Luro et al. (2012) reported high diversity among citron varieties in the Mediterranean region, which could be attributable to inter-varietal pollination and seed introductions from Asia. Using RAPD and cleaved amplified polymorphic sequence markers, Nicolosi et al. (2000) reported high genetic diversity among twelve varieties of citron. These studies are based on a limited number of *C. medica* accessions and the genetic diversity of citrons in their native habitat remains unknown.

The present study, based on an extensive sampling from northeast India, is the first to assess the genetic variability of *C. medica* in its natural habitat. The main objective of the present study is to assess the genetic diversity and structure of wild and domesticated populations of *C. medica* over a broad geographical area. The specific objectives of the present study are to (1) determine the levels of genetic diversity in wild and domesticated populations of *C. medica,* (2)

determine whether the domestication process led to a reduction in genetic diversity (3) assess genetic structure and diversity of *C. medica* in its native habitat and (4) assess genetic relationships among wild and domesticated populations.

Materials and methods

Leaf samples from 219 individuals of *C. medica* representing four wild and eight domesticated home garden populations in Assam, Arunachal Pradesh and Mizoram (Figure 2.1, Table 2.1) were collected and stored dry until further analysis. A total of 20 individuals per population, with the exception of Neairgram and Namasi populations where 15 and 4 individuals respectively were available, were sampled. Morphological features including tree height, leaf length and width, fruit shape, size and weight were recorded during sampling.

The total genomic DNA from leaves was extracted following the methods of Doyle and Doyle (1987) and Dayanandan et al. (1997). The quality of extracted DNA was tested through electrophoresis on 0.5% agarose gel and staining with ethidium bromide. The PCR amplification of simple sequence repeat (SSR) loci were carried out following Barkley et al. (2006, 2009) and Ollitrault et al. (2010) in 15 µl reactions containing 2.0 µl template DNA, 0.2 µl Taq polymerase, 1.5 µl of 10 X PCR buffer, 1.5 µl of 2.5 mM $MgCl_2$, 1.5 µl of 0.2 mM dNTP, 0.5 µl of the forward and reverse oligonucleotide primers (2.5 pmol each) and 0.5 µl of the M13 universal forward primer (1 pmol/µl), 0.5 µl DMSO and 6.3 µl sterile dH_2O. Thermal cycling parameters consisted of initial denaturation at 94 °C for 4 minutes followed by 35 cycles of 94 °C for 1 minute, 50–55 °C for 45 s (primer specific annealing temperature, Table 2.2), and 72 °C for 1 minute and final extension at 72 °C for 7 minutes. PCR reactions were performed on a GeneAmp PCR System 9700 thermal cycler.

Each forward oligonucleotide primer consisted of M13 tail sequence (5′-CACGACGTTGTAAAACGAC-3′) at the 5′ end for visualization of the PCR product using M13 primers labelled with IRD700 and IRD800. The amplified PCR products were diluted (1:20) with loading dye (Formamide and Bromophenol blue), denatured at 94 °C for 5 minutes and cooled on ice before loading onto the 6% polyacrylamide gel on a LI-COR IR^2 DNA analyzer. About 1 µl aliquot of each PCR product was loaded onto each lane of the gel along with 3 lanes containing a 50 -350 bp size standard (LI-COR). The fragment size corresponding to each SSR

marker of each sample was scored using the e-seq software and the bands recorded as 1 (present) or 0 (absent) on an EXCEL sheet for further analysis.

Microsatellite data analysis

The obtained allele frequency data for all populations and markers were tested for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) using POPGENE Version 1.31 (Yeh et al. 1999). The average number of alleles per locus (Na), the observed heterozygosity (Ho), the expected heterozygosity (He) as well as the mean number of alleles (MNA), allelic richness (A_R), private allele (A_P), genetic differentiation (F_{ST}) and inbreeding coefficient (F_{IS}) in each population and locus were calculated using software programs POPGENE Version 1.31 (Yeh et al. 1999), FSTAT version 2.9.3.2 (Goudet 2001) and Arlequin Version 3.0 (Excoffier et al. 2005). The Polymorphic Information Content (PIC) for each SSR microsatellite locus based on the entire set of accessions was calculated using Power Marker V3.25 (Liu and Muse 2005). Pairwise standard genetic distances (D_S) among the 12 domestic and wild populations were calculated following Nei's unbiased measures of genetic distance (Nei 1978) using the POPGEN 32 software package and the resulting genetic distance matrix was used for a cluster analysis through unweighted pair-group method with arithmetic averages (UPGMA). The F-statistics (F_{IS} = inter-individuals, F_{IT} = subpopulations and F_{ST} = total population; Wright 1978) were computed to estimate genetic differentiation among the twelve *C. medica* populations. POPGENE Version 1.31 (Yeh et al. 1999) was used to estimate the significance of genotypic differentiation between population pairs. All probability tests were based on the Markov chain method (Guo and Thompson 1992, Raymond and Rousset 1995) using 1000 dememorization steps, 100 batches and 1000 iterations per batch. When the null hypothesis was rejected, the F_{IS} statistic of Wright (1951) was estimated following Weir and Cockerham (1984) and used as an indicator of heterozygote excess or deficit. The F_{ST} statistic (Wright 1951) was estimated following Weir and Cockerham (1984) and pairwise tests of differentiation were performed in FSTAT. Permutation tests were performed in FSTAT, where genotypes were randomized among samples and the significance of the P-values from the pairwise tests of differentiation were determined using standard Bonferroni corrections.

Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was performed in Arlequin 3.0 software (Excoffier et al. 2005) to test the differentiation of the accessions in

various groups with the probability of non-differentiation (F_{ST} = not > 0) over 10000 randomizations. The distribution of genetic variation within and among wild and domesticated populations was estimated using Nei's standard genetic variation (Nei 1987). Pairwise F_{ST} values between all pairs of populations were calculated and differentiations were tested between the populations in Arlequin. To examine the geographic structure of genetic variation among the *C. medica* populations, I tested for correlations between genetic distance and geographic distance using a Mantel test based on a pairwise matrix of Nei's (1978) unbiased genetic distances, Rousset (1997) genetic differentiation $F_{ST}/(1-F_{ST})$ and a pairwise matrix of geographic distances (Mantel 1967). Gene flow (Nm) among populations was estimated as the number of migrants per generation between pairs of populations. Nm was estimated according to Slatkin (1993) by using the formula $Nm = 1/(4F_{ST}) – 1/4$.

Genetic bottlenecks among the populations were identified using the program BOTTLENECK version 1.2.02, under three different models, the infinite allele and stepwise mutation models (Cornuet and Luikart 1996), and the two-phased model of mutation (Luikart et al. 1998). Both the Wilcoxon signed-rank test and a sign test were used to assess significance of whether the observed He is greater than expected under an equilibrium model.

The software program STRUCTURE version 2.1 (Pritchard et al. 2000) was used for the analysis of population structure and identification of ancestral and hybrid forms. This method follows a Bayesian clustering approach to assign individuals into clusters using multilocus genotype data and allele frequencies. This approach works on the principle that the loci selected for investigation are unlinked, independent and at linkage equilibrium among the populations under the Hardy-Weinberg principle (Pritchard et al. 2000). Different accessions were assigned to probable clusters under the assumption that all accessions were from a common ancestor and that admixing of individuals among the populations had occurred. The posterior probabilities were estimated using a Markov Chain Monte Carlo (MCMC) method. The admixture of individuals independent of the geographic locations was used for clustering all individuals from the study populations and 15 independent runs of STRUCTURE were carried out for the total data set for *K* (number of clusters) values of 1 to 15. Simulations were carried out with the following settings: admixture model, correlated allele frequencies and MCMC repetitions of 10,000 iterations. The final results were based on a run length of 100,000 and five iterations for each *K* using admixture model with the independent frequency and correlation model. I

examined ΔK values, which are derived from the second-order rate of change of the likelihood function used to determine K (Evanno et al. 2005), to provide a better estimate of the number of clusters in such conditions. For the number of clusters best represented by the data, only individuals with probabilities above the threshold q = 0.75 for a specific cluster were retained in that population.

Results

Characteristics of the seven SSR markers used to assess genetic diversity of the 219 *Citrus medica* individuals are given in Table 2.2. Five of the seven primer pairs described by Barkley et al. (2006, 2009) and Ollitrault et al. (2010) were used for genetic analysis. Two of the seven markers, cAGG9 and CCTO1, were excluded from the analysis due to their low polymorphism and poor amplifications. All SSR loci used in the present study were polymorphic and none of the loci deviated from Hardy-Weinberg equilibrium. No significant linkage disequilibrium was found in any pairs of loci, so all five SSR microsatellite loci provided independent information. A total of 67 alleles were detected within the citron individuals, with allele frequencies across all loci ranging from 2.50% to 82.50%. The number of alleles generated by each SSR marker varied from 8 to 20 with an average of 13.4 alleles per locus (Table 2.3). The highest number of alleles was scored at locus CiBE3936 (20 alleles) and lowest number of alleles at locus CiBE4796 (8 alleles) (Table 2.3). The effective number of alleles (Ne) for each locus ranged from 3.66 to 6.25 with an average value of 4.85. The amplified fragment size of the alleles varied from 131 (CiBE3936) to 248 (CiBE3298) bp for all five loci. The PIC values ranged between 0.829 (CiBE3936) and 0.694 (CiBE0753) with a mean PIC value of 0.762 for all loci (Table 2.3).

The total number of alleles across all loci ranged between 13 in the Namsai wild population and 36 in the Banskandi domestic population. The mean allelic richness (A_R), independent of sample size, ranged between 3.83 in the Tinsukia wild population to 2.48 in the Sairang[2] domestic population (Table 2.4). Overall genetic diversity varied significantly within wild and domesticated populations located in different geographic locations. The MNA across all populations was 2.77 ± 0.17, varying between 2.60 ± 0.55 in the Namsai wild population, which had the lowest number of individuals (4), and 7.20 ± 2.95 in the domesticated Banskandi population. In general, a higher MNA was observed in the domesticated populations. Most of the alleles present in domestic populations were also present in wild populations. Private alleles,

unique to a specific population, were observed in the Itanagar domestic population ($A_P = 4$), as well as in the Tinsukia wild, Banskandi domestic, Aizawl domestic and Sairang[1] wild populations, each with two private alleles, and in the Sairang[2] and Motinagar[1] domestic populations, which had one private allele each. No private alleles were found in any of the other populations (Table 2.4). The frequencies of these private alleles ranged between 2.50 - 12.50%.

The mean observed (Ho) and expected (He) heterozygosity values varied significantly ($P < 0.001$) within the populations (Table 2.4). The highest value for Ho = 0.540 ± 0.251 was observed in the domesticated Banskandi population, while the lowest Ho = 0.220 ± 0.160 occurred in the Tinsukia wild population. The highest He within the populations was found in the Tinsukia wild population (He = 0.733 ± 0.093), while the lowest occurred in the Sairang[2] domestic population (He = 0.438 ± 0.217). The He values for wild populations ranged from 0.500 - 0.733, and for domestic populations it ranged from 0.438 - 0.706. This wide range of heterozygosity values indicates high diversity within the populations. In all cases, average observed heterozygosities were lower than the expected heterozygosities under HWE (Table 2.4).

Population differentiation F_{ST} values were calculated for each locus and population separately and slight variation was observed among loci (Table 2.3) and populations (Table 2.4). The F_{ST} values ranged between 0.174 - 0.252 in wild populations and 0.193 - 0.294 in domestic populations, with slightly greater values in domestic populations. The F_{ST} values and their level of significance for pairs of populations were also calculated (Table 2.5). Among the twelve pairs of populations, only two pairs were not significantly differentiated, viz., Banskandi (domesticated) and Tinsukia (wild), Aizawl (domesticated) and Itanagar (domesticated). All other population pairs were significantly differentiated and the significance level in the most of the population pairs was $P < 0.001$ (Table 2.5). The greater and significant F_{ST} values between these population pairs may indicate greater genetic divergence in citron populations among these pairs. Inbreeding coefficient (F_{IS}) values were significantly positive ($F_{IS} = 0.204 – 0.705$; $0.001 < P < 0.05$) for all the populations except for one wild population in which it was positive but insignificant ($F_{IS} = 0.115$; $P > 0.05$) (Table 2.4). In all loci, significantly positive F_{IS} values were obtained and these ranged between 0.204 – 0.548. The average value of F_{IS} for all loci was 0.334 and F_{IT} was 0.511 for all accessions (Table 2.3). The gene flow (Nm) was calculated according

to genetic differentiation and it ranged between 0.600 in the Sairang^2 domestic population to 1.187 in the Tinsukia wild population (Table 2.4).

The pair-wise Nei's genetic distance (D_S) values are summarized in Table 2.5. In general, domestic populations showed close genetic relationships as compared to wild populations. The pairwise D_S vales between populations ranged from 0.062 between the Sairang^1 wild and Sairang^2 domestic populations in Mizoram to 2.091 between two domestic populations, Sairang^2 (Mizoram) and Neairgram (Assam). Similar results were observed when the genetic distances of the populations in the study were determined using Nei's D_A index (Nei's unbiased genetic distance) of genetic distance. The smallest D_A was observed between the Sairang^1 wild and Sairang^2 domestic populations (0.049) and largest D_A was observed between the Neairgram and Sairang^2 domestic populations (2.074) (Data not shown here). The AMOVA showed significant total genetic variation among the populations and individuals ($P < 0.001$) for all variance components. The genetic differences were 27.49% among individuals within populations, 24.98% among populations, and 47.53% at the individual level (Table 2.6).

Genetic relatedness between wild and domesticated populations was determined using Nei's standard and unbiased genetic distances and UPGMA methods. The UPGMA dendrogram showed five different clusters of *C. medica* accessions for all twelve populations and there was an admixture of individuals between wild and domestic populations. The first cluster comprised two geographically isolated populations, Tinsukia (wild) and Banskandi (domestic); the second cluster consisted of distant populations Itanagar and Aizawl (both domestic); the third cluster contained the Sairang^2 (wild), Sairang^1 (domestic) and Motinagar^1 (domestic) populations, which are located in the same geographic region; the fourth cluster was formed by the Motinagar^2 (wild) and Lakhipur (domestic) populations; and the fifth cluster was made up of two proximate domestic populations Sonai and Neairgram and the distant, wild Namsai population (Figure 2.2).

The STRUCTURE analysis revealed five distinct clusters ($K = 5$) represented by the individuals having posterior probability values above the threshold value q =0.75 (Figure 2.3). The assignment of individuals into different wild and domestic population groups are presented in Figure 2.4. Bayesian clustering analysis assigned 219 accessions into five genetically inferred clusters. Cluster 1 mainly comprises individuals of three different populations, among them one wild population, #1 (34%), and two geographically isolated domestic populations, #6 (36%) and, #7 (30%). Cluster 2 is dominated by individuals of three geographically isolated domestic

populations #2 and 3 (26% each) and #4 (24%), and one distant wild population, #1 (24%). Cluster 3 contains individuals belonging to same geographic location of four domestic populations # 2 and 7 (6% each), # 9 (36%) and one wild population # 8 (36%). Cluster 4 has individuals from one distant wild population, #1 (24%), and three distantly located domestic populations, #2 (18%), #3 (26%) and #4 (29%). In cluster 5, the majority of the accessions were contributed by two geographically isolated domesticated and wild populations #11 (38.5%) and #12 (38.5%) and two other populations # 1(4%) and #10 (19%). (Table 2.7; Figure 2.4).

Correlation between geographic distance (km) and Nei's genetic distance among the citron populations of NE India was insignificant (Figure 2.5). The geographic distance among the populations ranges from 0.01 km to 535 km. A Mantel test also showed no significant correlation between geographic distance and genetic differentiation [$F_{ST}/(1–F_{ST})$] for *C. medica* populations in the region (Figure 2.6). Thus, genetic distances between populations are independent of the corresponding geographical distances.

Discussion

The present study is the first to quantify the amount and distribution of genetic variability in *C. medica* within its native geographical range. The results, based on genotypes of five selected SSR loci, demonstrate that domesticated citron populations possess a slightly higher genetic diversity than wild populations and the difference between those populations were insignificant. High levels of polymorphism in the five selected SSR markers allowed me to unambiguously distinguish 219 accessions belonging to twelve geographically isolated populations.

Overall diversity values obtained in the present study differ from those found by Ollitrault et al. (2010), who reported low genetic diversity (He = 0.15, 1.44 alleles per locus). A prior study by Barkley et al. (2006) also reported lower diversity indices between citron individuals. These differences in genetic diversity between the present and previous studies are probably at least partly due to sample size as far fewer individuals were sampled in these earlier studies. More importantly, current sampling from different regions throughout its native range, rather than from small numbers of accessions in *ex situ* germplasm banks may have resulted in a better reflection of the genetic diversity present in *C. medica*. This results show that there is abundant genetic variation at the molecular level among the 219 citron individuals from four

wild and eight domestic populations throughout northeast India, where the species thought to have originated.

The domesticated populations of *C. medica* have slightly higher genetic diversity as compared to those wild populations. In general, all the populations have lower observed heterozygosity values then the expected heterozygosity suggesting inbreeding. Slightly higher genetic diversity among the domesticated populations suggest that movement of cultivated individuals through a large geographic distances resulting in allele combinations which would not occur naturally (Miller and Gross 2011). The exchange of such highly valued medicinal plants in the form of seed, seedlings and mature plant cuttings, sometimes over long distances, is a common practice among tribal and non-tribal communities in the region. Most likely farmers may have selected individuals with desirable traits, which may have contributed to the increased genetic diversity in domesticated populations. This may have lead to increased mixing and gene flow among geographically isolated populations.

An average $F_{ST} = 0.275$ for overall loci revealed significant genetic differentiation between populations. Similar moderate-to-high F_{ST} values are consistent with the relatively high genetic differentiation observed in some other tropical trees *Caryocar brasiliense* (Collevatti et al. 2001), *Swietenia macrophylla* (Novick et al. 2003) and *Dalbergia monticola* (Andrianoelina et al. 2009). These results also reflect genetically distinct populations in the region differing simultaneously in allele frequencies and allele sizes, and suggest that new mutations may be contributing to the allelic diversity found in wild and domestic citron populations. In general, wild and domestic citron populations showed strong genetic differentiation. Domestic populations showed a higher proportion of genetic differentiation among populations (F_{ST} = 0.193 – 0.294) than among wild populations (F_{ST} = 0.174 – 0.252). Similarly, Hamrick and Godt (1996) reported that the mean value of genetic differentiation among populations of crop species (domestic) is higher than that of non-crop (wild) species. This pattern of higher F_{ST} values for cultivated populations can be explained by distinct sources of germplasm used in establishing domestic populations with limited exchange of genetic material, resulting in a lower degree of gene flow among cultivated populations and increasing their genetic differences to some extent. The results are supported by the long cultivation history of citron species in the region. Some of the domestic populations are not far from wild habitats; therefore, migration from wild to cultivated populations by natural or artificial mechanisms may be an ongoing process. Abundant

occurrences of wild and primitive relatives of citron, e.g., *C. nana* (Wester) Yu.Tanaka, *C. odorata* (Wester) Tanaka and species under the subgenus *Papeda* in the eastern Himalayan areas (Tanaka 1969), as well as my recent *Citrus* germplasm collection in northeast India indicate their persistence and diversification in the region of origin. Favourable environmental conditions in this area, currently in the 'Indo-Burma biodiversity hot spot' favoured its growth and further spreading to other parts of the world (Tanaka 1969). In a very recent palynological study by Langgut (2014) stated that citron originated in Asia, particularly India and then gradually dispersed to other areas.

The AMOVA results revealed a high level of genetic variation among individuals (47.53% of the total variation) and significantly ($P < 0.001$) low level of variation among populations (24.98%). In most of the citron populations seeds or cuttings of one or a few individuals were brought from the wild population, transferred to and grown in the farmers' home gardens or local agroforestry systems, and maintained for generation after generation. In clonally propagated plants, separation from the wild ancestor during the domestication process reduces the chances of sexual crossing in subsequent populations (McKey et al. 2010, Zohary and Spiegel-Roy 1975). However, in many perennial plant species heterozygosity also maintained through clonal propagation (Petit and Hampe 2006). Clonal propagation methods might have increased the homogeneity at the population level and observed greater individual differences (47.53%) could not be expressed at the population levels. The citron populations showed significant inbreeding coefficients (F_{IS}) ($P < 0.001$ to 0.01), with the single exception of the Namsai wild population.

The indirect estimates of geneflow (Nm) based on population differentiation among populations showed significant variation ($P < 0.001$) and ranged between 0.600 to 1.187. Population differentiation and effective population size corresponded to three different categories of Nm values: high ($Nm \geq 1.000$), intermediate (0.250 – 0.990) and low (0.000 – 0.249) (Slatkin 1981, 1985). One wild population, Tinsukia, and three domesticated populations, Banskandi, Itanagar and Aizawl, displayed relatively high gene flow ($Nm > 1.000$) and in the other populations it was intermediate ($Nm = 0.600 – 0.918$). The relatively high through intermediate levels of gene flow among populations indicates the movement of genetic material among farmers in the region. Genetic distances between wild and domesticated populations are smaller and admixture is more common between sympatric populations of wild and domesticated

populations than between allopatric populations, which is indicative of gene flow between sympatric populations. The presence of a few private alleles (1 – 4) in most of the wild and domestic populations also shows the existence of gene flow among populations (Slatkin 1985). A review by Ellstrand et al. (1999) of thirteen globally important crops including wheat, rice and maize concluded that gene flow among wild and domestic relatives is common and unintentional, and occurs naturally whenever these relatives come into contact with each other. Viard et al. (2004) and Scurrah et al. (2008) reported similar results of gene flow among the wild and domestic annual crop plants beet and potato species through seeds and clonal propagation. Similar results have also been reported for many perennial food plants (Miller and Gross 2011).

Significant ($P < 0.001$ to 0.05) heterozygosity was observed in the allele frequency data under three different mutation models analysed using the BOTTLENECK software. This result indicates no bottleneck event occurred in any of the citron populations of the region. It is possible that slight or past bottleneck effects may have gone undetected. A number of natural citron populations in the region have diminished, due to natural and anthropogenic disturbances and overexploitation. Until now, such disturbances have had no identifiable consequences in terms of overall genetic diversity and effective population size. Citron populations in the region are growing and maintaining their allelic richness without any reduction in genetic diversity through either natural processes or farming methods. Future studies on larger populations and a wider selection of markers and methods may detect bottleneck events that this study did not, which may be helpful in determining whether conservation measures are required.

The STRUCTURE analysis showed a probable shared ancestry between the wild and domestic citron populations, suggesting that gene flow has occurred between these populations. Overall, the STRUCTURE results suggest five subpopulations within the twelve wild and domestic populations. The grouping of individuals into five distinct clusters is also supported by the highest ΔK values, confirming the presence of five genetically distinct groups (Figure 2.3 and Table 2.7). This is further supported by AMOVA, which showed that most of the total variance occurred among individuals (47.53%) and among individuals within populations (27.49%). The overall genetic structure of populations is not entirely represented in the geographical proximity of individuals. A number of individuals from some populations that are not distributed in the same geographic locations, however, a few of them are genetically structured with the other populations of the region having similar genetic characteristics. Although these are located in

isolated locations; these clustering results may be due to unidentified gene flow among the populations. The genetic diversity observed among the wild and domestic populations did not affect the clustering of the species at population levels. Grouping wild and domestic individuals in the same clusters indicates their admixture due to the long history of cultivation in the region and proximity to farmers' lands. The domestic Banskandi population and the wild Tinsukia population showed similarly large amounts of genetic diversity; however, most of the individuals from these two distant populations clustered together (Cluster-1 and 3, Figure 2.4). Such clustering explains the admixture of individuals among the far distant populations and might be due to their long history of exchange of genetic material. Citron individuals may have spread from wild sources, i.e., the site of origin, to other farmer-managed systems through the movement of the people or sharing of seeds. Further, the UPGMA dendrogram (Figure 2.2) discriminated twelve populations into five putative populations from the 219 accessions. However, cluster analysis could not clearly differentiate the wild and domesticated populations. Thus, there has been mixing of wild and domestic populations. The non-significant ($P < 0.05$) relationship between geographic and genetic distances between populations indicates that their genetic differences are independent of corresponding geographical distances.

Conclusion

There is great diversity in the citron germplasm and this may act as baseline for sustainable utilization and conservation of this valuable genetic resource. The Himalayan northeast region of India is believed to be a centre of diversity for the genus *Citrus* and this study supports the hypothesis that the region harbors a high level of genetic diversity in *Citrus medica*. This also supports the views of Vavilov (1951) who stated that generally plant species show high diversity at species and varietal level in their original place of origin and particularly in the regions that harbour a large number of wild relatives of crop plants. A few individual showed mixed ancestry and there were no clear demarcation between the wild and domesticated populations. Further genetic analyses with more markers and wild populations from primary forest may help in clear distinction between true wild and domestic populations. The observed intraspecific genetic variation in the citron germplasm may help in selecting the most diverse populations for further improvement of fruit quality through breeding programmes, for wider acceptance and

commercialization. There exists a vast genetic resource in the genus *Citrus*, but only a very few species or varieties were commercially exploited throughout the world.

Table 2.1. Northeast India *C. medica* populations sampled during the present study.

Population/Locality Number	Source	Habitat	No. of Individuals	Latitude (North) (°. ′. ″)	Longitude (East) (°. ′. ″)	Altitude (meter)
01. Tinsukia Assam	Wild	Secondary forest	20 (CM1-20)	27.29.32.70	95.22.16.62	12
02. Banskandi Assam	Domestic	Home garden	20 (CM21-40)	24.48.43.80	92.54.58.98	31
03. Itanagar Arunachal Pradesh	Domestic	Home garden	20 (CM41-60)	27.06.10.41	93.41.22.32	146
04. Aizawl Mizoram	Domestic	Home garden	20 (CM61-80)	23.43.13.45	92.42.33.46	1036
05. Sairang[1] Mizoram	Wild	Secondary forest	20 (CM81-100)	23.48.30.29	92.39.30.96	197
06. Sairang[2] Mizoram	Domestic	Home garden	20 (CM101-120)	23.48.35.19	92.39.05.12	102
07. Motinagar[1] Assam	Domestic	Home garden	20 (CM121-140)	24.38.38.52	92.57.51.98	35
08. Motinagar[2] Assam	Wild	Secondary forest	20 (CM141-160)	24.38.38.24	92.57.50.54	35
09. Lakhipur Assam	Domestic	Home garden	20 (CM161-180)	24.47.33.74	93.00.23.13	31
10. Sonai Assam	Domestic	Home garden	20 (CM181-200)	24.44.02.63	92.53.29.43	27
11. Neairgram Assam	Domestic	Home garden	15 (CM201-215)	24.45.51.24	92.50.38.21	28
12. Namsai Arunachal Pradesh	Wild	Secondary forest	04 (CM216-219)	27.40.06.48	95.51.35.13	149

Values in parentheses are the accession numbers.

Table 2.2. Microsatellite SSR loci used in the study.

Locus	Repeat motifs	Annealing temp. (°C)	Primer sequence 5′-3′	Reference
cAGG9	AGG	50	F-AATGCTGAAGATAATCCGCG R-TGCCTTGCTCTCCACTCC	Barkley et al. 2009
CCTO1	CCT	50	F-TCAACACCTCGAACAGAAGG R-CCCACATGCTAGCACAAAGA	Barkley et al. 2006, 2009
GT03	GT	50	F-GCCTTCTTGATTTACCGGAC R-TGCTCCGAACTTCATCATTG	Barkley et al. 2006, 2009
CiBE3298	(AG)15	55	F-TTCTCCTCCACTACACAACAC R-CTTGAATCCCATTTCCAAC	Ollitrault et al. 2010
CiBE3936	(TC)16	55	F-GTAATGATAGCCGTTGGTCTT R-TATGAGATGCCTTGTATTGCT	Ollitrault et al. 2010
CiBE4796	(AG)10	55	F-GATGAGAACGCTGATGCT R-TTCAACCACACTGACGATAA	Ollitrault et al. 2010
CiBE0753	(AAT)13	55	F-TCTCCTTGCCATTATTTATTT R-CAGTTCTCAGTTGCCCGA	Ollitrault et al. 2010

Table 2.3. Diversity statistics of the five polymorphic SSR loci used among 219 *Citrus medica* individuals. Statistics include number of alleles (Na), polymorphic information content (PIC), effective number of alleles (Ne), observed (Ho) and expected (He) heterozygosity, Nei's standard genetic distance (D_S), local inbreeding coefficient (F_{IS}), overall inbreeding coefficient (F_{IT}), genetic differentiation (F_{ST}) and gene flow (Nm).

Locus	Na	PIC	Ne	Ho	He	D_S	F_{IS}	F_{IT}	F_{ST}	Nm
GT03	12	0.773	4.91	0.369	0.798	0.796	0.373	0.556	0.292	0.606
CiBE3298	9	0.752	4.67	0.438	0.788	0.786	0.281	0.438	0.219	0.891
CiBE3936	20	0.829	6.25	0.532	0.842	0.84	0.266	0.375	0.149	1.426
CiBE4796	8	0.761	4.78	0.379	0.793	0.791	0.204	0.461	0.323	0.522
CiBE0753	18	0.694	3.66	0.196	0.728	0.727	0.548	0.725	0.391	0.387
Mean		0.762	4.85	0.383	0.79	0.788	0.334	0.511	0.275	0.767
±SD		0.048	5.36	0.123	0.041	0.04	0.133	0.136	0.093	0.184

Table 2.4. Diversity statistics by *C. medica* population. Statistics include allelic richness (A_R), number of private alleles (A_P), mean number of alleles (MNA), polymorphic information content (PIC), observed (Ho) and expected (He) heterozygosity, genetic differentiation (F_{ST} = average of pairwise F_{ST}), local inbreeding coefficient (F_{IS} = 1 – Ho/He) and gene flow (Nm = (1 – F_{ST})/4F_{ST}).

Population	A_R	A_P	MNA	PIC	Ho	He	F_{ST}	F_{IS}	Nm
01	3.83 ±0.99	2	5.60 ±2.30	0.672	0.220 ±0.160	0.733 ±0.093	0.174 ±0.092	0.705***	1.187
02	3.76 ±1.08	2	7.20 ±2.95	0.640	0.540 ±0.251	0.706 ±0.107	0.193 ±0.099	0.239***	1.045
03	3.43 ±1.16	4	6.00 ±3.08	0.583	0.390 ±0.249	0.658 ±0.127	0.199 ±0.117	0.413**	1.006
04	3.19 ±1.01	2	6.00 ±2.35	0.538	0.470 ±0.299	0.603 ±0.202	0.217 ±0.119	0.224***	0.902
05	3.06 ±0.65	2	5.80 ±2.05	0.505	0.400 ±0.158	0.555 ±0.158	0.236 ±0.132	0.285***	0.809
06	2.48 ±0.68	1	4.00 ±1.41	0.389	0.330 ±0.279	0.438 ±0.217	0.294 ±0.177	0.252***	0.600
07	3.26 ±1.15	3	5.40 ±2.61	0.539	0.269 ±0.213	0.600 ±0.165	0.218 ±0.128	0.559***	0.902
08	3.11 ±0.51	-	4.40 ±0.55	0.552	0.372 ±0.333	0.622 ±0.117	0.252 ±0.106	0.399**	0.742
09	2.81 ±0.85	-	4.80 ±1.92	0.458	0.410 ±0.185	0.512 ±0.188	0.268 ±0.132	0.204**	0.683
10	3.01 ±0.98	-	4.40 ±1.52	0.515	0.360 ±0.225	0.580 ±0.223	0.214 ±0.126	0.385**	0.918
11	2.83 ±0.70	-	3.80 ±1.48	0.507	0.453 ±0.321	0.604 ±0.094	0.258 ±0.129	0.256**	0.719
12	2.60 ±0.54	-	2.60 ±0.55	0.375	0.450 ±0.326	0.500 ±0.152	0.249 ±0.146	0.115^{ns}	0.754

±: standard deviation.
Significance levels: **P < 0.01, ***P < 0.001; ns: non-significant.

(1) Tinsukia-Assam, (2) Banskandi-Assam, (3) Itanagar-A.P., (4) Aizawl-Mizoram, (5) Sairang[1]-Mizoram, (6) Sairang[2]-Mizoram, (7) Motinagar[1]-Assam, (8) Motinagar[2]-Assam, (9) Lakhipur-Assam, (10) Sonai-Assam, (11) Neairgram-Assam, (12) Namsai-A.P.

Table 2.5. Pairwise genetic differentiation (F_{ST}) (below the diagonal) and Nei's standard genetic distance (D_S) (above the diagonal) among the twelve *C. medica* populations.

	1	2	3	4	5	6	7	8	9	10	11	12
1	-	0.202	0.355	0.391	0.797	0.801	0.689	1.597	1.174	0.641	0.558	0.577
2	0.048^{ns}	-	0.324	0.349	0.886	0.828	0.712	1.375	1.007	0.806	1.138	0.841
3	0.103***	0.101***	-	0.078	0.600	0.527	0.467	1.559	0.962	0.968	1.752	1.331
4	0.130***	0.124**	0.022^{ns}	-	0.774	0.703	0.579	1.475	0.773	0.717	1.487	1.013
5	0.230***	0.253***	0.219***	0.276***	-	0.062	0.146	1.011	1.255	1.147	1.386	1.262
6	0.285**	0.299***	0.252***	0.315***	0.041*	-	0.079	1.387	1.768	1.792	2.091	1.952
7	0.192***	0.206***	0.169***	0.217***	0.067***	0.058**	-	1.174	1.236	1.297	1.893	1.640
8	0.273***	0.271***	0.305***	0.325***	0.305***	0.400***	0.297***	-	0.207	0.483	0.964	0.751
9	0.295***	0.288***	0.302***	0.295***	0.382***	0.477***	0.355***	0.112***	-	0.185	0.953	0.570
10	0.194***	0.231***	0.272***	0.255***	0.339***	0.445***	0.328***	0.194***	0.108***	-	0.337	0.156
11	0.165***	0.257***	0.322***	0.335***	0.351***	0.452***	0.355***	0.276***	0.326***	0.151***	-	0.249
12	0.174**	0.234***	0.317***	0.314***	0.373***	0.503***	0.375***	0.263***	0.275***	0.055*	0.101*	-

Significance levels: *P < 0.05; **P < 0.01, ***P < 0.001; ns: non-significant.

(1) Tinsukia-Assam, (2) Banskandi-Assam, (3) Itanagar-A.P., (4) Aizawl-Mizoram, (5) Sairang^{1}-Mizoram, (6) Sairang^{2}-Mizoram, (7) Motinagar^{1}-Assam, (8) Motinagar^{2}-Assam, (9) Lakhipur-Assam, (10) Sonai-Assam, (11) Neairgram-Assam, (12) Namsai-A.P.

Table 2.6. Summary of analysis of molecular variance (AMOVA) for twelve populations and 219 individuals.

Source of variation	DF	Sum of squares	Variance components	Percentage of variation (%)	Fixation indices	P value
Among populations	11	222.886	0.501	24.98	F_{ST} = 0.249	0.001
Among individuals within populations	207	426.092	0.552	27.49	F_{IS} = 0.366	0.001
Within individuals	219	209.00	0.950	47.53	F_{IT} = 0.524	0.001

Table 2.7. Proportion of ancestry of each population in each of the gene pools as defined using the model-based clustering method from Pritchard et al. (2000).

	Proportion of individuals in each gene pool (%)											
Populations/ Clusters	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Cluster 1	-	-	-	-	34	36	30	-	-	-	-	-
Cluster 2	24	26	26	24	-	-	-	-	-	-	-	-
Cluster 3	-	6	-	-	-	-	6	36	36	16	-	-
Cluster 4	24	18	26	29	3	-	-	-	-	-	-	-
Cluster 5	4	-	-	-	-	-	-	-	-	19	38.5	38.5

(1) Tinsukia-Assam, (2) Banskandi-Assam, (3) Itanagar-A.P., (4) Aizawl-Mizoram, (5) Sairang[1]-Mizoram, (6) Sairang[2]-Mizoram, (7) Motinagar[1]-Assam, (8) Motinagar[2]-Assam, (9) Lakhipur-Assam, (10) Sonai-Assam, (11) Neairgram-Assam, (12) Namsai-A.P.

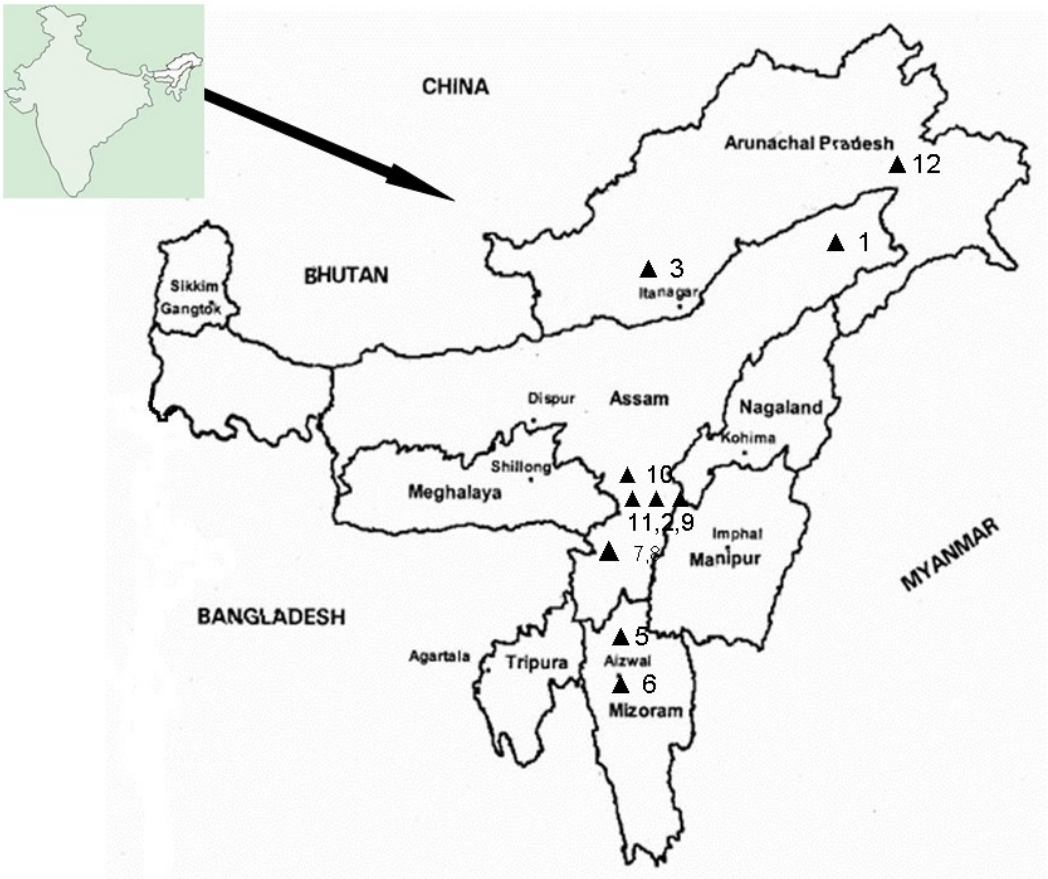


Figure 2.1. Sampling sites of *C. medica* populations in Northeast India. Characteristics of these populations are provided in Table 2.1.

(1) Tinsukia-Assam, (2) Banskandi-Assam, (3) Itanagar-A.P., (4) Aizawl-Mizoram, (5) Sairang[1]-Mizoram, (6) Sairang[2]-Mizoram, (7) Motinagar[1]-Assam, (8) Motinagar[2]-Assam, (9) Lakhipur-Assam, (10) Sonai-Assam, (11) Neairgram-Assam, (12) Namsai-A.P.

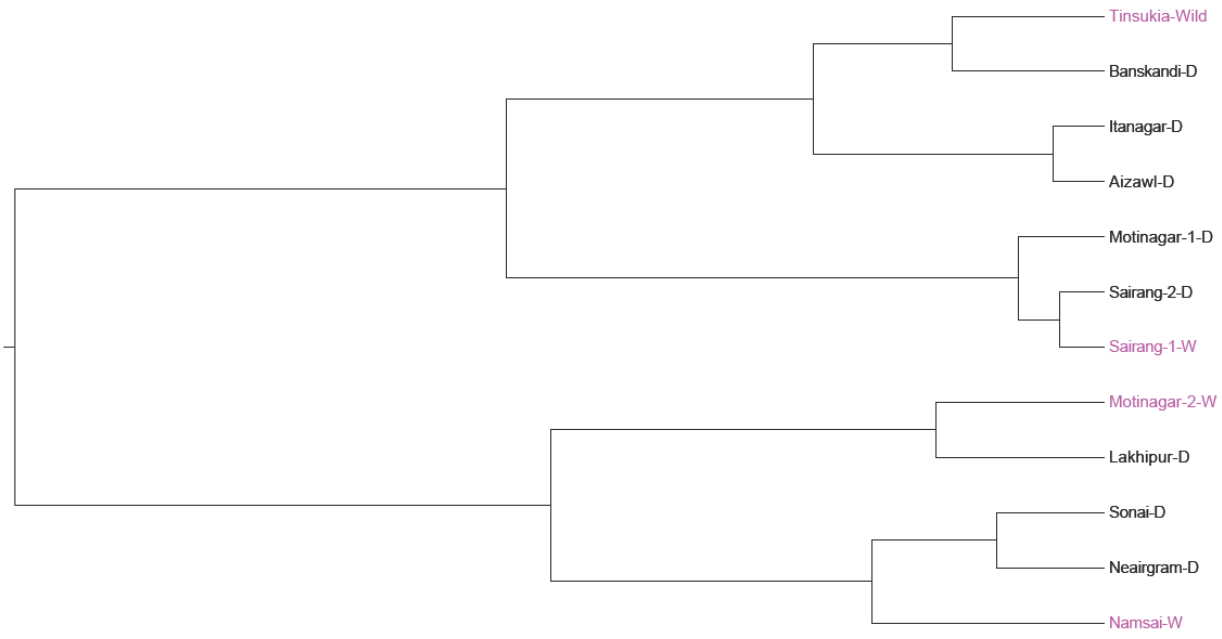


Figure 2.2. UPGMA dendrogram representing genetic relationships among the twelve *C. medica* populations, constructed using Nei's genetic distance calculated from allele frequencies observed at five microsatellite loci.

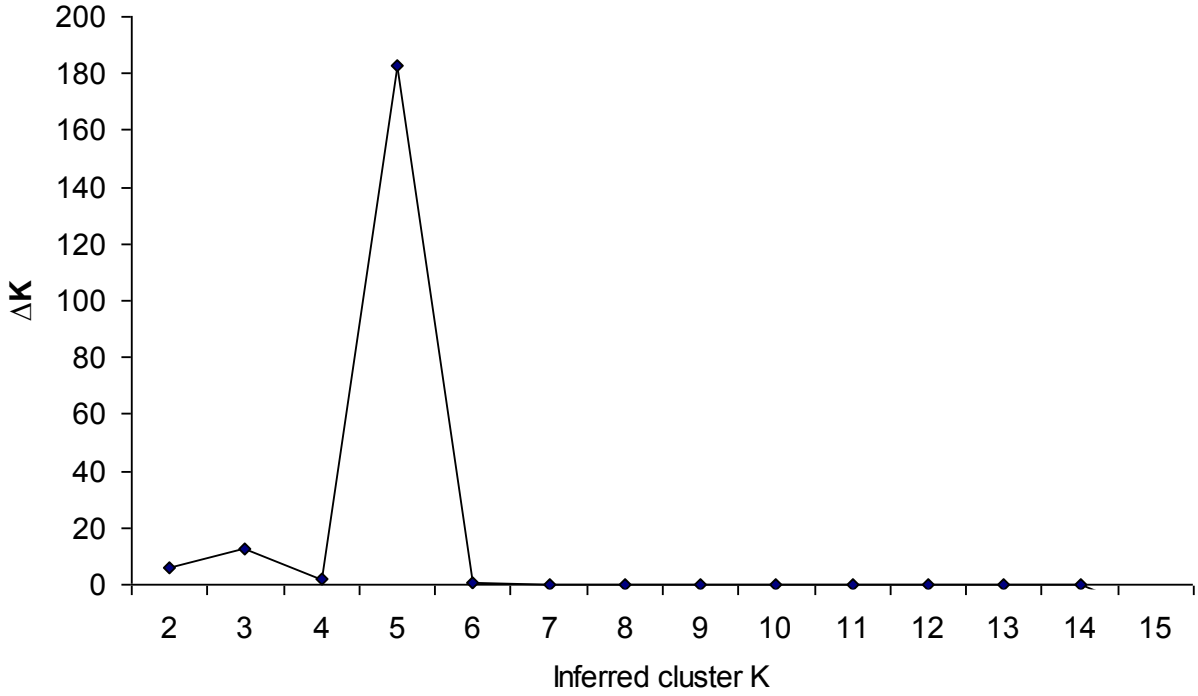


Figure 2.3. The number of inferred clusters K based on mean log likelihood probability values (ΔK) (K= 1-15) obtained from STRUCTURE analysis. The most likely value for putative population identified at K=5.

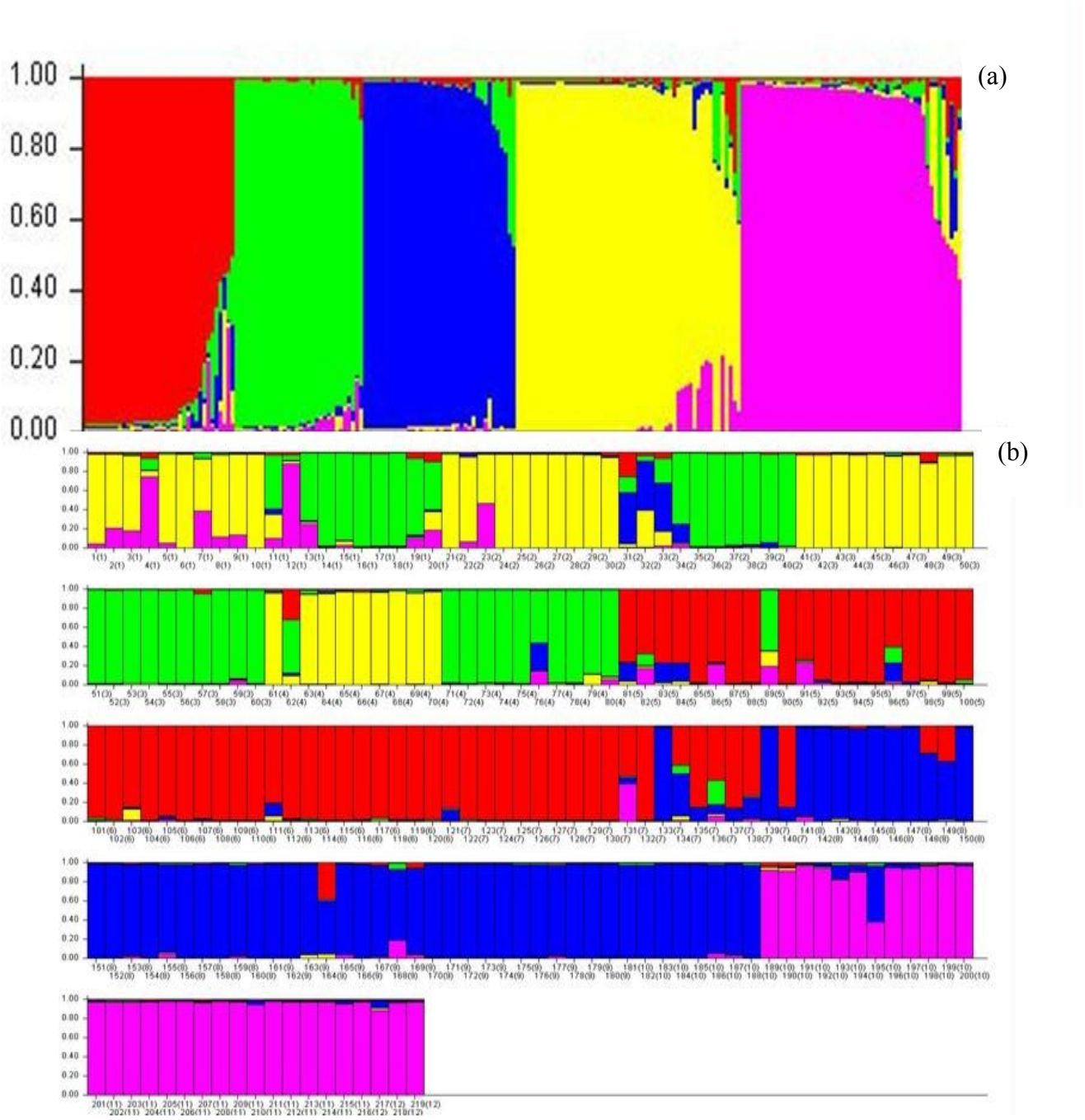


Figure 2.4. Population assignments by STRUCTURE. (a) Clustering of populations at $K = 5$. The X-axis shows population numbers as defined in Table 2.1; the Y-axis shows the proportion of alleles derived from each population. Accession assignments are as follows (population numbers and proportion): **Cluster1**: #5 (34%), #6 (36%) and #7 (30%) **Cluster2**: #1 (24%), #2 & #3 (26% each) and #4 (24%); **Cluster3**: #2 & #7 (6% each), #8 & #9 (36% each) and #10 (16%). **Cluster4**: #1 (24%), #2 (18%), #3 (26%), #4 (29%) and #5 (3%); and **Cluster5**: #1 (4%), #10 (19%) and #11 & #12 (38.5% each) (b) Assignment of 219 individual (population number in brackets) *C. medica* accessions to into five distinct clusters. The Y-axis shows the proportion of alleles derived from each individual. Individuals of the same color belong to the same cluster. An individual with more than one color shares a percentage of its among multiple clusters, according to the admixture proportions.

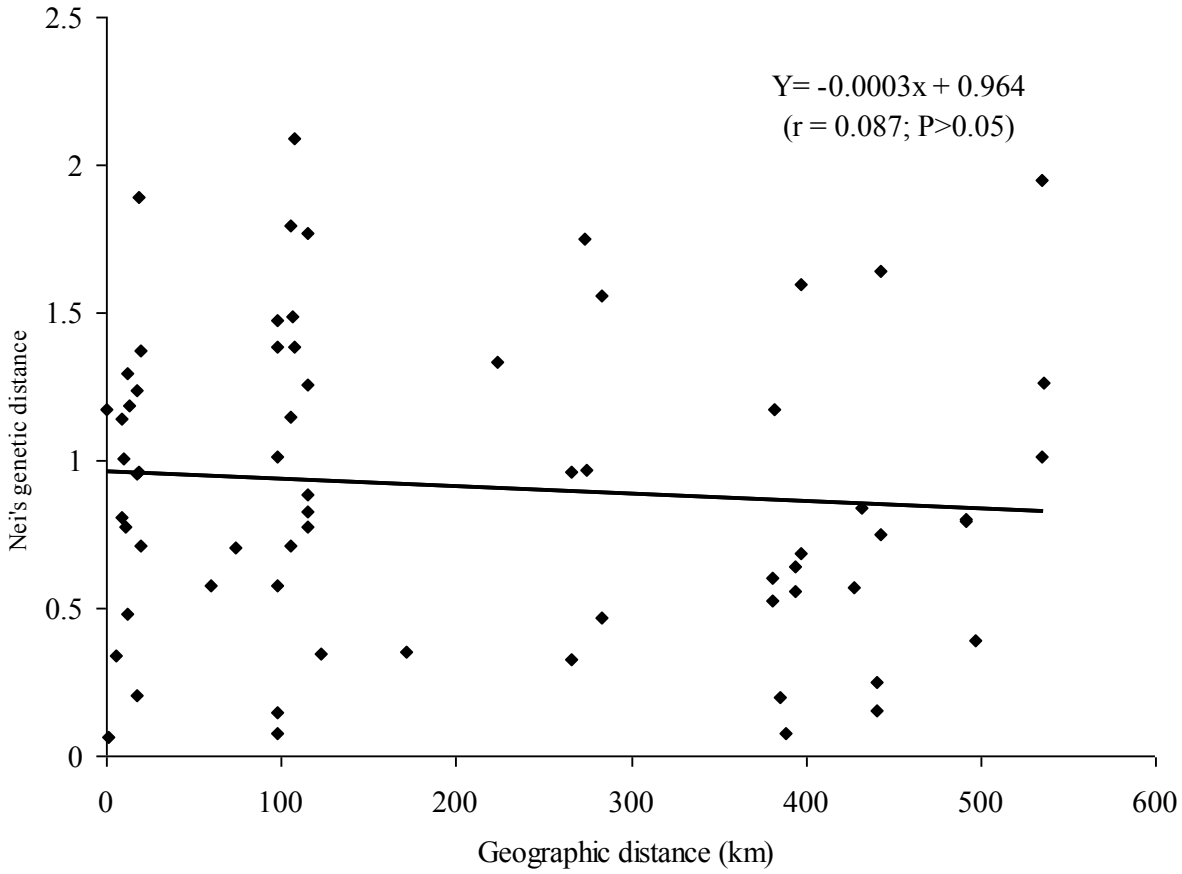


Figure 2.5. Relationship between geographic distance and Nei's genetic distance among the twelve populations of wild and domestic *C. medica*.

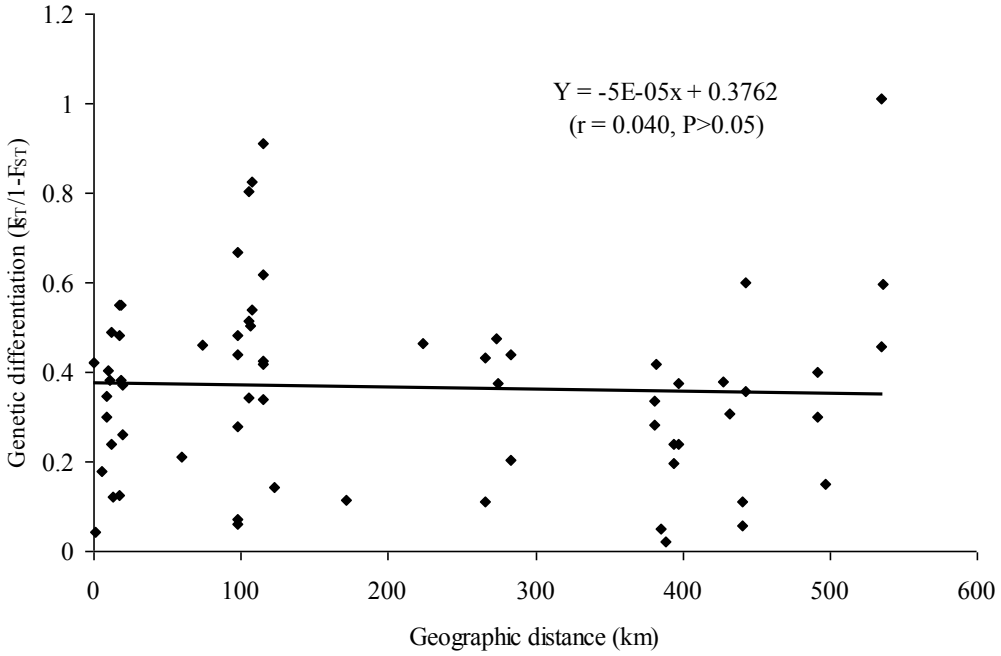


Figure 2.6. Relationship between geographic distance and genetic differentiation [$F_{ST}/(1 - F_{ST})$] among the twelve populations of wild and domestic *C. medica*. F_{ST} was calculated according to Weir and Cockerham (1984).

Chapter 3: Plant diversity in the indigenous home gardens in Mizoram, Northeast India.

Abstract

The eastern Himalayan region of northeast India is well known for its traditional home gardens, which are considered to play important roles in the maintenance of livelihoods of indigenous communities and conservation of biological diversity. I studied 90 home gardens located in 6 villages in Aizawl and Serchhip districts in Mizoram, northeast India to determine a) plant species composition in home gardens, b) correlation between home garden size and plant species diversity, c) common uses of plants in home hardens and d) the role of home gardens in conservation of plant genetic resources. A total of 333 plant species (133 trees, 92 shrubs and 108 herbs) belonging to 122 families with an average of 78 species per home garden were recorded. The size of home gardens ranged between 0.10 – 0.60 ha and showed significant ($P<0.001$) positive correlation between the garden size and plant species diversity. The species diversity index for trees, shrubs and herbs was 4.76, 4.39 and 4.58 respectively. The species similarity within each life-form was high with 50% for trees, 38% for shrubs and 49% for herbs. Plant species in the home gardens could be grouped into 11 major use categories and majority of plants were of medicinal or multiple use categories. These home gardens are reservoirs of plant genetic resources and play a vital role in sustaining the livelihood of local inhabitants.

Key words: Diversity, homegarden, indigenous, northeast India, plant, people, tribe.

Introduction

Home gardens are considered as one of the oldest subsistence farming systems practiced by rural communities in many parts of the world, and can include multi-layer systems of trees, shrubs and herbs around homesteads (Idohoua et al. 2014, Kabir and Webb 2008, Kumar and Nair 2004, Salako et al. 2014). An estimated 15–36% of residential land in the UK, India, Africa, and China is occupied by home gardens (Baudry et al.1999, Cilliers et al. 2013, Huai et al. 2011, Jaganmohan et al. 2012, Loram et al. 2011). Home gardens are generally multifunctional and play key roles in providing ecosystem services and numerous benefits for sustaining the livelihood of local inhabitants (Calvet-Mir et al. 2012, Clarke et al. 2014, Galluzzi et al. 2010, Reyes-Garcia et al. 2012, Schupp and Sharp 2012). These ecosystems are increasingly becoming

the focus of human-natural systems research with increased demand for precise quantification of plant species abundance, community diversity and ecosystem functioning (Bernholt et al. 2009, Jaganmohan et al. 2012, Kabir and Webb 2009). Home gardens are important as a means of maintaining plant genetic resources (Agelet et al. 2000, Sunwar et al. 2006), as potential hotspots of agricultural biodiversity (Galluzzi et al. 2010, Kumar and Nair 2004, Taylor and Lovell 2014), and as natural resources for alleviating poverty (Fraser et al. 2011, Reyes-Garcia et al. 2012, Salako et al. 2014, Shackleton et al. 2008). In addition, they represent a viable solution for biodiversity conservation as *ex-situ* and *in-situ* conservation areas for rare and threatened species and may play a significant role in sustaining regional biodiversity (Kabir and Webb 2009, Rico-Gray et al. 1990, Roy et al. 2013).

The home gardens in the eastern Himalayan region of northeast India are known to have played an important role in the domestication of many plants and traditional crop varieties. Wild relatives of several crops and other commercially used plants are conserved in these home gardens (Galluzzi et al. 2010, Hammer et al. 1999), and serve as an invaluable genetic resource for breeding and improvement of crops and horticultural plants. Home gardening in the region is believed to have evolved with the local practice of jhum agriculture, the slashing and burning of the forest at the village outskirts. Jhum is a labour intensive cultivation system which requires minimal capital and nutrient input. Its practice results in the loss of topsoil and nutrients, leading to habitat degradation. Farmers of the region have recognized the adverse impacts of jhum agriculture and consequently developed a preference for home gardening over jhum for the maintenance of crop diversity, household food security, nutrition and subsistence income generation. Since most of the landscapes in the region are steep slopes, home gardening land use system is a more suitable approach for minimizing soil erosion, and is easily adaptable for ecological rehabilitation and an agricultural productivity increase in marginal lands (Sahoo 2007). The home garden systems in the region resemble the agroforestry systems practiced in many parts of the world, and serve as an important source of food, timber, fodder, fruits, and herbal medicine for local inhabitants.

The objectives of the present study are to gain insights into plant diversity and their importance in conservation of plant genetic resources through utilization in northeast India. I use home gardens in the Mizoram province as representative home gardens of northeast India, which are maintained by tribal communities of the region and often located next to their primary

dwellings. The specific questions addressed are, 1) What is the plant species composition in home gardens? 2) Is there a correlation between home garden size and plant species diversity? and 3) What are the uses of plants in home gardens and what is their relevance in conservation of plant genetic resources?

Materials and methods

Study site

This study was conducted in six villages located in Aizawl and Serchhip districts in Mizoram, northeast India. Mizoram or 'land of the hill people' is located within the Indo-Burma biodiversity hotspot at the far end of the Himalayan mountain range. The total land area of Mizoram is 21,081 km^2 and approximately 91% of the area is under forest cover. It lies between 92°15' and 93°26'E longitude and between 21°58' and 24°35'N latitude, with an altitudinal range of 21 to 2157 m above the mean sea level. Mizoram is surrounded by three states (Assam, Manipur and Tripura) and shares international boundaries with Bangladesh on the west and Myanmar on the east and south. The climate of the area is moist tropical to sub-tropical and the temperature ranges between 20˚- 30˚C and 7˚- 18˚C during summer and winter respectively and receives an annual rainfall of 2000 - 3200 mm, with high rainfall during the wet summer months of April to September and low rainfall in the dry and cold months of October to March. The topography of the study sites is highly undulated, and most agricultural practices are performed in the upland areas. The indigenous tribal communities in Mizoram practice home gardening for their livelihood. I studied 90 indigenous home gardens located in six selected villages in Mizoram. Three villages (Selesih, Sairang and Thingsulthliah) in Aizawl district, while the other three villages (Serchhip, Keitum and Chaitlang) in Serchhip district (Figure 3.1). Data describing the extent and elevation of the areas encompassing the home gardens in each village are given in Table 3.1.

Data collection and analysis

I requested voluntary participation from home garden owners and field surveys were conducted during March to October 2008. After a preliminary survey of 35 home gardens (about 23% of the home gardens in each village), 15 gardens in each village (a total of 90 home gardens) were chosen for detailed study. Home garden owners provided information of the social customs surrounding home gardening practices, technical details such as tool and fertilizer use, as well as

watering techniques. Information on plant species composition and uses of each species was collected through direct observation and discussion with the farmers. I measured the area of total plant cover in each garden after excluding the dwelling area. Data collection was conducted in each home garden during the peak sowing and growing season (April-June) and harvesting (June-September) season of the year. In each garden, species composition was enumerated by randomly placing five 10m x 10m quadrats for trees. Within each of these quadrats, another 5m x 5m quadrat for shrubs and a 1m x 1m quadrat for herbs were established. Species richness was calculated as the number of species encountered in all quadrats grouped by habit forms (trees, shrubs, herbs and climbers). The local names of all plants were recorded, and each was identified to species level in consultation with the herbarium at the Mizoram University and taxonomists at regional herbaria of the Botanical Survey of India in Shillong, Meghalaya. Plants with multiple uses were classified by main use, into categories including fencing, food, fuel-wood, fruits, medicinal, ornamental, roofs, timber, trade and spice. Plant species with several uses other than the above mentioned categories were included in the "other" category, which includes a variety of uses including shade, timber, fiber, and soil fertility.

The plants in each quadrat were counted, and a t-test was performed to identify the significant differences in the mean values of species richness in six different sites. The diversity and abundance of plants in home gardens between villages were examined using analyses of variance (ANOVA) (SPSS 16.0) at two scales, garden and village. Garden level plant diversity and abundance were compared within the home garden in each village and overall among six villages. The data collected in the quadrats were used to determine the frequency, density and dominance, following Phillips (1959). Species diversity was calculated using the Shannon-Weaver index of diversity: $H' = -\sum \{(n_i / N) \ln(n_i / N)\}$, where n_i = importance value index (IVI) of a species, N = total IVI of the community (i.e., 300). The importance value index (IVI) was calculated following Salako et al. (2014) to analyze the importance of each species in each home garden and in each village. For a species i of a given home garden, the IVI was computed as IVIi = RDi + RFi + RDoi, where, RDi is the relative density of the species i; $RD_i = N_i / \sum_{i=1}^{P} Ni$ (where P is the total number of species recorded in the each village and Ni is the mean density of the species i in that village). RFi is the relative frequency of the species i: $RF_i = f_i / \sum_{i=1}^{P} fi$, where $f_i = \frac{ji}{k}$ (f_i frequency of the species i, j_i the number of home gardens at which the species i was counted, and k is the total number of home gardens (k=90). RDo is the relative dominance of the species i:

$RDo_i = Do/\sum_{i=1}^{P} Doi$ (Doi, is the mean dominance of the species i in the home gardens). The IVI-value is an overall estimation of the level of importance of a species in the home gardens in a village.

The dominance index (Simpson 1949) of the community was calculated as: $C = \sum \{(n_i/N)^2\}$, where n_i and N are same as for Shannon's index. Pielou's (1966) evenness index was calculated as: $e = H' / \log(S)$, where H'= the Shannon-Weaver index of diversity, and S = total number of species. Sorensen's similarity index (Sorensen 1948) was calculated as, [2C / (A + B)] x 100], where A and B are the total species content (trees, shrubs, herbs and climbers) in stand A and B respectively, while C is the number of species common to both stands.

Results

The physical location and sociological characteristics of the study villages are given in Table 3.1. The human population densities of the two districts Aizawl and Serchhip, are 113 and 46 persons per square kilometer respectively. The population density in these districts represents the lowest in the country with the majority of individuals living in rural areas (Census of India 2011). The selected home gardens in Aizawl and Serchhip districts are located approximately 40 and 100 km respectively from Aizawl, the state capital city of Mizoram. The household sizes of the study area varied between 5-8 people with 2-3 income earning members in the family. The average number of households in six villages was 663 with the highest number of households (1051) in Sairang village of Aizawl district and the lowest (308) in Chhiahtlang village of Serchhip district. Among the 3981 households in six villages, only 441 households (11%) had home gardens (Table 3.1). Although random and small shops are found in all villages, most of the produce from home gardens is transported for sale at weekly (Saturday) market in the district capital. These home gardens are mostly rainwater fed, and water harvesting technology in the villages is almost non-existent due to steep slopes coupled with poor water-holding capacity of the soil. Almost all gardeners in the study areas use traditional tools such as khurpi (hand-held iron hoe), shovel, spade, sickle, knife and other traditional practices of manual weeding and pest control. Soil fertility of the home gardens is maintained through natural means using organic manures produced at home through composting leftover crops and other household organic materials in concrete tanks. A few gardeners use manure from their small scale pig and poultry farms. In general, all adult family members contribute equal labor to the overall maintenance and

management of gardens. In general, men select cash crops, trees and fruit species and obtain and sow seed materials, while women mainly grow and manage vegetables, spices, medicinal plants, and harvest and market superfluous crops. Most gardening activities are performed under the supervision of elderly family members with traditional knowledge for garden.

A wide variation in home garden sizes was observed and the area of each home garden ranged between 1421- 6027 m^2 ± S.E. 330 in Sairang, 1047 - 5462 m^2 ± S.E. 295 in Selesih, 1064 - 4321 m^2 ± S.E. 223 in Thingsulthliah, 1127 - 4867 m^2 ± S.E. 240 in Serchhip, 1245 - 3891 m^2 ± S.E. 207 in Keitum and 1098 - 3245 m^2 ± S.E. 179 in Chhitahlang. In general, home gardens located in the Serchhip district are relatively smaller ($P<0.001$; $t(44)=5.085$) then the home gardens of Aizawl district (Table 3.1).

Species richness and diversity

A total of 122 plant families were recorded in the present study (Table 3.2). The most common plant families (Figure 3.2) were Fabacece, Rutaceae, Zingiberaceae, Lamiaceae and Solanaceae, Euphorbiaceae, Asteraceae and Cucurbitaceae, which contained 30, 18, 14, 13 each, 11 and 10 species respectively. The highest numbers of food plants were from the family Fabaceae and the family Rutaceae contributed maximum number of fruits and medicinal plants. The most abundant tree species included *Areca cathechu, Artocarpus heterophyllus, Mangifera indica*, *Parkia timoriana* and several *Citrus* species. The most dominant shrubs species were *Amaranthus viridis*, *Cajanus cajan*, *Calamus erectus*, *Capsicum annum*, *Carica papaya*, *Clerodendrum colebrookianum*, *Hibiscus macrophyllus*, *Murraya koenigii* and large number of *Musa* and *Solanum* species. The dominant herbaceous and climber species were *Ageratum conyzoides* a few number of *Allium, Brassica, Cucurbita* species etc (Table 3.2).

The number of plant species in each home garden ranged from 36 to 167, with an average of 78 species in each garden suggesting a high inter-garden variation in overall species composition and richness. The importance value indices curve based on 90 gardens sampled in the area did not reach an asymptote, indicating that home gardens in the region may contain more number of plant species than I was able to identify in this study. The lack of an asymptote further indicates that multiple plant species share dominance in the overall structural composition of the home gardens (Figure 3.3). The occurrences of species in the studied garden are not normally distributed. Most (85%) of the species were represented in a broad range of frequency (5 - 40%) classes and only a few species (15%) in high frequency classes (41 - 75%) (Figure 3.4)

indicating occurrence of large number of species in those gardens. A total of 333 plant species were found in the 90 home gardens studied. Of these, trees were the most abundant, with 133 species (40%), followed by 108 (32%) species of herbs and 92 (28%) species of shrubs. Overall, 96 genera of trees belonging to 52 families, 59 genera of shrubs belonging to 36 families and 59 genera of herbaceous plants belonging to 52 families were found (Table 3.3). Species richness varied significantly [Mean=1124.55, SD=1292.65; t(44)=5.83, P=0.001] among villages with highest number of species in the gardens of Sairang village (Aizawl district) followed by the gardens in Serchhip village (Serchhip district) and the lowest was in Thingsulthliah village (Aizawl district) (Table 3.3).

Species diversity indices for trees, shrubs and herbs varied significantly [Mean=4.11, SD=0.288; t(17)=60.41, P=0.001] within gardens in different villages. Overall, the tree species diversity was higher (F=6.84, P=0.01; ANOVA) than the herb and shrub species diversity. Evenness index for trees, shrubs and herbs also showed a trend similar to the diversity index values and varied slightly within home gardens ($P<0.05$). The evenness values were higher in the small home gardens in Selesih and lower in the large gardens in Sairang villages of Aizawl district (Table 3.3). The similarity indices of trees, shrubs and herbs were high (91%) between gardens in Seleisih and Sairang followed by Thingsulthliah and Sairang (88%) of Aizawl district. The lowest similarity values of plant species (68%) were observed among the gardens of Serchhip and Selisih of Serchhip and Aizawl districts. The tree species similarity indices showed significant variation [Mean=70.76, SD=6.33; t(14)=43.23, P=0.001] among gardens with highest similarity (87%) between gardens in Seleisih and Sairang and the lowest similarity between Keitum and Chhiahtlang (51%) (Table 3.4). In general, 66 trees (50%), 35 shrubs (38%) and 53 herb (49%) species were common to all home gardens.

Stratification and functional diversity

All home gardens were composed of a mixture of herb, shrub and tree species forming multiple layers of different plant species with three to four distinct vertical stratifications. The uppermost canopy consisted of trees and therefore was a perennial layer. Species commonly found in this layer included *Alstonia scholaris*, *A. cathechu*, *Bombax ceiba*, *Borassus flabellifer*, *Canarium bengalense*, *Castanopsis indica*, *Grevillea robusta*, *Mesua ferrea*, *P. timoriana*, *Quercus griffithii*, *Sterculia villosa* and *Tectona grandis*. This layer was followed by individuals of *Aegle marmelos*, *A. heterophyllus*, *Dillenia indica*, *Elaeocarpus floribundus*, *Lagerstroemia speciosa*,

M. indica, *Oroxylum indicum*, *Psidium guajava*, *Schima wallichii* and *Tamarindus indica*. Annual and perennial plants are found immediately below this layer. The most common and important species were *Acacia nilotica*, *Albizia procera*, *Averrhoa carombola*, *Bauhinia variegata*, *C. aurantifolia*, *C. grandis*, *C. macroptera*, *C. reticulata*, *C. medica*, *C. rugulosa*, *Persea americana*, *Phyllanthus acidus*, *Ziziphus jujuba*. The third storey consisted of a variety of shrub species including a large number of perennial medicinal and crop plants including *A. viridis*, *C. cajan*, *C. colebrookianum*, *Chenopodium album*, *Ocimum sanctum*, *Hibiscus sabdariffa*, *Manihot esculenta*, *Solanum khasiana*, *S. melongena* and also climbing crops like *Sechium edule*, *Piper betle*, *Glycine max*, *Momordica charantia*, *Dolichos tetragonolobus*, *Vitis vinifera* and a variety of *Musa* species. The lowest ground storey consisted of species that were 20 cm or less in height, such as *A. conyzoides*, *Allium cepa*, *A. hookeri*, a few species and varieties of *Brassica*, *Colocasia* and *Cucurbita* species, *Curcuma longa*, *Ipomoea batatas* and *Zingiber officinale.*

Based on uses, the overall plant species were broadly categorized into eleven groups (Figure 3.5). The species under different use category were well represented in each surveyed garden. Under different use category, medicinally important plants had the major (33%) constituents in home gardens followed by food plants (16%), fruits species (10%), ornamental (6%), timber (5%) and fuel wood (2%), trade and spice plants (2%) and 1% each of roofing and fencing category, with a large proportion of plants (22%) having multiple uses (Figure 3.5).

Discussion

The mountainous region of Mizoram in the Indo-Burma biodiversity hotspot is home to many indigenous communities with unique life styles who are accustomed to live in steep slopes using locally available natural resources. Increased population and urbanization in many parts of India lead to reduction in forest cover. However, the mountain areas of Mizoram have not experienced extensive deforestation except for shifting cultivation, a prevalent system of cultivation as a main source of livelihood of indigenous communities. The indigenous tribal communities of the region experienced and realized the adverse effects of slash and burn shifting cultivation and majority of inhabitants accepted home gardening as an alternative and sustainable hill farming system. These home gardens are the only type of agricultural land in the region and the source of year round supply of food and other daily necessities including medicine, fuel wood, timber and cash

income through the sale of surplus products. The ownership of these gardens passed from one generation to the next and maintained as permanent family gardens, sustaining productivity for many generations without major changes in the composition of plant communities. Other factors including the ban of slash and burn agriculture, low household income, lack of industries, high market prices of essential commodities and food products and poor access to the urban market area also promoted the maintenance of home gardens. In addition, maintenance of large number of species in home gardens provide indirect benefits and ecological services such as habitats to birds, butterflies, and bees. Similar services from home gardens throughout the world has been reported (Calvet-Mir et al. 2012, Clarke et al. 2014, Fernandes and Nair 1986, Idohoua et al. 2014, Kabir and Webb 2009, Mendez et al. 2001, Sunwar et al. 2006).

The size of home gardens in Mizoram ranged between 0.10 – 0.60 ha, which is similar to global average home garden size of 0.10 – 0.50 ha (Brierley 1985, Das and Das 2005, Fernandes and Nair 1986, Kumar et al. 1994). The plant diversity and home garden productivity is largely a function of the garden size and according to the farmers in Mizoram and these observations suggest that large home gardens provide sufficient products for the own consumption of households as well as significant financial gains through sale of extra products in local markets. This study has shown significant positive correlation ($R=0.820$, $P<0.001$) between the size and total species diversity (Table 3.5). The farmers constrained with land shortage concentrate on fewer species with high usage and allocate more land area for food crops as evident by the significant ($R=0.650$, $P<0.001$) positive correlation between garden size and plants used for food (Table 3.5). This pattern of increasing tree species richness with increasing land holding also reported in other home garden systems (Abebe et al. 2013, Huai et al. 2011, Kumar et al. 1994, Loram et al. 2011, Mendez et al. 2001, Zhang and Jim 2014).

The species composition of these home gardens is similar to general floristic profile reported from other tropical home gardens. Several taxa such as *Allium*, *Annona, Brassica*, *Calamus*, *Citrus*, *Dioscorea*, *Carica*, *Capsicum*, *Curcuma*, *Mangifera, Psidium, Spondias* have been reported in many tropical home gardens in many regions of the world (Albuquerque et al. 2005, Das and Das 2005, Kabir and Webb 2009, Padoch and De Jong 1991, Rico-Gray et al. 1990, Shastri et al. 2002, Sunwar et al. 2006, Wezel and Bender 2003). Representation of over three hundred species in diverse plant families and genera with an average of 78 species per garden highlights the rich biodiversity in gardens (Table 3.3). In general plant richness estimated

in home gardens in Mizoram is relatively higher than plant richness reported in home gardens from other parts of India including Assam in Northeast India (Das and Das 2005), Karnataka (Shastri et al. 2002) and Kerala (Kumar et al. 1994). Several home garden surveys in the other areas of the world (Ahmad and Abood 1990) reported 44 species in Malaysia, 278 species in China (Clarke et al. 2014), 281 in Mexico (Larios et al. 2013), 200 species in Thailand (Makaraphirom 1989) and 62 species in Bangladesh (Roy et al. 2013). High species richness and diverse plant composition provide a wide range of choices of plant material to meet diverse needs of home garden owners. Fernandes and Nair (1986) pointed out that tropical home gardens harbor diversity equivalent to tropical forests. Other studies (Alvarez-Buylla Roces et al. 1989, Michon et al. 1983) also highlighted the importance of home gardens for the maintenance and conservation of plant genetic diversity. The species diversity index for tree, shrub and herb in the present study was 4.76, 4.39 and 4.58 respectively (Table 3.3). The species diversity index values are higher than the corresponding values of home gardens in various parts of the world: 0.50 – 3.30 in Hong Kong (Zhang and Jim 2014), 1.007 – 3.153 in Tehuaccn Valley, Mexico (Larios et al. 2013), 1.9 – 2.7 in Thailand (Gajaseni and Gajaseni 1999), 2.30 – 3.39 in Bangladesh (Roy et al. 2013), 2.43 – 3.84 in up and low lands areas of Mexico (Gliessman 1990a), 3.21 in Karnataka, India (Shastri et al., 2002), 3.55 in Costa Rica (Gliessman 1990b) and 3.93 in Sri Lanka (Kharal 2000). The species diversity index of home gardens in Mizoram are similar to the values (4.03 – 4.42) reported from home gardens in western Nepal (Sunwar et al. 2006). The high diversity values found in those gardens highlights their richness and are related to several factors such as varied geography, favorable microclimates, long history, introduction of species from the nearby forest to fulfill community needs of plant species, exchange and sharing of resources by the communities. Multiple nutritional demands and year round needs of various products also increased the diversity in those home gardens. Dominance index values ranged between 0.164 – 0.373 among the gardens and tree species have lower values then herbs and shrubs (Table 3.3). Overall low dominance indices explain the heterogeneity and richness in species composition with greater dominance of trees followed by herbs and shrubs respectively. The greater evenness values of 0.970 – 0.978 among different plant categories and gardens indicate that greater percentage (ca. 97%) of the species is uniformly distributed in different gardens in the area. In general, high evenness and low dominance values in the gardens confirm that those gardens are not occupied by limited number of species, however, abundant number of

species. Greater species similarities among the gardens of different villages are due to the reason that tribal communities in all the villages are from same ethnic groups. They have almost the similar management and conservation strategies. In general, the household requirements for food, spices and other plant species, traditional agricultural systems, culture and indigenous knowledge are also very much similar among the communities residing in different villages. Some variation arises may be due to individual family species preference, size of home garden, altitude and soil fertility status.

In regard to vertical structure, different species composition and perennial habits of large number of plants make these gardens resemble to tropical forests with multi layered vegetation structure. Smith et al. (2005) stated that different stratifications and dynamic architecture make home gardens a sustainable and resilient ecosystem. Vertical stratification in vegetation makes such system more productive by capturing light sources and uptake of soil nutrients by different root systems. On the other hand, many shade loving crop plants receive optimum environment for their growth and yield. Different climbing crops such as grapes, squash, piper and pumpkin receive physical support from other plants and act as host for a number of epiphytes, such as Orchids. The indigenous tribal communities of the region have developed and learned similar management strategies through generations. Furthermore, similar practices may have evolved through direct observations and cultural experiences through living in association with natural forests for many generations.

The year round and regular services of different plant products are due to combinations of large variety of crops of different habits viz., annual, biennial and perennial. The presence of crops with different functions and habits fulfills the nutritional and financial needs of the farmer. Home garden plants are used for food and fruit production as well as medicinal products. Similar results were observed in other studies (Padoch and De Jong 1991, Rico-Gray et al. 1990). These results are also consistent with findings from other studies that highlighted the importance of home gardens in producing healthy food and economic support to the gardener (Calvet-Mir et al. 2012, Reyes-Garcia et al. 2012, Shackleton et al. 2008). The perennial nature of these home garden and combination of herbaceous vegetables, shrubs and trees form mixed and balanced production system. This might play an important role in ecological sustainability and stability through effective management strategies by the owner of these home gardens. Dietary changes have resulted in increased in the diversity of cultivated vegetable species, including exotic and

improved varieties of species such as cauliflower, broccoli, spinach, radishes etc. Nevertheless, gardeners reported that they continue to grow landraces of their preferred staple food as a preferred choice for traditional dishes. Thus, traditional knowledge associated with the cultivation of indigenous wild crop varieties are maintained (e.g., *A. viridis*, *A. spinosus, C. esculenta*, *C. mannii*, *C. gigantea*, *D. tetragonolobus*, *H. macrophyllus*, *M. esculanta*, *Solanum anguivi, S. khasiana*, *Polygonum convolvulus*, *P. orientale* etc.) and landraces of many crops along with a few domesticated and improved varieties of crops viz., *A. cepa*, *A. sativum*, *Abelmoschus esculentus, Brassica capitata*, *B.rapa*, *C. papaya*, *Coriandrum sativum*, *Daucas carota, Phaseolus vulgaris*, *Raphanus sativa*, *Solanum melongena*, *Vinga mungo*.

Different tree species have been found to be associated with various socio-economic and ecological roles in the site. As an example, a large number of timber species (5%) such as *Artocarpus chama*, *Chukrasia velutina*, *Cinnamomum tamala*, *M. indica, M. ferrea*, *Magnolia champaca, S. villosa*, *S. wallichii* are used for the construction of houses and furniture. Many of these species also serve multiple functions. Species such as *Trema orientalis, Calamus acanthospathus*, *Lantana camara*, *Erythrina arborescens* and *A. nilotica* were planted as living fences between home gardens and to protect crops from wild animals. As per garden owner knowledge and information sharing during the survey a few evergreen and perennial tree species viz., *A. scholaris*, *Azadirachta indica*, *P. timoriana*, *S. wallichii*, *S. villosa* also have a number of ecological importance besides their timber and fuel wood supply. Particularly those ecological services includes shade for the under canopy trees, shrubs and herbs and improved soil fertility through leaf litter decomposition. According to farmer perspectives many annual crops shows better yield when they are in association with a few tree species like *Albizia myriophylla*, *Cassia javanica* subsp. *nodosa*, *Erythrina indica* and *Duabanga grandiflora*. This may be due to better nitrogen fixing abilities of those plants. *P. timoriana*, locally known as 'Jongtra', is found to be common in almost all of the home gardens because of its wide economic and ecological roles in these systems; this species provides good economic return every year through the sale of its long, tender pods as a delicious vegetable throughout the region, particularly among the tribal community. Furthermore, occasionally this plant is harvested for timber and used for making furniture and fulfills other domestic needs. Varieties of *Cucurbita* species locally known as 'Maien', is used for its tender shoot, flower and fruits. 'Iskut'(*S. edule* Jacq. Sw.) is used for its tender shoot and fruits. Taro and yam-like roots (locally called 'Kochu' and 'Dawl'),

representing a few *Colocassia* and *Dioscorea* species used for leaf, petiole, corm and rhizomes, and several *Musa* species known as 'Balah' are harvested as fruits and vegetables. In general, these home gardens are the potential source of different bio-products for the overall and basic need of the practicing families of the hill region.

The high intra-specific diversity observed in many species of different plant families viz., Araceae (6 *Colocasia* species), Musaceae (8 *Musa* species), Polygonaceae (5 *Polygonum* species), Rutaceae (14 *Citrus* species), Solanaceae (10 *Solanum* species) and Zingiberaceae (8 *Curcuma* species) could be attributable to the introduction of crop plants from wild sources, preference of the farmer and selection for desired traits. This also suggests that these gardens maintain wild crop relatives and could serve as an important center of plant domestication. Hammer et al. (1999) pointed out that genetic exchange through natural crosses among wild and domestic crops is a common phenomenon in the home gardens. Human regulated back yard and kitchen gardens always play important role in domestication and further utilization of wild crop relatives through hybridization (Hughes et al. 2007). These hybrid landraces will have higher capacity to overcome environmental challenges than highly exploited commercial crops (Jackson et al. 2007, Negri 2005). Other workers also reported maintenance of landraces and a wide range of genetic diversity to be a highly valued ecosystem service provided by home gardens from different region of the world (Calvet-Mir et al. 2012, Sandhu et al. 2010, Swinton et al. 2007). The importance of intra-specific diversity is highly recognized in various ecological and biological phenomenons like adaptation, survival and breeding (Feuillet et al. 2008, Nunney and Campbell 1993).

Although a very limited number of species recorded from home gardens are commercialized (e.g. *A. cathechu*, *Citrus macroptera*, *C. reticulata*, *C. sinensis*, *M. indica*, *C. papaya*, *C. colebrookianum*, *Musa paradisiaca*, *M. acuminata*), many of the species are endemic to the region (e.g. *A. chama*, *A. lakoocha*, *C. bengalense*, *C. indica*, *Cinnamomun tamala*, *C. macroptera*, *C. colebrookianum*, *M. champaca*, *O. indicum*, *S. khasiana*, *Curcuma amada*, *Zingiber zerumbet*. As per IUCN endangered and threatened categories, the species like *Bombax insignae*, *B. flabellifer*, *Centella asiatica*, *C. macroptera*, *C. rugulosa*, *Garcinia cowa*, *Hedychium spicatum*, *Livistona chinensis*, *Mangifera sylvatica* and *Rauvolfia serpentina*, were also encountered in the different home gardens. Which suggest that home gardens also appeared to host many endangered and threatened species along with high endemic species of the region.

Conclusion

Home gardening in the hilly region of Mizoram is an important agricultural system for food, fruits, vegetables, and medicine. The diversity and incorporation of native and introduced species, and cultural practices make the home gardens in the region a sustainable agricultural system. Home gardens in the region are effective reservoirs of diverse plant genetic resources. The diversity found in these home gardens are similar to forests of the region. These gardens serve as an important means of conservation of native plants through use and reducing pressure on wild resources. The availability of wild relatives of crops, abundant genetic diversity, and landraces provide a unique opportunity for crop improvement.

Table 3.1. Survey results describing physical and sociological characteristics of the villages (study sites) in Aizawl and Serchhip districts of Mizoram. Population information from Census of India (2011).

	Aizawl district			**Serchhip district**		
	Sairang	Selesih	Thingsulthlia	Serchhip	Keitum	Chhiahtlang
Population	5034	4779	3402	3865	2022	4142
No. of households	1051	873	724	613	412	308
No. of adult males	2829	2409	1663	1947	1007	2137
No. of adult females	2205	2370	1739	1918	1015	2005
Average garden size (m^2)	4297	3887	2874	3159	2556	2211
Distance from market (km)	19	12	47	4	16	10

Table 3.2. List of plant species in the home gardens [density (relative percentage of occurances), IVI (RDi + RFi + RDoi)]

Family	Species	Habit	Density	IVI	Uses
Acanthaceae	*Justica adhatoda* L.	S	11.1	2.3	Other
	Strobilanthes flaccidifolius Nees.	S	5.6	1.5	Medicinal
	Thunbergia grandiflora Roxb.	H	40.0	3.7	Ornamental
Adoxaceae	*Viburnum mullaha* Buch-Ham. Ex D.Don	S	25.6	2.8	Medicinal
Amaranthaceae	*Amaranthus caudatus* L.	S	2.2	2.1	Food
	Amaranthus viridis L.	S	25.6	4.0	Food
	Amaranthus spinosus L.	H	30.0	2.7	Food
	Chenopodium album L.	H	6.7	1.8	Food
Amaryllidaceae	*Allium sativum* L.	H	10.0	1.6	Spice
Anacardiaceae	*Mangifera indica* L.	T	67.8	5.2	Other
	Mangifera sylvatica Roxb.	T	31.1	2.7	Other
	Rhus semialata Murray.	T	6.7	0.9	Medicinal
	Semecarpus anacardium Roxb.	T	21.1	2.0	Fruit
	Spondias pinnata (L.) Kurz.	T	21.1	1.9	Other
Apiaceae	*Trachyspermum roxburghianum* (D.C)	T	2.2	0.7	Medicinal
	Centella asiatica (L.) Urban.	H	53.3	5.4	Medicinal
	Coriandrum sativum L.	H	22.2	3.2	Spice
	Daucas carota L.	H	34.4	3.7	Food
	Eryngium foetidum L.	H	22.2	2.9	Other
	Trachyspermum roxburghianum (D.C.)	H	6.7	1.6	Spice
Apocynaceae	*Alstonia scholaris* (L.) R.Br.	T	21.1	1.9	Other
	Anodendron paniculatum D.C.	T	16.7	1.7	Medicinal
	Wrightia antidysenterica (L) R.Br.	T	26.7	2.3	Medicinal
	Wrightia angustifolia Thwaites.	T	3.3	0.8	Medicinal
	Catharanthus roseus (L.) G.Don	S	23.3	3.5	Medicinal
	Rauvolfia serpentina (L.) Benth. Ex. Kurz.	S	5.6	1.9	Medicinal
Araceae	*Colocasia esculenta* (L). Schott.	H	23.3	2.8	Food
	Colocasia gigantea (Blume ex. Hassk.)	H	7.8	1.7	Food
	Colocasia lihengiae Long & Liu	H	34.4	3.3	Medicinal
	Colocasia macrorrhiza (L.) Schott.	H	32.2	3.4	Food
	Colocasia mannii Hook.	H	17.8	2.6	Food
	Colocasia obtusiloba (L.) Kunth.	H	22.2	2.9	Food
Araliaceae	*Trevesia palmata* (Roxb.) Vis.	S	7.8	0.9	Medicinal

	Aralia racemosa L.	H	2.2	1.1	Medicinal
Arecaceae	*Areca cathechu* L.	T	70.0	7.8	Other
	Borassus flabellifer L.	T	25.6	2.4	Fruits
	Calamus acanthospathus Griff.	S	23.3	2.9	Fencing
	Calamus erectus Roxb.	S	11.1	4.6	Food
	Calamus guruba Buch-Ham.	S	4.4	1.3	Food
	Licula peltata Roxb. ex Buch-Ham.	S	5.6	1.9	Roofing
	Livistona chinensis L.	S	10.0	1.9	Roofing
Asteraceae	*Artemisia vulgaris* L.	S	16.7	2.9	Food
	Helianthus annuus A. Cunn. Ex. R.Br.	S	33.3	4.5	Food
	Tithonia diversifolia (Hemsley) A. Gray	S	5.6	2.3	Ornamental
	Ageratum conyzoides L.	H	58.9	5.6	Medicinal
	Bidens biternata (Lour) Merr.	H	17.8	2.4	Medicinal
	Blumea alata D.Don.	H	8.9	1.4	Medicinal
	Chromolena odorata (L.) King. & Rob.	H	25.6	2.9	Medicinal
	Mikania micrantha Kunth.	C	26.7	3.1	Medicinal
	Spilanthes acmella (L.) Murr.	H	35.6	3.8	Medicinal
	Spilanthes oleracea L.	H	55.6	5.1	Medicinal
	Stevia rebaudiana Bertoni	H	5.6	1.3	Medicinal
Balsiminaceae	*Impatiens balsamina* L.	H	11.1	1.9	Ornamental
Bignoniaceae	*Oroxylum indicum* (L.) Kurz	T	23.3	2.1	Medicinal
Boraginaceae	*Cordia dichotoma* L.	T	8.9	1.2	Fruit
Bromeliaceae	*Annanas comosus* (L.) Merrill	H	28.9	3.3	Food
Burseraceae	*Bursera serrata* Wall.ex. Colebr.	T	11.1	1.2	Timber
	Canarium bengalense Roxb.	T	11.1	1.2	Other
Brassicaceae	*Brasica juncea* L.	H	34.4	3.5	Food
	Brassica botrytis L.	H	26.7	3.4	Food
	Brassica capitata L.	H	50.0	4.8	Food
	Brassica compestris L.	H	58.9	5.7	Food
	Brassica oleracea L.	H	53.3	5.2	Ornamental
	Brassica rapa L.	H	53.3	5.1	Food
	Raphanus sativa L.	H	38.9	3.9	Food
Cannabaceae	*Chukrasia velutina* M. (Roem.)	T	36.7	2.8	Other
	Trema orientalis (L.) Blume	T	11.1	1.2	Fencing
	Canabis sativa L.	S	36.7	4.2	Medicinal
Cannaceae	*Canna orientalis* Bouche.	H	22.2	3.1	Medicinal
Caricaceae	*Carica papaya* L.	S	48.9	5.8	Fruits

Combretaceae	*Anogeissus acuminate* (Roxb.) Wall.	T	12.2	1.4	Timber
	Terminalia bellerica (Gaertn.) Roxb.	T	22.2	2.2	Other
	Terminalia chebula Retz.	T	8.9	1.2	Other
Convolvulaceae	*Ipomoea batatas* (L.) Lam.	H	47.8	4.6	Food
	Ipomoea aquatica L.	H	11.1	1.4	Food
Cornaceae	*Alangium begoniifolium* Roxb.	T	16.7	1.8	Medicinal
Costaceae	*Costus speciosus* Smith.	H	3.3	1.3	Medicinal
	Costus variegata L.	H	12.2	2.1	Medicinal
Cucurbitaceae	*Benincasa hispida* (Thunb.) Cogn.	C	6.7	1.7	Food
	Cucumis sativa L.	C	20.0	2.5	Food
	Cucurbita maxima Duchesne	C	30.0	3.6	Food
	Cucurbita siceraria Molina.	C	40.0	4.3	Food
	Cucumis melo L.	C	32.2	3.2	Food
	Luffa cylindrica (L.) Roem.	C	13.3	1.8	Food
	Momordica charantia L.	C	24.4	3.0	Food
	Sechium edule (Jacq.) Sw.	C	68.9	6.1	Food
	Thladiantha cordifolia (Blume) Cogn.	C	11.1	2.1	Food
	Trichosanthes anguina L.	C	33.3	4.0	Food
Cupressaceae	*Cryptomaria japonica* (L.) D.Don.	T	16.7	1.8	Timber
Cyperaceae	*Cyperus rotundus* L.	H	42.2	4.1	Medicinal
Dennstaedtiaceae	*Microlepia strigosa* (Thunb.) C. Presl.	H	23.3	2.8	Medicinal
Dilleniaceae	*Dillenia indica* L.	T	26.7	2.3	Other
	Dillenia pentagyna Roxb.	T	13.3	1.6	Other
Dioscoreaceae	*Dioscorea alata* L.	C	26.7	2.8	Food
	Dioscorea glabra Roxb.	C	8.9	1.3	Medicinal
Elaeagnaceae	*Elaeagnus caudata* Schlecht. Ex Momiy	T	14.4	1.5	Fruit
Elaeocarpaceae	*Elaeocarpus aristatus* Roxb	T	17.8	1.6	Fruit
	Elaeocarpus floribundus Blume	T	36.7	2.7	Fruit
Ericaceae	*Rhododendron arboreum* Sm.	S	2.2	1.7	Ornamental
	Rhododendron formosum Wall	S	13.3	1.8	Ornamental
	Rhododendron veitchianum Hook.	S	5.6	1.2	Ornamental
	Vaccinium sprengelii G. Don.	S	16.7	2.4	Medicinal
Euphorbiaceae	*Croton wallichi* Muell. Arg.	T	16.7	1.7	Medicinal
	Emblica officinalis Gaertn.	T	7.8	1.0	Other
	Phyllanthus acidus (L.) Skeels	T	30.0	2.7	Medicinal
	Vernicia fordii Shaw.	T	11.1	1.2	Trade
	Codiaeum variegatum (L.) Rumph.	S	18.9	2.8	Medicinal

	Euphorbia royleana D.C.	S	1.1	2.3	Ornamental
	Jatropha curcas L.	S	16.7	3.0	Trade
	Ricinus communis L.	S	14.4	2.0	Medicinal
	Securinega virosa Roxb. ex.Willd.	S	4.4	1.3	Medicinal
	Manihot esculenta Crantz.	H	10.0	1.5	Food
	Phyllanthus niruri L.	H	2.2	1.4	Medicinal
Fabaceae	*Acacia nilotica* (L)	T	21.1	1.9	Fencing
	Albizia myriophylla Benth.	T	44.4	3.2	Other
	Albizia saman F.Muell.	T	34.4	2.8	Other
	Albizzia lebbeck (L.) Benth.	T	33.3	2.5	Other
	Albizia procera (Roxb.) Benth	T	20.0	1.9	Other
	Acacia pennata (L.) Willd.	S	7.8	1.4	Medicinal
	Bauhinia variegata L.	T	15.6	1.4	Other
	Bauhinia scandens L.	S	11.1	1.6	Ornamental
	Butea monosperma (Lam.) Kuntze	T	15.6	1.7	Ornamental
	Cassia alata (L.) Roxb.	T	23.3	2.1	Fuel wood
	Cassia fistula L.	T	20.0	1.7	Fuel wood
	Cassia nodosa L.	T	21.1	1.9	Fuel wood
	Cassia tora (L.) Roxb.	T	16.7	1.6	Fuel wood
	Tamarindus indica L.	T	31.1	2.3	Other
	Delonix regia L.	T	21.1	2.2	Other
	Dalbergia spinosa Roxb.	T	30.0	2.4	Other
	Erythrina arborescens Roxb.	T	12.2	1.2	Other
	Erythrina indica Lam.	T	30.0	2.5	Medicinal
	Erythrina stricta Roxb.	T	11.1	1.2	Medicinal
	Parkia timoriana (DC.) Merr.	T	72.2	5.4	Other
	Cajanus cajan (L.) Millsp.	S	53.3	5.9	Food
	Crotalaria juncea L.	S	28.9	4.2	Medicinal
	Desmodium gyroides (Roxb.) D.C.	S	27.8	3.4	Medicinal
	Canavalia ensiformis (L.) D.C.	H	12.2	2.1	Medicinal
	Mimosa pudica L.	H	21.1	2.5	Medicinal
	Dolichos tetragonolobus L.	C	38.9	3.8	Food
	Glycine max (L.) Merr.	C	32.2	3.4	Food
	Phaseolus vulgaris L.	C	37.8	4.0	Food
	Pueraria montana (Lour.) Merr.	C	13.3	2.2	Medicinal
	Vinga mungo (L.) Hepper	C	35.6	3.8	Food
	Vinga unguiculata (L.) Walp.	C	30.0	3.2	Food

Fagaceae	*Castanopsis indica* (Roxb.) D.C.	T	28.9	2.6	Timber
	Quercus griffithii Hook & Th.	T	25.6	2.2	Timber
Guttiferae	*Mesua ferrea* L.	T	21.1	1.9	Other
	Garcinia cowa Roxb.	T	24.4	2.1	Medicinal
	Garcinia lancifolia (G.Don.) Roxb.	T	1.1	0.6	Medicinal
Hypoxidaceae	*Curculigo crassiflora* (Baker) Hook.	H	20.0	2.6	Medicinal
	Itea macrophylla Wall.	T	17.8	1.6	Timber
Iteaceae	*Itea chinensis* Hook. & Arn.	T	2.2	1.0	Timber
Lamiaceae	*Mentha viridis* L.	S	10.0	2.0	Medicinal
	Ocimum americanum L.	S	5.6	1.9	Medicinal
	Ocimum sanctum L.	S	63.3	7.6	Medicinal
	Vitex negundo L.	S	2.2	1.3	Medicinal
	Elsholtzia communis Coll.	H	6.7	1.3	Medicinal
	Mentha spicata L.	H	20.0	2.8	Food
	Leucosceptrum canum Smith.	T	12.2	1.4	Medicinal
	Premna racemosa Wall. Ex. Sch.	T	8.9	1.2	Timber
	Gmelina arborea Roxb.	T	24.4	2.4	Other
	Tectona grandis L.f.	T	28.9	2.4	Timber
	Clerodendron colebrookianum Walp.	S	60.0	6.9	Medicinal
	Clerodendron infortunatum L.	S	41.1	4.7	Medicinal
	Vitex peduncularis Wall. ex Sch.	T	16.7	2.4	Medicinal
Lauraceae	*Cinnamomum verum* J. Presl	T	17.8	1.6	Other
	Cinnamomun tamala (Buch.-Ham.) T.Nees. & Eberm.	T	13.3	1.4	Other
	Cinnamoumun glanduliferum Meisn	T	5.6	1.0	Other
	Persea americana Mill.	T	63.3	5.4	Fruit
	Phoebe attenuata Nees.	T	32.2	3.2	Other
Liliaceae	*Allium cepa* L.	H	60.0	5.6	Spice
	Allium hookerii Thwaites	H	68.9	6.5	Spice
	Asparagus gonoclados Baker	C	28.9	2.7	Medicinal
	Asparagus racemosus Willd.	C	6.7	1.6	Medicinal
Lythraceae	*Duabanga grandiflora* (D.C.) Walp.	T	34.4	2.7	Other
	Lagerstroemia speciosa (L.) Pers.	T	10.0	1.2	Other
	Punica granatum L.	T	34.4	2.8	Other
	Lawsonia inermis L.	S	6.7	1.7	Medicinal
Magnoliaceae	*Magnolia champaca* (L.) Bail.	T	36.7	2.8	Other
Malvaceae	*Pterygota alata* (Roxb.) R.Br.	T	8.9	1.0	Medicinal

	Bombax ceiba L.	T	34.4	2.8	Other
	Bombax insignae Wall.	T	4.4	0.8	Timber
	Sterculia villosa Roxb.	T	64.4	4.7	Other
	Abelmoschus esculentus (L.) Moench.	S	56.7	6.4	Food
	Gossypium arboreum L.	S	7.8	1.9	Trade
	Hibiscus macrophyllus Roxb. ex Hormen	S	35.6	4.6	Medicinal
	Hibiscus rosa chinensis L.	S	6.7	1.6	Ornamental
	Hibiscus sabdariffa L.	S	43.3	5.6	Medicinal
	Hibiscus surattensis L.	S	15.6	2.6	Other
	Urena lobota L.	H	36.7	3.6	Medicinal
Marantaceae	*Phrynium capitatum* Willd.	H	15.6	1.9	Other
	Phrynium placentarium Lour.	H	8.9	1.3	Other
Melastomaceae	*Melastoma nepalensis* Lodd.	S	18.9	2.4	Medicinal
Meliaceae	*Azadirachta indica* A. Juss.	T	48.9	3.6	Other
	Melia azedaratchta L.	T	57.8	4.3	Other
	Toona ciliata M.Roem.	T	17.8	2.1	Timber
Moraceae	*Artocarpus chama* Buch-Ham.	T	18.9	1.8	Other
	Artocarpus nitidus Griff.	T	14.4	1.4	Other
	Artocarpus heterophyllus Lam.	T	62.2	4.9	Other
	Artocarpus lakoocha Roxb.	T	11.1	1.3	Other
	Ficus cunia Buch-Ham. Ex. Roxb.	T	6.7	0.9	Medicinal
	Ficus elestica Roxb.	T	35.6	3.0	Trade
	Ficus geniculata Kurz.	T	8.9	1.1	Medicinal
	Ficus recemosa L.	T	17.8	1.8	Fuel wood
	Morus australis Poir.	T	17.8	1.6	Other
Musaceae	*Ensete glaucum* Roxb.	S	23.3	3.0	Ornamental
	Musa acuminata Colla.	S	31.1	4.0	Fruit
	Musa balbisiana Colla.	S	18.9	2.8	Other
	Musa glauca Roxb.	S	32.2	4.2	Other
	Musa nagensium Prain.	S	8.9	1.6	Other
	Musa paradisiaca L.	S	54.4	6.0	Fruit
	Musa sanguinea Hook.	S	14.4	2.1	Other
	Musa velutina Wendl.	S	15.6	2.4	Food
Myricaceae	*Myrica esculenata* L.	T	28.9	2.2	Other
Myrtaceae	*Callistemon lanceolatus* D.C.	T	40.0	3.2	Ornamental
	Eucalyptus globulus Labill.	T	13.3	1.4	Other
	Psidium guajava L.	T	71.1	5.7	Fruit

	Syzigium jambos (L). Alston	T	17.8	1.8	Other
	Syzygium cumini (L.) Skeels	T	22.2	2.1	Other
Nyctaginaceae	*Bougainvellea spectabilis* Willd.	C	20.0	3.2	Ornamental
Oleaceae	*Ligustrum robustum* (Roxb.) Blume.	T	10.0	1.1	Medicinal
	Olea diocea Roxb.	T	5.6	1.0	Other
Orchidaceae	*Arundina graminifolia* (D.Don.) Hochr.	H	17.8	2.4	Ornamental
	Dendrobium chrysotoxum Lindl.	H	7.8	1.3	Ornamental
	Vanda coerulea Griff. Ex. Lindl.	H	5.6	1.7	Ornamental
Oxalidaceae	*Averrhoa carombola* L.	T	44.4	3.4	Medicinal
	Oxalis corniculata L.	H	3.3	1.5	Medicinal
Pandanaceae	*Pandanus pseudofoetidus* Roxb.	S	8.9	1.6	Medicinal
Passifloraceae	*Passiflora edulis* Sims.	C	46.7	5.3	Fruit
	Passiflora nepalensis Walp.	C	10.0	2.4	Fruit
Phyllanthaceae	*Baccaurea ramiflora* Lour.	T	18.9	1.9	Other
Piperaceae	*Piper diffusum* Vahl.	C	34.4	4.4	Medicinal
	Piper betle L.	C	21.1	2.9	Other
	Piper boehmerifolia (Micq.) D.C.	C	5.6	1.4	Other
	Piper thomsonii Hook.	C	28.9	2.8	Other
Plantaginaceae	*Plantago major* L.	H	12.2	1.5	Medicinal
Poaceae	*Arundo donax* L.	S	15.6	3.0	Roofing
	Saccharum officinarum L.	S	13.3	2.6	Medicinal
	Thysanolaena maxima (Roxb.) O. Ktze	S	35.6	4.8	Trade
	Cynodon dactylon (L.) Pers.	H	7.8	1.3	Medicinal
	Imperata cylindrica (L.) P. Beauv.	H	15.6	2.3	Roofing
	Zea mays L.	S	74.4	8.9	Food
Polygonaceae	*Polygonum barbatum* L.	S	26.7	3.6	Medicinal
	Polygonum convolvulus L.	S	20.0	3.1	Medicinal
	Polygonum orientale L.	S	34.4	3.9	Medicinal
	Polygonum plebium R.Br.	S	16.7	2.5	Medicinal
	Polygonum nepalense Meisn.	H	31.1	2.8	Medicinal
Portulaceae	*Portulacca oleracea* L.	H	4.4	0.9	Medicinal
Proteaceae	*Grevillea robusta* A.Cunn. ex R.Br.	T	21.1	2.0	Other
Pteridaceae	*Adiantum caudatum* Klotzsch	H	3.3	1.3	Food
	Adiantum phillippense L.	H	21.1	2.5	Food
	Pteris amoena Bl.	H	22.2	2.6	Medicinal
Rhamnaceae	*Ziziphus jujuba* Mill.	T	38.9	3.1	Fruit
Rhizophoraceae	*Carallia brachiata* (Lour.) Merr.	T	34.4	2.8	Timber

Rosaceae	*Malus pumila* Mill.	T	10.0	1.2	Fruit
	Prunus cerasoides D. Don	T	10.0	1.1	Fruit
	Prunus domestica L.	T	8.9	1.0	Fruit
	Pyrus communis L.	T	24.4	2.3	Fruit
	Rubus lasiocarpus Hook.	S	2.2	2.1	Fruit
	Rubus niveus Thunb.	S	2.2	1.7	Fruit
Rubiaceae	*Hymenodictyon excelsum* Wall.	T	2.2	0.7	Timber
	Psychotria calocarpa Kurz.	T	4.4	0.8	Fruit
	Saprosma ternatum (Wall) Hook.	T	20.0	2.1	Timber
	Wendlandia grandis (Hook) Cowan	T	28.9	2.2	Medicinal
	Wendlandia tinctoria (Roxb.) D.C.	T	12.2	1.2	Medicinal
	Coffea arabica L.	S	6.7	1.3	Trade
	Psychotria calocarpa Kurz.	S	3.3	1.3	Medicinal
	Paederia foetida L.	H	28.9	3.4	Medicinal
	Paederia scandes Lour.	H	8.9	1.3	Medicinal
Rutaceae	*Aegle marmelos* (L.) Correa	T	25.6	2.1	Medicinal
	Atalantia monophylla (L.) Correa	T	13.3	1.5	Medicinal
	Citrus aurantiifolia (Christm.) Swingle	T	55.6	4.6	Fruit
	Citrus grandis L.	T	60.0	4.7	Fruit
	Citrus jambhiri Lush.	T	41.1	3.1	Fruit
	Citrus karna Raff.	T	21.1	2.4	Fruit
	Citrus limon (L.) Burm.f.	T	53.3	4.3	Fruit
	Citrus limonia Osbeck	T	38.9	3.2	Fruit
	Citrus macroptera Montr.	T	65.6	6.1	Other
	Citrus medica L.	T	58.9	5.4	Other
	Citrus nobilis Loureio	T	12.2	1.7	Fruit
	Citrus paradisi Macf.	T	41.1	3.5	Fruit
	Citrus reshni Tanaka	T	7.8	1.2	Fruit
	Citrus reticulata Blanco	T	40.0	3.4	Fruit
	Citrus rugulosa Tanaka	T	17.8	2.1	Fruit
	Citrus sinensis Osbeck	T	13.3	1.4	Fruit
	Zanthoxylum budrunga Wall ex. D.C.	T	10.0	1.3	Medicinal
	Murraya koenigii (L.) Spreng	S	35.6	4.3	Medicinal
Santalaceae	*Pyrularia edulis* (Wall.) D.C.	T	4.4	1.0	Medicinal
Sapindaceae	*Lepisanthes senegalensis* Juss. ex Poir.	T	3.3	1.6	Medicinal
	Litchi chinensis Sonn.	T	31.1	2.7	Fruit
Sapotaceae	*Mimusops elengi* L.	T	62.2	4.7	Other

Scrophulariaceae	*Lindernia ruellioides* (Colsm.) Pennell	H	14.4	2.2	Medicinal
	Torenia peduncularis Benth. ex Hook.	H	5.6	1.3	Medicinal
Smilacaceae	*Smilax glabra* Roxb.	C	10.0	1.9	Medicinal
Solanaceae	*Capsicum annum* L.	S	73.3	9.3	Spice
	Capsicum frutescens L.	S	37.8	4.7	Spice
	Nicotiana tabacum L.	S	67.8	8.3	Medicinal
	Solanum aethiopicum L.	S	38.9	4.6	Medicinal
	Solanum anguivi Lam.	S	12.2	2.6	Other
	Solanum esculentum Mill.	S	21.1	3.6	Other
	Solanum khasiana Clarke	S	51.1	6.1	Other
	Solanum melongena L.	S	56.7	6.5	Food
	Solanum nigrum L.	S	36.7	4.6	Food
	Solanum spinosum L.	S	5.6	1.9	Other
	Solanum torvum Sweet	S	22.2	3.3	Food
	Solanum villosum Miller.	S	21.1	2.8	Other
	Solanum violaceum Ort.	S	11.1	1.6	Medicinal
Symplocaceae	*Symplocos laurina* Jacq.	T	5.6	0.9	Timber
Tetramelaceae	*Tetrameles nudiflora* R.Br.	T	24.4	2.2	Timber
Theaceae	*Schima wallichii* (D.C.) Korth.	T	40.0	3.1	Other
	Camellia sinensis L.	S	2.2	1.7	Food
Ulmaceae	*Celtis timorensis* Span.	T	14.4	1.5	Timber
	Ulmus lancefolia Roxb. ex. Wall.	T	10.0	1.1	Medicinal
Urticaceae	*Boehmeria penduliflora* Wedd.ex. D.G.	S	16.7	2.1	Medicinal
	Boehmeria rugulosa Wedd.	S	22.2	2.9	Medicinal
Verbenaceae	*Lantana camara* L.	S	7.8	2.3	Fencing
Vitaceae	*Vitis vinifera* L.	C	45.6	4.1	Food
Zingiberaceae	*Alpinia nigra* (Gaertn.) Burtt.	H	5.6	1.4	Food
	Amomum dealbatum L. Roxb.	H	13.3	2.3	Medicinal
	Curcuma caesia Roxb.	H	2.2	1.1	Medicinal
	Curcuma grandiflora Wall. ex. Bake	H	41.1	4.3	Medicinal
	Curcuma latiflora Valeton	H	24.4	2.8	Medicinal
	Curcuma longa L.	H	5.6	1.1	Medicinal
	Curcuma longispicata Valeton	H	14.4	2.1	Other
	Curcuma amada Roxb.	H	27.8	3.0	Other
	Curcuma trichosantha Gagnep.	H	18.9	2.2	Other
	Curcumorpha longiflora Roxb.	H	25.6	2.6	Medicinal
	Hedychium spicatum Buch-Ham. ex. Sm.	H	14.4	1.6	Medicinal

Kaempferia rotunda L.	H	11.1	1.8	Medicinal
Zingiber officinale Roscoe	H	66.7	6.1	Other
Zingiber zerumbet (L.) Roscoe ex. Sm.	H	20.0	2.3	Other

H-Herb, S-Shrub, T-Tree, C-Climber, IVI-Importance Value Index.

Table 3.3. Species richness and community indices (Shannon diversity= $H' = -\sum \{(n_i/N) \log_e (n_i/N)\}$; Dominance index= $C=\sum \{(n_i/N)^2\}$; Pielou's evenness index= e= H'/logs) of home gardens located in six different villages in Mizoram, northeast India.

Parameters	Sairang	Selesih	Thingsulthliah	Serchhip	Keitum	Chhiahtlang	Overall
Total number of species							
Trees	110	94	93	97	96	99	133
Shrubs	63	52	40	61	67	58	92
Herbs	88	61	35	74	55	66	108
Number of genera							
Trees	83	66	64	74	75	63	96
Shrubs	44	36	29	43	45	40	59
Herbs	69	45	26	57	49	51	59
Number of families							
Trees	42	42	39	42	45	41	48
Shrubs	26	25	21	28	29	26	36
Herbs	33	27	19	31	26	28	38
Diversity index							
Trees	4.05	4.44	4.42	4.45	4.43	4.44	4.76
Shrubs	4.06	3.87	3.61	4.11	4.00	3.94	4.39
Herbs	4.40	4.04	3.48	4.23	3.94	4.15	4.58
Dominance index							
Trees	0.164	0.237	0.240	0.237	0.239	0.239	0.200
Shrubs	0.287	0.316	0.355	0.280	0.296	0.305	0.246
Herbs	0.241	0.290	0.373	0.264	0.304	0.274	0.220
Evenness index							
Trees	0.863	0.978	0.975	0.972	0.971	0.966	0.971
Shrubs	0.980	0.978	0.978	0.968	0.952	0.971	0.970
Herbs	0.983	0.982	0.979	0.982	0.984	0.990	0.978

Table 3.4. Species composition similarity index based on Sorensen's similarity index [2C/(A+B)] x 100] of the overall species below the vertical line and tree species above the vertical line within the six villages in Mizoram.

Villages	Sairang 1	Selesih 2	Thingsulthliah 3	Serchhip 4	Keitum 5	Chhiahtlang 6
1. Sairang	-	91.18	88.67	75.36	71.84	78.47
2. Selesih	86.67	-	86.63	68.06	73.68	75.65
3. Thingsulthliah	69.55	74.38	-	76.84	70.90	73.96
4. Serchhip	64.56	62.74	68.83	-	83.94	85.71
5. Keitum	67.37	65.29	65.66	76.76	-	82.05
6. Chhiahtlang	70.63	69.41	66.92	78.10	51.17	-

Where, A and B are the total species in stand A and B respectively, while C is the number of species common to both stands.

Table 3.5. Correlation matrix between the garden size, total number of specie and different use categories of species.

	Garden size	Total # of species	Fencing	Food	Fruits	Fuel wood	Medicinal	Other uses	Ornamental	Roofing	Spice	Timber	Trade
Garden size	-												
Total # of species	0.820***	-											
Fencing	0.499***	0.524***											
Food	0.650***	0.846***	0.434**	-									
Fruits	0.609***	0.646***	0.432**	0.471***	-								
Fuel-wood	0.441**	0.531***	0.104^{ns}	0.355*	0.274*	-							
Medicinal	0.678***	0.879***	0.452***	0.711***	0.498***	0.393**	-						
Other uses	0.731**	0.858***	0.366*	0.626***	0.471***	0.509***	0.642***	-					
Ornamental	0.419**	0.540***	0.227*	0.361**	0.133^{ns}	0.348**	0.519***	0.440**	-				
Roofing	0.175^{ns}	0.310*	0.009^{ns}	0.296*	0.048^{ns}	0.153^{ns}	0.305**	0.282*	0.208^{ns}	-			
Spice	0.321*	0.407**	0.149^{ns}	0.265*	0.202*	0.325*	0.272*	0.407**	0.174^{ns}	0.071^{ns}	-		
Timber	0.678***	0.748***	0.475***	0.614***	0.523***	0.430**	0.524***	0.652***	0.290*	0.187^{ns}	0.217^{ns}	-	
Trade	0.366*	0.422**	0.257*	0.212*	0.291*	0.186^{ns}	0.401**	0.269*	0.416**	0.073^{ns}	0.233*	0.267*	-

***P<0.001, **P<0.01, *P<0.05, ns=not significant (N=90).

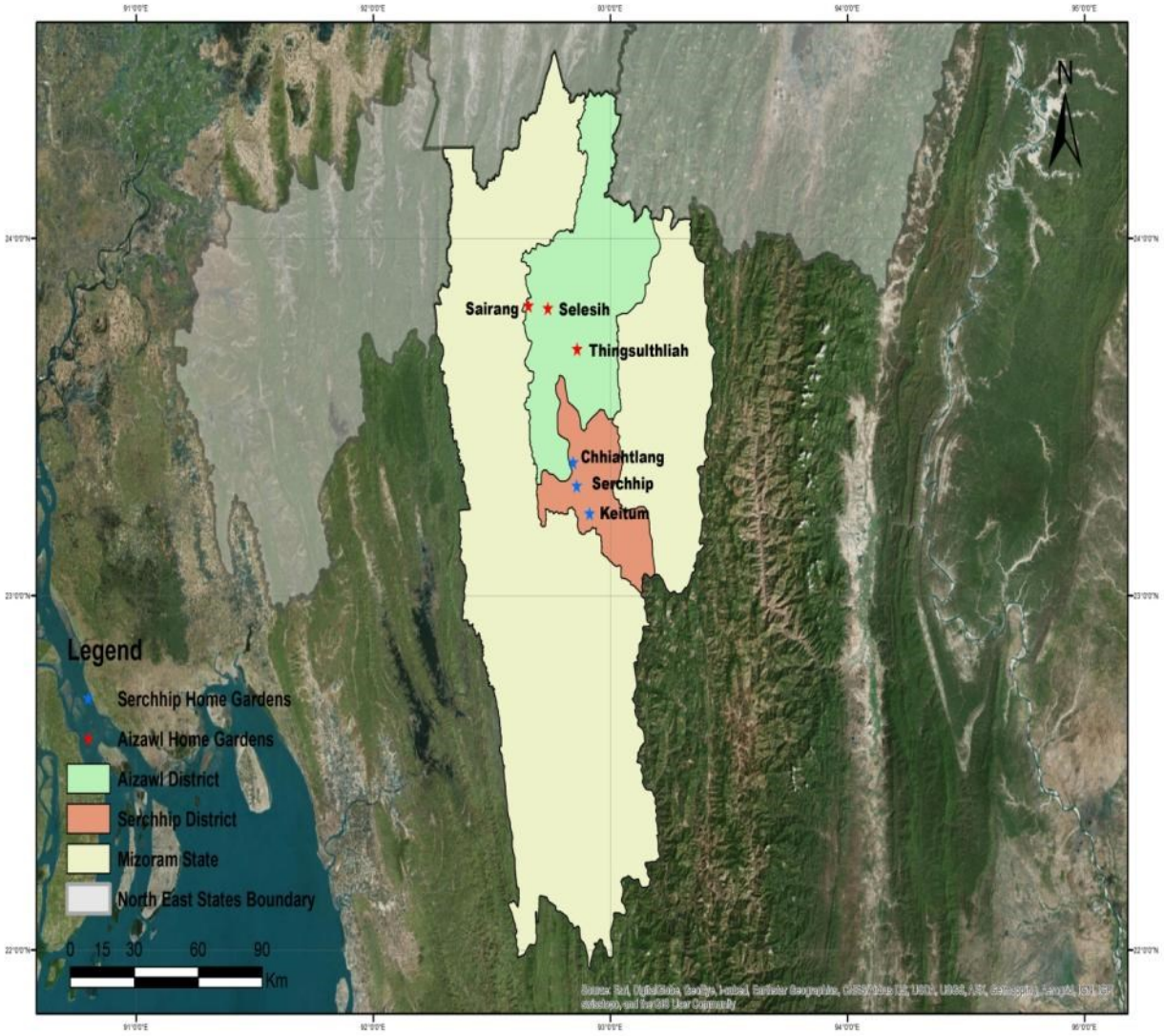


Figure 3.1. Map showing the villages where home gardens are located in Aizawl and Serchhip districts, Mizoram, northeast India.

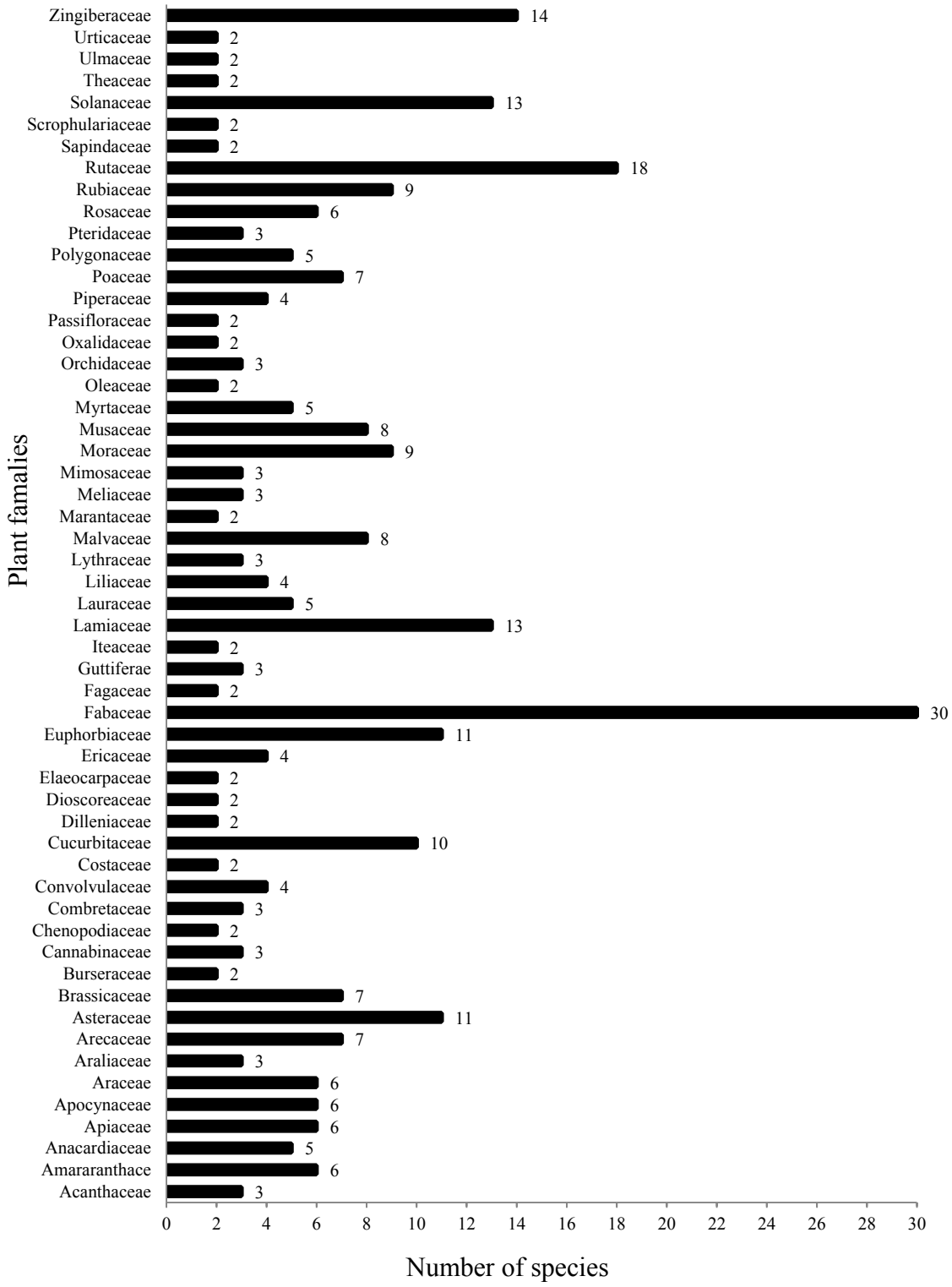


Figure 3.2. Species-rich plant families (≥ 2) in the home gardens in Mizoram.

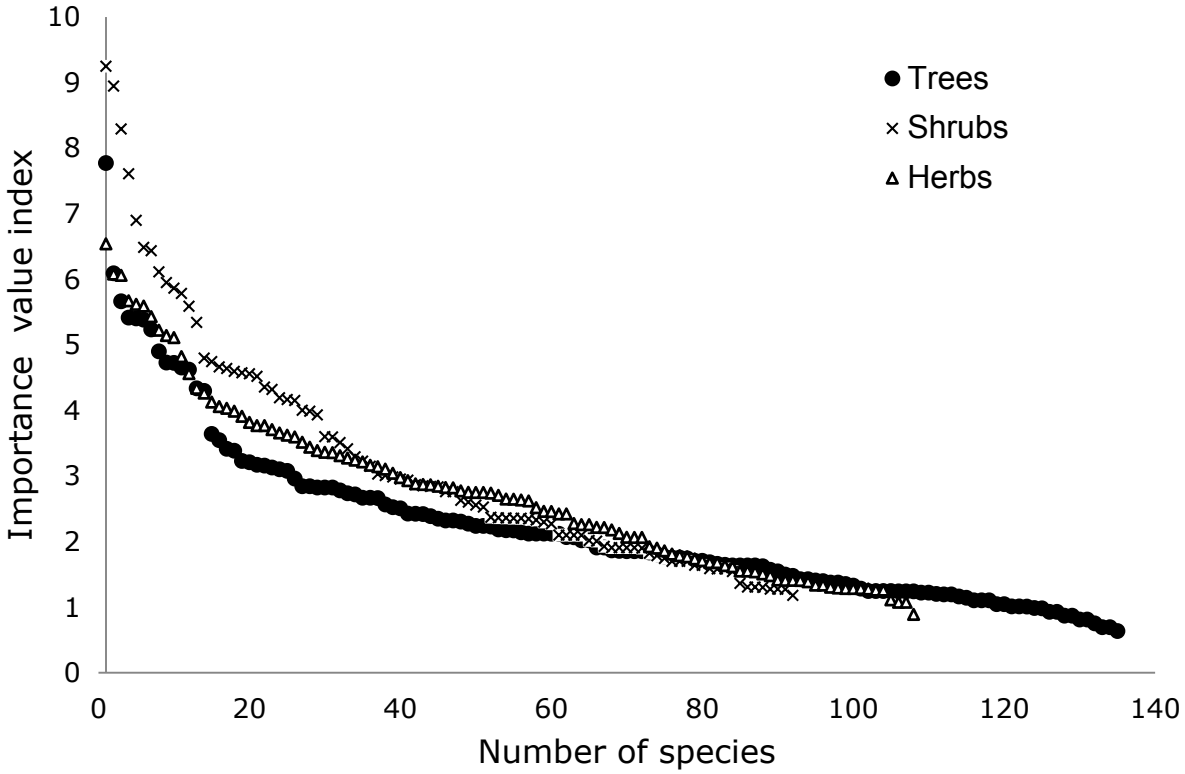


Figure 3.3. Importance value distribution curve of tree, shrubs and herbs species in the six home-gardens in Mizoram.

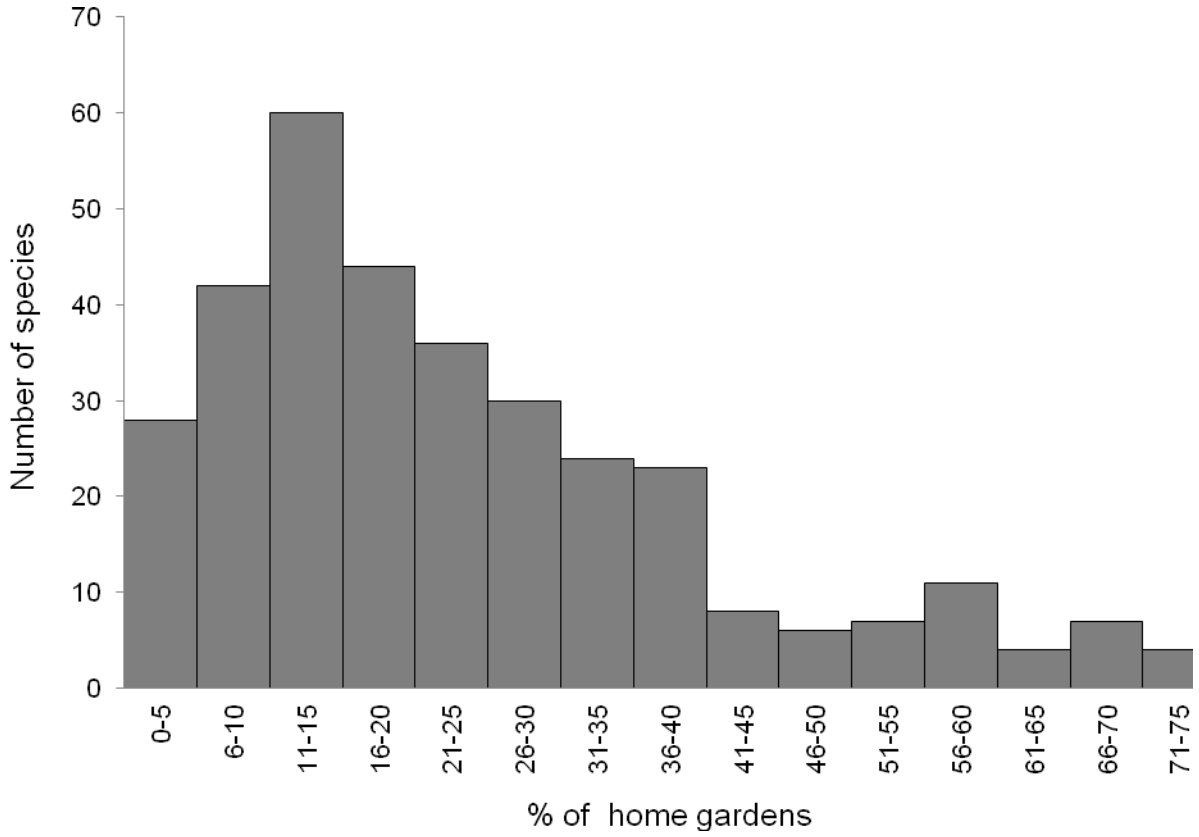


Figure 3.4. Frequency distribution of species richness.

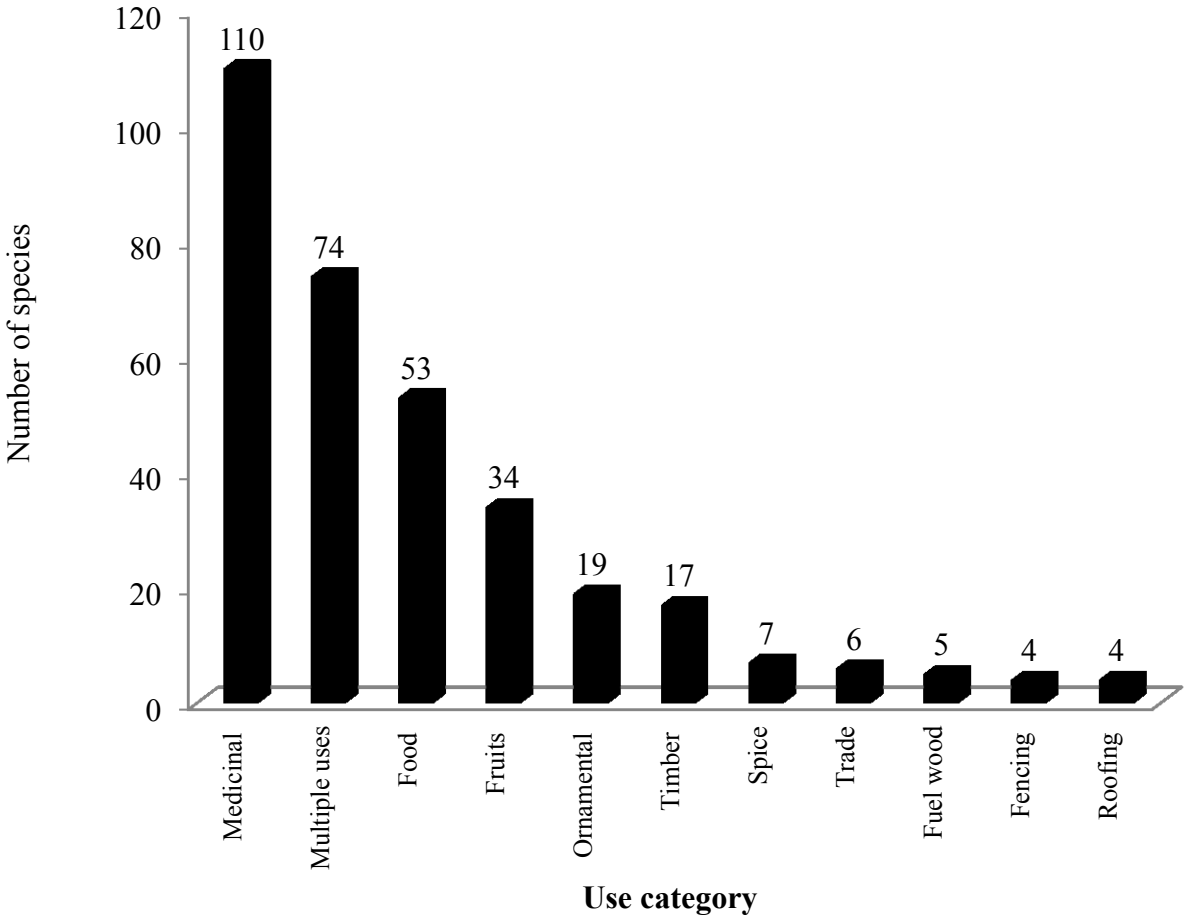


Figure 3.5. Use category of species.

SUMMARY

Phylogeny of *Citrus* species in northeast India was reconstructed using three chloroplast (trnL-trnF, trnS-trnG and rps16) and one nuclear (ITS2) DNA regions. Three different methods viz., Maximum parsimony (MP), Maximum likelihood (ML) and Bayesian (BI) inferences were used to reconstruct the phylogeny of 24 *Citrus* species. The phylogenetic trees inferred from the concatenated chloroplast and nuclear data matrix showed better resolution and resulted in well resolved phylogenetic grouping of *Citrus*. The different analyses grouped the morphologically distinct 24 *Citrus* species into five phylogenetic groups. Besides the three well recognized true species (*C. grandis, C. medica* and *C. retiuclata*), the other two species (*C. indica* and *C. assamensis*) may also be considered as true species that require further study with more accessions and molecular markers. In all the analyses, Clade I comprises of two species *C. indica* and *C. medica*. Clade II comprised of only a single wild and endemic species *C. assamensis*. Clade III comprises seven species including 6 acid members (*C.aurantifolia, C. limonia, C. volkameriana, C. limettioides, C. pseudolimon* and *C. limon*) and one wild *Papeda* (*P. trifoliata*). Clade IV comprises six species, including all five mandarin species (*C. nobilis, C. reticulata, C. aurantium, C. sinensis* and *C. reshni*) and one endangered and endemic *Papeda* species (*C. macroptera*). Clade V comprises eight species including four pomelo (*C. megaloxycarpa, C. grandis, C. rugulosa* and *C. paradisi*), two wild *Papeda* (*C. ichangensis* and *C. latipes*), and two acid members (*C. karna* and *C. jambhiri*). *Citrus* species of the eastern Himalayan regions of northeast India are morphologically variable but have low level of genetic divergence. There may not be as many as 24 or more true biological species that were described on the basis of horticultural/ morphological characteristics. The phylogenetic relationships obtained by three different analyses revealed polyphyletic groupings of acid and *Papeda* members.

Citron (*Citrus medica* L.) one of the primitive and true *Citrus* species commonly occur in wild and domestic habitats in northeast India. Genetic diversity and structure of 219 citron individuals of 8 domestic and 4 wild populations were assessed using 5 polymorphic microsatellite markers. In total 67 alleles were detected with an average of 13.4 alleles per locus. The mean number of alleles (2.60 to 7.20) varied significantly within the wild and domesticated populations. The mean observed (0.220 to 0.540) and expected (0.438 to 0.733) heterozygosity values also varied significantly among populations. In general, domesticated populations exhibited slightly higher level of genetic diversity than wild populations and the difference between them was insignificant. Population differentiation (F_{ST}) values ranged between 0.174 - 0.252 in wild and 0.193 - 0.294 in domestic populations. The AMOVA results

revealed significant (P<0.001) diversity with high (47.53%) among individual and low (24.98%) among population variability. The pairwise Nei's genetic distances among domesticated populations were low as compared to the genetic distances among wild populations. The indirect estimates of gene flow (Nm) among populations varied significantly (P<0.001) and ranged between 0.600 to 1.187. The UPGMA analyses of Nei's genetic distance and Bayesian clustering in STRUCTURE assigned 219 individual into five genetically distinct clusters and showed mixed ancestry of wild and domestic populations. The exchange of plants among home gardens in the form of seed, seedlings and cuttings is a common practice in the region, and may have lead to mixing of genotypes among populations. The overall genetic diversity of *Citrus medica* in the region is high and may serve as an important genetic resource for sustainable use.

The size of indigenous home gardens in Mizoram, northeast India ranged from 0.10 to 0.60 ha and harbor high biodiversity composed of annual, biennial and perennial plants with structural similarity to tropical forests. A total of 333 plant species (133 trees, 92 shrubs and 108 herbs) belonging to 122 families with an average of 78 species per home garden were found. The dominant plant families were Fabacece, Rutaceae, Zingiberaceae, Solanaceae, Euphorbiaceae, Asteraceae and Curcurbitaceae. The plant family Fabaceae produced majority of the food plants, and the highest number of fruits and medicinal plants were from Rutaceae. The majority of the fruit plants were represented by tree species and vegetables were mainly from herbs and shrubs. Overall, the number and diversity of tree species was higher than the herb and shrub species. The species diversity index values were 4.76, 4.39 and 4.58 for trees, shrubs and herbs respectively. Dominance index values ranged between 0.164 – 0.373 and mostly dominated by tree species. The high evenness values indicate uniform distribution of species within the gardens. These home gardens serve as a year round production system for food, vegetables, medicine, fruits, fuel wood and timber. The wide range of crop plants fulfill varying nutritional needs of the community. The presence of intra-specific diversity in a variety of plant groups such as *Citrus, Colocasia, Curcuma, Musa, Polygonum* and *Solanum* could be attributable to existence of wild relatives near domesticated sites. The home gardens in the region are reservoirs of diverse plant genetic resources including wild relatives and serve as important genetic resource for the improvement of horticultural and crop plants.

REFERENCES

Abbo S, Van-Oss RP, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant Science*, **19(6),** 351–360. doi:org/10.1016/j.tplants.2013.12.002.

Abebe T, Starch FJ, Wiersum KF, Bongers F (2013) Diversity, composition and density of trees and shrubs in agroforestry homegardens in Southern Ethiopia. *Agroforestry Systems,* **87**, 1283–1293.

Agelet A, Bonet MA, Valles J (2000) Home gardens and their role as a main source of medicinal plants in mountain regions of Catalonia (Iberian Peninsula). *Economic Botany,* **54**, 295–309.

Ahmad AM, Abood F (1990) Selected forest tree with potential application in Malaysian Agroforestry. BIOTROP Special Publication No. 39. BIOTROP, Indonesia. pp.77–89.

Albuquerque UP, Andrade LHC, Caballero J (2005) Structure and floristic of home gardens in Northeastern Brazil. *Journal of Arid Environment,* **62**, 491–506.

Allendorf FW, Luikart G, Aitken SN (2013) Conservation and the Genetics of Populations, 2nd Edition, John Wiley and Sons, Ltd, UK.

Alvarez-Buylla Roces ME, Lazos Chavero E, Garcia-Barrios JR (1989) Home gardens of humid tropical region in Southeast Mexico: an example of an agro-forestry cropping system in a recently established community. *Agroforestry Systems,* **8**, 133–156.

Andrianoelina O, Favreau B, Ramamonjisoa L, Bouvet JM (2009) Small effect of fragmentation on the genetic diversity of *Dalbergia monticola*, an endangered tree species of the eastern forest of Madagascar, detected by chloroplast and nuclear microsatellites. *Annals of Botany,* **104**, 1231–1242.

Araujo EF, Queiroz LP, Machado MA (2003) What is *Citrus*? Taxonomic implications from a study of cp-DNA evolution in the tribe Citreae (Rutaceae subfamily Aurantioideae). *Organism Diversity and Evolution,* **3**, 55–62.

Asadi Abkenar A, Isshiki S, Tashiro Y (2004) Phylogenetic relationships in the ‘‘true *Citrus* fruit trees’’ revealed by PCR-RFLP analysis of cpDNA. *Scientia Horticulturae,* **102,** 233–242.

Avise JC (1994) Molecular Markers, Natural History, and Evolution. Chapman & Hall, New York.

Barkley NA, Krueger RR, Federici CT, Roose ML (2009) What phylogeny and gene genealogy analyses reveal about homoplasy in *Citrus* microsatellite alleles. *Plant Systematics and Evolution*, **282**, 71–86.

Barkley NA, Roose ML, Krueger RR, Federici CT (2006) Assessing genetic diversity and population structure in a *Citrus* germplasm collection utilizing simple sequence repeat markers (SSRs). *Theoretical and Applied Genetics*, **112**, 1519–1531.

Barrett HC, Rhodes AM (1976) A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. *Systematic Botany*, **1**, 105–136.

Baudry J, Yu Z, Liewan C (1999) Landscape patterns changes in two subtropical Chinese villages as related to farming policies. *Critical Review in Plant Sciences,* **18(3)**, 373–380.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: Hoch PC, Stephenson AG (Eds.). Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis.pp. 289–303.

Bayer RJ, Mabberley DJ, Morton C, Miller CH, Sharma IK, Pfeil BE, Rich S, Hitchcock R, Sykes S (2009) A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. *American Journal of Botany,* **96**, 668–685.

Bernholt H, Kehlenbeck K, Gebauer J, Buerkert A (2009) Plant species richness and diversity in urban and peri-urban gardens of Niamey, Niger. *Agroforestry Systems,* **77**, 159–179.

Bhattacharya SC, Dutta S (1956) Classifications of *Citrus* fruits of Assam. *ICAR Science Monograph*, **20**, 1–110.

Bock WJ (2004) Species: the concept, category and taxon. *Journal of Zoological Systematics and Evolutionary Research* **42**, 178–190.

Brierley JS (1985) West Indices kitchen gardens: A historical perspective with current insights from Grenada. *Food and Nutrition Bulletin,* **7**, 52–60.

Calvet-Mir L, Gomez-Baggethun E, Reyes-Garcia V (2012) Beyond food production: ecosystem services provided by home gardens. A case study in Vall Fosca, Catalan Pyrenees, Northeastern Spain. *Ecological Economics,* **74**, 153–160.

Census of India (2011) Registrar Publication. New Delhi: Government of India.

Chase W, Morton CM, Kallunki JA (1999) Phylogenetic relationships of Rutaceae: a cladistic analysis of the subfamilies using evidence from rbcL and atpB sequence variation. *American Journal of Botany,* **8,** 1191–1199.

Cilliers SS, Cilliers J, Lubbe CE, Siebert S (2013) Ecosystem services of urban green spaces in African countries-perspectives and challenges. *Urban Ecosystem,* **16**, 681–702.

Clarke LW, Li L, Jenerette GD, Yu Z (2014) Drivers of plant biodiversity and ecosystem service production in home gardens across the Beijing Municipality of China. *Urban Ecosystem.* doi:10.1007/s11252-014–0351-6.

Collevatti RG, Grattapaglia D, Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. *Molecular Ecology*, **10**, 349–356.

Corazza-Nunes MJ, Machado MA, Nunes WMC, Cristofani M, Targon MLPN (2002) Assessment of genetic variability in grapefruits (*Citrus paradisi* Macf.) and pummelos (*C. maxima* (Burm.) Merr.) using RAPD and SSR markers. *Euphytica*, **126**, 169–176.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics,* **144**, 2001–2014.

Cracraft J (1989) Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Otte DJ, Endler JA (Eds.). Speciation and its Consequences. Sinauer Associates, Sunderland, MA. pp. 28–59.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods,* **9***,* 772.

Das T, Das AK (2005) Inventorying plant biodiversity in home gardens: a case study in Barak Valley, Assam, North East India. *Current Science,* **89**, 155–163.

Davis JI (1996) Phylogenetics, molecular variation and species concepts. *Bioscience*, **46(7)**, 502–511.

Dayanandan S, Bawa KS, Kesseli RV (1997) Conservation of microsatellite among tropical trees (*Leguminosae*). *American Journal of Botany*, **84**, 1658–1663.

Donoghue MJ (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist,* **88**, 172-181.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2009) Geneious v4.7. Available from http:www.geneious.com.

Ellstrand NC (2003) Current knowledge of gene flow in plants. *Philosophical Transactions of the Royal Society, Biological Sciences,* **358**, 1163–1170.

Ellstrand NC (2014) Is gene flow the most important evolutionary force in plants. *American Journal of Botany,* **101(5)**, 737–753.

Ellstrand NCH, Prentice C, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539–563.

Endler JA (1986) Natural Selection in the Wild. Princeton, NJ: Princeton University Press.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN Version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplo types: application to human mitochondrial DNA restriction data. *Genetics,* **131**, 479–491.

Fang, DQ, Krueger RR, Roose ML (1998) Phylogenetic relationships among selected *Citrus* germplasm accessions revealed by inter-simple sequence repeat (ISSR) markers. *Journal of the American Society for Horticultural Science,* **123,** 612– 617.

Farris JS, Kallersjo M, Kluge AG, Bult C (1994) Testing significance of congruence. *Cladistics,* **10**, 315–320.

Federici CT, Fang DQ, Scora RW, Roose ML (1998) Phylogenetic relationships within the genus *Citrus* (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. *Theoretical and Applied Genetics*, **96**, 812–822.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution,* **17**, 368–376.

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

Fernandes ECM, Nair PKR (1986) An evaluation of the structure and function of tropical homegardens. *Agricultural System,* **21**, 279–310.

Feuillet C, Langridge P, Hopugh R (2008) Cereal breeding takes a walk on the wild side. *Trends in Genetics,* **24**, 24–32.

Finet C, Timme, RE, Delwiche CF, Marletaz F (2010) Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology,* **20**, 2217–2222.

Fraser JA, Junqueira AB, Clement CR (2011) Homegardens on Amazonian dark earths, non anthropogenic upland, and floodplain soils along the Brazilian middle Madeira River exhibit diverging agro-biodiversity. *Economic Botany*, **65**, 1-12. doi:10.1007/s12231-010-9143.

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG, Purugganan MD (2014) Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceeding of the National Academy of Sciences*, **111(17)**, 6147–6152.

Gajaseni J, Gajaseni N (1999) Ecological rationalities of the traditional home garden system in the Chao Phraya Basin, Thailand. *Agroforestry Systems*, **46**, 3–23.

Galluzzi G, Eyzaguirre P, Negri V (2010) Home gardens: neglected hotspots of agro-biodiversity and cultural diversity. *Biodiversity and Conservation,* **19**, 3635–3654.

Garcia-Jacas N, Susanna A, Garnatje T, Vilatersana R (2001) Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Annals of Botany*, **87**, 503–515.

Garcia-Lor A, Curk F, Snoussi-Trifa H, Morill R, Ancillo G, Luro F, Navarro L, Ollitrault P (2013) A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true *Citrus* fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Annals of Botany*, **111**, 1–19.

Gliessman SR (1990a) Integrating trees into agriculture: the home garden agro-ecosystem as an example of agro-forestry in the tropics. In: Gliessman SR (Eds.). Agroecology: Researching the Ecological Basis for Sustainable Agriculture. Springer-Verlag, New York, pp. 160–168.

Gliessman SR (1990b) Understanding the basis for sustainability for agriculture in the tropics: experiences in Latin America. In: Edwards CA, Lal R, Madden P, Miller RH, House G (Eds.). Sustainable Agricultural Systems. St. Lucie Press, Delray Beach, Florida, pp. 378–390.

Goudet J (2001) FSTAT Version 2.9.3, a program to estimate and test gene diversities and fixation indices. Available from http:www2.unil.ch/popgen/softwares/fstat.htm

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–72.

Grant V. 1981. Plant Speciation. Colombia University Press, New York.

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology,* **8,** 513–525.

Hammer K, Laghetti G, Perrino P (1999) A checklist of the cultivated plants of Ustica (Italy). *Genetic Resources and Crop Evolution,* **46**, 95–106.

Hamrick JL, Godt MJW *(*1996*)* Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of Londo, Series B, Biological Sciences,* **351**, 1291–1298.

Hardy GH (1908) Mendelian proportions in a mixed population. *Science,* **28**, 49–50.

Hennig W (1966) Phylogenetic Systematics. Urbana: University of Illinois Press.

Herrero R, Asins MJ, Pina JA, Carbonell EA, Navarro L (1996) Genetic diversity in the orange subfamily Aurantioideae. II. Genetic relationships among genera and species. *Theoretical and Applied Genetics,* **93**, 1327–1334.

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology,* **66**, 411–453.

Hooker JD (1875) Rutaceae. The Flora of British India, Volume 1. Reeve and Company, London. pp. 484–517.

Huai H, Xu W, Wen G, Bai W (2011) Comparison of the Home gardens of eight cultural groups in Jinping County, Southwest China. *Economic Botany*, **65(4)**, 345–355.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* **17**, 754–755.

Hughes CE, Govindarajulu R, Robertson A, Filer DL, Harris SA, Bailey CD (2007) Serendipitous backyard hybridization and the origin of crops. *PNAS*, **104**, 14389–14394.

Hynniewta M, Malik SK, Rao SR (2011) Karyological studies in ten species of *Citrus* (Linnaeus, 1753) (Rutaceae) of North-East India. *Comparative Cytogenetics,* **5(4)**, 277–287.

Hynniewta M, Malik SK, Rao SR (2014) Genetic diversity and phylogenetic analysis of *Citrus* (L) from north-east India as revealed by meiosis, and molecular analysis of internal transcribed spacer region of rDNA. *Meta Gene*, **2,** 237–251.

Idohoua R, Fandohanabc B, Salakoa VK, Kassaa B, Gbedomona RC, Yedomonhana H, Lucas R, Kakaia G, Assogbadjo AE (2014) Biodiversity conservation in home gardens: traditional knowledge, use patterns and implications for management. *International Journal of Biodiversity Science and Ecosystem Service Management*, **10(2)**, 89–100.

Jackson LE, Pascual U, Brussaard L, de-Ruiter P, Bawa KS (2007) Biodiversity in agricultural landscapes: Investing without losing interest. *Agriculture Ecosystem and Environment*, **121**, 193–195.

Jaganmohan M, Vailshery LS, Gopal D, Nagendra H (2012) Plant diversity and distribution in urban domestic gardens and apartments in Bangalore, India. *Urban Ecosystem*, **15**, 1-15. Doi:10.1007/s11252-012-0244-5.

Jarvis DI, Hodgkin T (1998) Strengthening the scientific basis of in situ conservation of agricultural biodiversity on-farm: options for data collecting and analysis. Proceedings of a workshop to develop tools and procedures for in situ conservation on-farm, Rome.

Jena SN, Kumar S, Nair KN (2009) Molecular phylogeny in Indian *Citrus* L. (Rutaceae) inferred through PCR-RFLP and trnL-trnF sequence data of chloroplast DNA. *Scientia Horticulturae*, **199**, 403–416.

Kabir ME, Webb EL (2008) Can home gardens conserve biodiversity in Bangladesh? *Biotropica*, **40(1)**, 95–103.

Kabir ME, Webb EL (2009) Household and home garden characteristics in southwestern Bangladesh. *Agroforestry Systems*, **75**, 129–145.

Kharal D (2000) Diversity and dynamics of tree species and its sustainability in the rural farm land: A case study in Chitawan districts in Nepal. M.Sc. Thesis, The Agricultural University of Norway, Norway.

Krug CA (1943) Chromosome numbers in the subfamily Aurantioideae with special reference to the genus *Citrus*. *Botanical Gazette*, **104(4),** 602–61.

Kumar BM, George SJ, Chinnamanis S (1994) Diversity, structure and standing stock of wood in the home gardens of Kerala in Peninsular India. *Agroforestry Systems*, **25**, 243–262.

Kumar BM, Nair PKR (2004) The enigma of tropical home gardens. *Agroforestry Systems*, **61**, 135–152.

Kumar K, Nair KN (2013) Genetic variation and phylogenetic relationships among Indian *Citrus* taxa revealed by DAMD-PCR markers. *Genetic Resources and Crop Evolution.* doi: 10.1007/s10722-013-9954-7.

Kumar S, Jena SN, Nair NK (2010) ISSR polymorphism in Indian wild orange (*Citrus indica* Tanaka, Rutaceae) and related wild species in North-east India. *Scientia Horticulturae*, **123**, 350–359.

Kumar S, Nair KN, Jena SN (2013) Molecular differentiation in Indian *Citrus L.* (Rutaceae) inferred from nrDNA- ITS sequence analysis. *Genetic Resources and Crop Evolution*, **60**, 59–75.

Langgut D (2014) Prestigious fruit trees in ancient Israel: first palynological evidence for growing *Juglans regia* and *Citrus medica*. *Israel Journal of Plant Sciences.* doi: 10.1080/07929978.2014.950067.

Larios C, Casas A, Vallejo M, Moreno-Calles AI, Blancas J (2013) Plant management and biodiversity conservation in Nahuatl homegardens of the Tehuacan Valley, Mexico. *Journal of Ethnobiology and Ethnomedicine*, **9**,74. doi:10.1186/1746-4269-9-74

Larson G, Piperno DR, Allaby RG, Purugganan M, Andersson L, Arroyo-Kalin M, Barton L, Vigueira CC, Denham T, Dobney K, et al. (2014) Current perspectives and the future of domestication studies. *PNAS*, **111(17),** 6139–6146, doi:10.1073/pnas.1323964111

Li YZ, Cheng YJ, Tao NG, Deng XX (2007) Phylogenetic analysis of mandarin landraces, wild mandarins, and related species in China using nuclear leafy second intron and plastid trnL-trnF sequence. *Journal of American Society of Horticultural Sciences*, **132**, 796–806.

Liang G, Xiong G, Guo Q, He Q, Li X (2007) AFLP analysis and the taxonomy of *Citrus*. *Acta Horticulture*, **760**, 137–142.

Linnaeus C (1753) Species Plantarum. Stockholm.

Liu K, Muse SV (2005) Power Marker: Integrated analysis environment for genetic marker data. *Bioinformatics,* **21**, 2128–2129.

Loram A, Warren P, Thompson K, Gaston K (2011) Urban domestic gardens: The effect of human interventions on garden composition. *Environmental Management*, **48**, 808–824.

Lu Z, Zhou Z, Xie RJ (2011) Molecular phylogeny of the true *Citrus* fruit trees group (Aurantioideae, Rutaceae) as inferred from chloroplast DNA sequences. *Agricultural Sciences in China*, **10**, 49–57.

Luikart G, Allendorf FW, Cornuet JM, William BS (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89***,* 238–247.

Lushington AW (1910) The genus *Citrus*. *Indian Forester,* **36**, 323–353.

Mabberley DJ (1998) Australian *Citreae* with notes on other Aurantioideae (Rutaceae). *Telopea*, **7**, 333–344.

Mabberley DJ (2004) *Citrus* (Rutaceae): A Review of recent advances in etymology, systematics and medical applications. *Blumea*, **49**, 481–498.

Maddison DR, Maddison WP (2001) MacClade 4. Sinauer, Sunderland, Massachusetts, USA.

Makaraphirom P (1989) Checklist of Species for Extension in Agroforestry Systems. Agroforestry Research 30, Royal Forestry Department, Bangkok, Thailand.

Malik SK, Kumar S, Singh IP, Dhariwal OP, Chaudhury R (2013) Socio-economic importance, domestication trends and in situ conservation of wild *Citrus* species of Northeast India. *Genetic Resources and Crop Evolution.* doi:10.1007/s10722-012-9948-x

Mallet J (2007) Hybrid speciation. *Nature,* **446**, 279–283.

Mantel N (1967) The Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Mayr E (1942) Systematics and the origin of species. New York: Columbia University Press.

Mayr E (1963) Animal Species and Evolution. Belknap Press, Cambridge, MA.

Mayr E (1996) What Is a Species, and What Is Not? *Philosophy of Science*, **63(2)**, 262-277.

McKey D, Elias M, Pujol B, Duputie A (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist*, **186**, 318–332.

Mendez VE, Lok R, Somarriba E (2001) Interdisciplinary Analysis of home gardens in Nicaragua: Micro-zonation, plant use and socioeconomic importance. *Agroforestry Systems*, **51**, 85–96.

Michon G, Bompard J, Hecketsweiler P, Ducatillon C (1983) Tropical forest architectural analysis to agro-forests in the humid tropics: The examples of traditional village-agro-forests in west Java. *Agroforestry Systems*, **1**, 117–129.

Miller AJ, Gross BL (2011) From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, **98**, 1389–1414.

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species *Molecular Ecology*, **17,** 1552–1563.

Moore GA (2001) Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends in Genetics*, **17**, 536–540.

Morton CM, Grant M, Blackmore S (2003) Phylogenetic relationships of the Aurantioideae inferred from cpDNA sequenced data. *American Journal of Botany,* **90**, 1463–1469.

Nair KN, Nayar MP (1997) Rutaceae. In: Hajra PK, Nair VJ, Daniel P (Eds.). Flora of India, Volume IV. Botanical Survey of India, Calcutta, India. pp. 229–407.

Negri V (2005) Agro-biodiversity conservation in Europe: ethical issues. *Journal of Agriculture Environment and Ethics*, **18**, 3–25.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* **89**, 583–590.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Nicolosi E, Deng ZN, Genetile A, La Malfa S, Ciontinella G, Tribulata E (2000) *Citrus* phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics*, **100**, 1155–1166.

Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. *Cladistics,* **6**, 211–223.

Novick RR, Dick CW, Lemes MR, Navarro C, Caccone A, Bermingham E (2003) Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Molecular Ecology*, **12**, 2885–2893.

Nunney L, Campbell KA (1993) Assessing minimum viable population size: demography meets population genetics. *Trends in Ecology,* **8**, 234–239.

Ollitrault F, Terol J, Pina JA, Navarro L, Talon M, Ollitrault P (2010) Development of SSR markers from *Citrus clementina* (Rutaceae) BAC end sequences and interspecific transferability in *Citrus*. *American Journal of Botany*, **97**, 124–129.

Ollitrault P, Terol J, Chen C, Federici CT, Lotfy S, Hippolyte I, Ollitrault F, Berard A, Chauveau A, Cuenca J, Costantino G, et al (2012) A reference genetic map of *C. clementina* Hort. ex Tan.; *Citrus* evolution inferences from comparative mapping. *BMC Genomics*, **13**, 593, doi: 10.1186/1471-2164-13-593.

Oxelman B, Liden M, Berglund D (1997) Chloroplast rpsl6 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution*, **206**, 393–410.

Padoch C, De Jong W (1991) The house gardens of Santa Rosa: diversity and variability in an Amazonian agricultural system. *Economic Botany,* **45**, 166–175.

Pang XM, Hu CG, Deng XX (2007) Phylogenetic relationships within *Citrus* and its related genera as inferred from AFLP markers. *Genetic Resources and Crop Evolution*, **54**, 429–436.

Penjor T, Anai T, Nagano Y, Matsumoto R, Yamamoto M (2010) Phylogenetic relationships of *Citrus* and its relatives based on rbcL gene sequences. *Tree Genetics and Genomes*, **6**, 931–939.

Penjor T, Yamamoto M, Uehara M, Ide M, Matsumoto N, Matsumoto R, Nagano Y (2013) Phylogenetic relationships of *Citrus* and its relatives based on matK gene sequences. *PLoS ONE.* doi:8(4)-e62574.

Pessina D, Gentili R, Barcaccia G, Nicole BS, Rossi G, Barbesti S, Sgorbatia S (2011) DNA content, morphometric and molecular marker analyses of *Citrus limonimedica*, *C. limon* and *C. medica* for the determination of their variability and genetic relationships within the genus *Citrus*. *Scientia Horticulturae,* **129**, 663–673.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics*, **37**, 187–214.

Phillips EA (1959) Methods of Vegetation Study. Henry Holt and Company, New York.

Pielou EC (1969) An Introduction to Mathematical Ecology. Wiley, New York.

Pridgeon MA, Solano R, Chase MW (2001) Phylogenetic relationships in Pleurothallidinae (Orchidaceae): combined evidence from nuclear and plastid DNA sequences. *American Journal of Botany,* **88(12)**, 2286–2308.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Purugganan MD, Fuller DQ (2011) Archaeological data reveal slow rates of evolution during plant domestication. *Evolution*, **65(1)**, 171–183.

Ramadugu C, Pfeil BE, Keremane ML, Lee RF, Maureira-Butler IJ, Roose ML (2013) A six nuclear gene phylogeny of *Citrus* (Rutaceae) taking into account hybridization and lineage sorting. *PLoS ONE*, **8(7)**:1–15.

Raymond M, Rousset F (1995) GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reyes-Garcia V, Aceituno L, Vila S, Calvet-Mir L, Garnatje T, Jesch A, Lastra JJ, Parada M, Rigat M, Valles J, Pardo-de-Santayana M (2012) Home gardens in three mountain regions of the Iberian Peninsula: description, motivation for gardening, and gross financial benefits. *Journal of Sustainable Agriculture*, **36**, 249–270.

Richardson JE, Fay MF, Cronk QCB, Bowman D, Chase MW (2000) A phylogenetic analysis of Rhamnaceae using *RBCL* and *TRNL-F* plastid DNA sequences. *American Journal of Botany*, **87(9)**, 1309–1324.

Rico-Gray V, Garcia-Franco JG, Chemas A, Puch A, Sima P (1990) Species composition, similarity and structure of Maya home gardens in Tixpeual and Tixcacatuyul, Yucatan, Mexico. *Economic Botany*, **44**, 470–487.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Roy B, Rahman MH, Fardusi MJ (2013) Status, diversity, and traditional uses of homestead gardens in Northern Bangladesh: A means of sustainable biodiversity conservation. *ISRN Biodiversity*, doi:org/10.1155/2013/124103.

Sahoo UK (2007) Agroforestry systems and practices prevailing in Mizoram. In: Puri S, Panwar P (Eds.). Agroforestry Systems and Practices. New India Publishing Agency, Pitam Pura, New Delhi, India. pp.367–383.

Salako VK, Fandohan B, Kassa B, Assogbadjo AE, Idohou AFR, Gbedomon RC, Chakeredza S, Dulloo ME, Kaka RG (2014) Home gardens: an assessment of their biodiversity and potential contribution

to conservation of threatened species and crop wild relatives in Benin. *Genetic Resources and Crop Evolution*, **61**, 313–330.

Sandhu HS, Wratten SD, Cullen R (2010) The role of supporting ecosystem services in conventional and organic arable farmland. *Ecological Complexity*, **7**, 302–310.

Schuh RT (2000) Biological Systematics: Principles and applications. Ithaca, NY7 Cornell University Press.

Schupp JL, Sharp JS (2012) Exploring the social bases of home gardening. *Agriculture and Human Values*, **29(1)**, 93–105.

Scora RW (1975) On the history and origin of *Citrus*. *Bulletin of the Torrey Botanical Club*, **102**, 369–375.

Scora RW (1988) Biochemistry, taxonomy and evolution of modern cultivated *Citrus*. *Proceeding of the International Society of Citriculture*, **1**, 277–289.

Scurrah M, Celis-Gamboa C, Chumbiauca S, Salas A, Visser RGF (2008) Hybridization between wild and cultivated potato species in the Peruvian Andes and biosafety implications for deployment of GM potatoes. *Euphytica*, **164**, 881–892.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Shackleton CM, Paumgarten F, Cocks ML (2008) Household attributes promote diversity of tree holdings in rural areas, South Africa. *Agroforestry Systems*, **72**, 221–230.

Shahsavar AR, Izadpanah K, Tafazoli E, Tabatabaei BES (2007) Characterization of *Citrus* germplasm including unknown variants by inter-simple sequence repeat (ISSR) markers. *Scientia Horticulturae*, **112**, 310–314.

Shannon EC, Weaver W (1963) The Mathematical Theory of Communication. Urbana, IL: University of Illinois Press.

Sharma BD, Hore DK, Gupta SG (2004) Genetic resources of *Citrus* of north-eastern India and their potential use. *Genetic Resources and Crop Evolution*, **51**, 411–418.

Shastri CM, Bhat DM, Nagaraja BC, Murali KS, Ravindranath NH (2002) Tree species diversity in a village ecosystem in Uttara Kannada district in Western Ghats, Karnataka. *Current Science*, **82**, 1080–1084.

Silvestro D, Michalak I (2012) RAxMLGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*, **12**, 335–337.

Simpson EH (1949) Measurement of diversity. *Nature*, **163**, 688.

Singh IP, Singh S (2003) Exploration, collection and mapping of citrus genetic diversity in India. Technical Bulletin No. 7, National Research Centre for Citrus, Nagpur, India.

Sites Jr JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, **18(9)**, 462-470. doi:10.1016/S0169-5347(03)00184-8.

Slatkin M (1981) Estimating levels of gene flow in natural populations. *Genetics,* **99**, 323–335.

Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution*, **39**, 53–65.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution,* **47**, 264–279.

Smith RM, Gaston KJ, Warren PH, Thompson K (2005) Urban domestic gardens (V): relationships between landcover composition, housing and landscape. *Landscape Ecology*, **20**, 235–253.

Sorensen T (1948) A method of estimating group of equal amplitude in plant society based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter,* **5**, 1–34.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Sunwar S, Thornstrom CG, Subedi A, Bystrom M (2006) Homegardens in western Nepal: opportunities and challenges for on-farm management of Agro-biodiversity. *Biodiversity and Conservation*, **15**, 4211–4238.

Swingle WT (1943) The botany of *Citrus* and its wild relatives. In: Webber HJ, Batchelor DL (Eds.). The Citrus Industry, Volume 1. University of California, Berkeley, pp. 128–474.

Swingle WT, Reece PC (1967) The botany of *Citrus* and its wild relatives. In: Reuther W, Batchelor LD, Webber HJ. (Eds.). The Citrus Industry. University of California Press, Berkeley, pp. 190–340.

Swinton SM, Lupi F, Robertson GP, Hamilton SK (2007) Ecosystem services and agriculture: cultivating agricultural ecosystems for diverse benefits. *Ecological Economics*, **64**, 245–252.

Swofford DL (2001) PAUP*: Phylogenetic analysis using parsimony. Version 4.0b8 for Macintosh. Sinauer, Sunderland, Massachusetts, USA.

Taberlet P, Gielly L, Patou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Tanaka T (1954) Species problem in Citrus. Japanese Society for promotion of Science, Ueno, pp. 152.

Tanaka T (1958) The origin and dispersal of *Citrus* fruits having their centre of origin in India. *Indian Journal of Horticulture*, **15**, 101–115.

Tanaka T (1969) Taxonomic problem of *Citrus* fruits in the Orient. *Bulletin of University Osaka Prefecture,* Series B*,* **21**, 133–138.

Tanaka T (1977) Fundamental discussion of *Citrus* classification. *Studia Citrologica,* 14, 1–6.

Taylor JR, Lovell ST (2014) Urban home gardens in the Global North: A mixed methods study of ethnic and migrant home gardens in Chicago, IL. *Renewable Agriculture and Food System*, doi:10.1017/S1742170514000180.

Templeton AR (1989) The meaning of species and speciation: A Genetic Perspective". In: Otte DJ and Endler JA (Eds.). Speciation and its Consequences. Sunderland, MA: Sinauer, pp. 3-27.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Tripolitsiotis C, Nikoloudakis N, Linos A, Hagidimitriou M (2013) Molecular characterization and analysis of the Greek *Citrus* germplasm. *Notulae Botanicae Horti Agrobotanici***, 41(2)**, 463–471.

Uzun A, Yesiloglu T, Aka-Kacar Y, Tuzcu, Gulsen O (2009) Genetic diversity and relationships within *Citrus* and related genera based on sequence related amplified polymorphism markers (SRAPs). *Scientia Horticulturae*, **121**, 306–312.

Vavilov N (1926) Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge, UK.

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants, Volume 13. Ronald Press, New York.

Viard F, Arnaud JF, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop-wild *Beta vulgaris* complex: Genetic distinctiveness vs. hot spots of hybridization over a regional scale. *Molecular Ecology*, **13**, 1357–1364.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wezel A, Bender S (2003) Plant species diversity of home gardens of Cuba and its significance for household food supply. *Agroforestry Systems*, **57**, 37–47.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Sninsky J, White T (Eds.). *PCR Protocols: A Guide to Methods and Applications*. Academic Press, Orlando, Florida, pp. 315–322.

Whittemore AT (1993) Species concepts: a reply to Ernst Mayr. *Taxon*, **42**, 573-583.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright S (1978) Evolution and the genetics of population. In: Variability within and among natural populations. Volume 4. University of Chicago Press, Chicago, USA.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J, Takita MA, et al (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, **32 (7)**, 656-663, doi:10.1038/nbt.2906

Yeh FC, Yang R, Boyle T (1999) POPGENE. Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis. University of Alberta, Edmonton, Alberta, Canada.

Zeven AC (1973) Dr. Th. H. Engelbrecht's views on the origin of cultivated plants. *Euphytica*, **22**, 279–286.

Zhang H, Jim CY (2014) Species diversity and performance assessment of trees in domestic gardens. *Landscape and Urban Planning*, **128**, 23–34.

Zhen-Hua Lu, Zhi-qin Z, Rang-jin X (2011) Molecular phylogeny of the “true Citrus fruit trees” group (Aurantioideae, Rutaceae) as inferred from chloroplast DNA sequence. *Agricultural Sciences in China*, **10(1)**, 49–57.

Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growing in the old World. *Science*, **187**, 319–327.

APPENDICES

Appendix 1: Aligned nucleotide sequences of the trnL-trnF gene

```
C.grandis           CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.medica            CCGGAAGGT-CTATACAATCTATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.limettioides      CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.limon             CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.aurantium         CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.ichangensis       CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.rugulosa          CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.reticulata        CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.pseudolimon       CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
A.marmelos          CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.assamensis        CCGTATGGTACTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 60
C.jambhiri          CCGTATGGT-CGAGACGATATATGTAAAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.indica            CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.paradisi          CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.latipes           CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.macroptera        CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.reshni            CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.karna             CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGATCAAGGAATCCCCGTTT 59
C.volkameriana      CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGATCAAGGAATCCCCGTTT 59
C.nobilis           CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGATCAAGGAATCCCCGTTT 59
C.sinensis          CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.megaloxycarpa     CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.limonia           CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
C.aurantifolia      CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
P.trifoliata        CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
M.paniculata        CCGTATGGT-CTATACGATATATGTAGAAATGAACACCTTTGAGCAAGGAATCCCCGTTT 59
                    *** * *** * * ** ** ****** ********** ********** **********

C.grandis           TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.medica            TAATGATTCCCAATCCATATCATTGCTCAGACTGAAACTTACCAAGTCTTCTTTTTGATG 119
C.limettioides      TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.limon             TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.aurantium         TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.ichangensis       TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.rugulosa          TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.reticulata        TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.pseudolimon       TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
A.marmelos          TAATGATTCCCAATCCATATTATTGCTCATACTTAAACTTACAAAGTCTTCTTTTTGATG 119
C.assamensis        TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 120
C.jambhiri          TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTAATG 119
C.indica            TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTAATG 119
C.paradisi          TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.latipes           TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.macroptera        TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.reshni            TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.karna             TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.volkameriana      TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.nobilis           TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.sinensis          TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.megaloxycarpa     TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.limonia           TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
C.aurantifolia      TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
P.trifoliata        TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
M.paniculata        TAATGATTCCCAATCCATATTATTGCTCATACTGAAACTTACAAAGTCTTCTTTTTGATG 119
                    ********************* ******** *** ******** ************* ***

C.grandis           ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.medica            ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.limettioides      ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.limon             ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.aurantium         ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.ichangensis       ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.rugulosa          ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.reticulata        ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
C.pseudolimon       ATTCAAGAAATGAAATTACCCTCCCAAGACTTTGAATCCCTGTTTATTTTTTAATTGACA 179
A.marmelos          ATTCAAGAAAAGAAATTCCCCTCCCAAGACTTTGAATCCCTTTTTCTTTTTTAATTGACA 179
C.assamensis        ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 180
C.jambhiri          ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
```

```
C.indica            ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.paradisi          ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.latipes           ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.macroptera        ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.reshni            ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.karna             ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.volkameriana      ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.nobilis           ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.sinensis          ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.megaloxycarpa     ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.limonia           ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
C.aurantifolia      ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTGTTTATTTTTTAATTGACA 179
P.trifoliata        ATTCAAGAAATGAAATTCCCCTCCCAAGACTTTTAATCCCTATTTATTTTTTAATTGACA 179
M.paniculata        ATTCAAGAAATGAAATTCCCCTCCCAAGAATTTTAATCCCA------TTTCCAATTAACA 173
                    ********** ****** *********** *** ******       ***  **** *** 

C.grandis           TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.medica            TAGACCCAAGTCATCTTGTAAGATGAAAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.limettioides      TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.limon             TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.aurantium         TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.ichangensis       TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.rugulosa          TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.reticulata        TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.pseudolimon       TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
A.marmelos          TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.assamensis        TAGACCCAACTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 240
C.jambhiri          TAGACCCAAGTCATCTAATAAGATGAGAACGGTGTGTCGGAAATGGGCGGGATAGCTCAA 239
C.indica            TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.paradisi          TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.latipes           TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.macroptera        TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.reshni            TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.karna             TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.volkameriana      TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.nobilis           TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.sinensis          TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.megaloxycarpa     TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.limonia           TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
C.aurantifolia      TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
P.trifoliata        TAGACCCAAGTCATCTAGTAAGATGAGAACGGTGTGTCGGAAATGGTCGGGATAGCTCAG 239
M.paniculata        AAGAACCAAGCAATCTAGTAAAATGAGAAACGTCCCGCGGAAAACGCCAGGATAGCTCAG 233
                     *** ****   ****  *** **** **  **    ******  * * **********

C.grandis           CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.medica            CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.limettioides      CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.limon             CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.aurantium         CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.ichangensis       CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.rugulosa          CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.reticulata        CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.pseudolimon       CTGGTAGAGCAGAGGACTGAAAATCCT 266
A.marmelos          CT-------------------------- 241
C.assamensis        CTGGTAGAGCACAGGACTGAAAATCCT 267
C.jambhiri          CTGGTAGAGCAAAGGACTGAAAATCCT 266
C.indica            CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.paradisi          CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.latipes           CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.macroptera        CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.reshni            CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.karna             CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.volkameriana      CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.nobilis           CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.sinensis          CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.megaloxycarpa     CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.limonia           CTGGTAGAGCAGAGGACTGAAAATCCT 266
C.aurantifolia      CTGGTAGAGCAGAGGACTGAAAATCCT 266
P.trifoliata        CTGGTAGAGCAGAGGACTGAAAATCCT 266
M.paniculata        CTGGTAGAGCAGAGGACTGAAAATCCT 260
                    **
```

Appendix 2: Aligned nucleotide sequences of the trnS-trnG gene

```
C.latipes            AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.indica             AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.megaloxycarpa      AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.limettioides       AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.rugulosa           AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.aurantifolia       AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.limon              AAACCGAACATGAAACTTTGGGGTCATTCGGCTCCCTTTATGGAAGTTTGCTATATTCCC 60
C.volkameriana       AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.ichangensis        AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.grandis            AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.assamensis         AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.jambhiri           AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.paradisi           AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.aurantium          AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.medica             AAACCGAACTTGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
P.trifoliata         AAACCGAACTTGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.reshni             AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.reticulata         AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.macroptera         AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.nobilis            AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.sinensis           AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.karna              AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
C.limonia            AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
M.paniculata         AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGATTGATATATTCCC 59
C.pseudolimon        AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGCTATATTCCC 59
A.marmelos           AAACCGAACATGAAACTTTGGGGTCATTCGGCTCC-TTTATGGAAGTTTGATATATTCCC 59
                     ********* ************************ ********** *** *********

C.latipes            AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.indica             AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.megaloxycarpa      AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.limettioides       AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.rugulosa           AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.aurantifolia       AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.limon              AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 120
C.volkameriana       AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.ichangensis        AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.grandis            AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.assamensis         AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.jambhiri           AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.paradisi           AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.aurantium          AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.medica             AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
P.trifoliata         AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.reshni             AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.reticulata         AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.macroptera         AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.nobilis            AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.sinensis           AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.karna              AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.limonia            AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
M.paniculata         AGAGAGAAGCGGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
C.pseudolimon        AGAGAGAAGCCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
A.marmelos           CGAGAGAAACCGGGAGCAAAATAACAAAAGTCGACCCATAACATCTATGTCAGCTTTTTT 119
                      ******* * ************************************************

C.latipes            TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.indica             TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.megaloxycarpa      TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.limettioides       TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.rugulosa           TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.aurantifolia       TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.limon              TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 180
C.volkameriana       TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.ichangensis        TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.grandis            TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.assamensis         TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.jambhiri           TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.paradisi           TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.aurantium          TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.medica             TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
```

```
P.trifoliata          TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.reshni              TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.reticulata          TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.macroptera          TGTCTGAATGAAATCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.nobilis             TGTCTGAATGAAATCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.sinensis            TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.karna               TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
C.limonia             TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
M.paniculata          TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGACGATCCCTCTAGAAGAGTGGGATTA 179
C.pseudolimon         TGTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAAGAGTGGGATTA 179
A.marmelos            -GTCTGAATGAATTCAAAACAATCCGCTTTCTAGATGATCCCTCTAGAANAGTGGGATTA 178
                       *********** ********************** ************ **********

C.latipes             TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.indica              TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.megaloxycarpa       TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.limettioides        TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.rugulosa            TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.aurantifolia        TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.limon               TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 240
C.volkameriana        TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.ichangensis         TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.grandis             TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.assamensis          TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.jambhiri            TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.paradisi            TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.aurantium           TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.medica              TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
P.trifoliata          TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.reshni              TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.reticulata          TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.macroptera          TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.nobilis             TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.sinensis            TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.karna               TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
C.limonia             TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
M.paniculata          TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAGAGAATCCTTAGG 239
C.pseudolimon         TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAAAGAATCCTTAGG 239
A.marmelos            TAACAATACCAATCTTTCTAGTTACTTCGTTCTCTATTTCTATTTGAGAGAATCCTTAGG 238
                      ********************************************** ************

C.latipes             AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.indica              AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.megaloxycarpa       AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.limettioides        AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.rugulosa            AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.aurantifolia        AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.limon               AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 294
C.volkameriana        AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.ichangensis         AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.grandis             AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.assamensis          AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.jambhiri            AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.paradisi            AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.aurantium           AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.medica              AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
P.trifoliata          AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
C.reshni              AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.reticulata          AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.macroptera          AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.nobilis             AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.sinensis            AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.karna               AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
C.limonia             AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 293
M.paniculata          AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATGAATATGTTGATGTC 299
C.pseudolimon         AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATACG------TTGATGTC 293
A.marmelos            AAAAGTTTTTTGTTTCCCCCGAGCTAAACTAAAAAAATAGAATATG------TTGATGTC 292
                      ******************************************** *      ********

C.latipes             TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.indica              TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.megaloxycarpa       TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.limettioides        TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.rugulosa            TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.aurantifolia        TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.limon               TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 354
C.volkameriana        TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
```

```
C.ichangensis        TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.grandis            TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.assamensis         TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.jambhiri           TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.paradisi           TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.aurantium          TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.medica             TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
P.trifoliata         TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.reshni             TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.reticulata         TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.macroptera         TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.nobilis            TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.sinensis           TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.karna              TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
C.limonia            TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
M.paniculata         TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 359
C.pseudolimon        TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAC 353
A.marmelos           TCTAGTAAACTAGAGTGATCATTTAATAGCTATTTTGCTTCAATCTAACCTATAAAAAAA 352
                     ************************************************** 

C.latipes            ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.indica             ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.megaloxycarpa      ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.limettioides       ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.rugulosa           ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.aurantifolia       ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.limon              ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 404
C.volkameriana       ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.ichangensis        ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.grandis            ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.assamensis         ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.jambhiri           ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.paradisi           ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.aurantium          ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.medica             ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
P.trifoliata         ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.reshni             ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.reticulata         ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.macroptera         ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.nobilis            ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.sinensis           ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.karna              ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
C.limonia            ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
M.paniculata         ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 409
C.pseudolimon        ---CAAAATTGAAGATTTAGTTACGATTCGAACTAGA-------CTTTTCTATCTTTATC 403
A.marmelos           AAACAAAATTGAAGATTTAGTTACGATTCGAACTAGAAACTAGACTTTTCTATCTTTATC 412
                        ********************************       ***************** 

C.latipes            CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.indica             CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.megaloxycarpa      CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.limettioides       CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.rugulosa           CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.aurantifolia       CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.limon              CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 463
C.volkameriana       CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
C.ichangensis        CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
C.grandis            CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
C.assamensis         CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
C.jambhiri           CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.paradisi           CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAACAAAAA 463
C.aurantium          CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.medica             CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
P.trifoliata         CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
C.reshni             CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.reticulata         CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAC-AAAA 462
C.macroptera         CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAACAAAAA 463
C.nobilis            CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAACAAAAA 463
C.sinensis           CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAACAAAAA 463
C.karna              CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAACAAAAA 463
C.limonia            CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
M.paniculata         CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAAAAAAAA 469
C.pseudolimon        CATGGATCCTTTACTTATACTTAAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 461
A.marmelos           CATGGATCCTTTACTTATACTGCAAAAATTGGAAGTATTGATCCAATTCAAAAA--AAAA 470
                     ********************  ******************************  **** 

C.latipes            AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTGGGA 522
```

```
C.indica            AACAAAA--A------------------------------------------------- 470
C.megaloxycarpa     AACAAAA---------------------------------------------------- 469
C.limettioides      AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 522
C.rugulosa          AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 522
C.aurantifolia      AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 522
C.limon             AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 523
C.volkameriana      AACAAAATTATGTTTC-------------------------------------------- 477
C.ichangensis       AACAAAAT--------------------------------------------------- 469
C.grandis           AACAAAA---------------------------------------------------- 468
C.assamensis        AACAAAA---------------------------------------------------- 468
C.jambhiri          AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 522
C.paradisi          AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 523
C.aurantium         AACAAAA---------------------------------------------------- 469
C.medica            AACAAAA---------------------------------------------------- 468
P.trifoliata        AACAAAA---------------------------------------------------- 468
C.reshni            AACAAAA--A--------AAT-------------------------------------- 473
C.reticulata        AACAAAA--A--------AA--------------------------------------- 472
C.macroptera        AACAAAA---------------------------------------------------- 470
C.nobilis           AACAAAA---------------------------------------------------- 470
C.sinensis          AACAAAAT--------------------------------------------------- 471
C.karna             AACAAAA---------------------------------------------------- 470
C.limonia           AACAAAA---------------------------------------------------- 468
M.paniculata        AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 529
C.pseudolimon       AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 521
A.marmelos          AACAAAATTATGTTTCGCAATTTCATAATCCAAATTGTCAATTTCGAATTGACTCTTGGA 530
                    *******
C.latipes           TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 582
C.indica            ------------------------------------------------------------
C.megaloxycarpa     ------------------------------------------------------------
C.limettioides      TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 582
C.rugulosa          TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 582
C.aurantifolia      TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 582
C.limon             TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTAAGAGTGAAAG 583
C.volkameriana      ------------------------------------------------------------
C.ichangensis       ------------------------------------------------------------
C.grandis           ------------------------------------------------------------
C.assamensis        ------------------------------------------------------------
C.jambhiri          TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 582
C.paradisi          TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTTATTTAAATTTGAGAGTGAAAG 583
C.aurantium         ------------------------------------------------------------
C.medica            ------------------------------------------------------------
P.trifoliata        ------------------------------------------------------------
C.reshni            ------------------------------------------------------------
C.reticulata        ------------------------------------------------------------
C.macroptera        ------------------------------------------------------------
C.nobilis           ------------------------------------------------------------
C.sinensis          ------------------------------------------------------------
C.karna             ------------------------------------------------------------
C.limonia           ------------------------------------------------------------
M.paniculata        TACAAATCACGAGAACGGATATTTTTCCCCGAATCTTTTATTTTCATTTTAGAGTGAAAG 589
C.pseudolimon       TACAAATCAC------------------------------------------------- 531
A.marmelos          TACAAATCACGAGAACGGATATTTTTCTCCGAATCTTTT------ATTTTAGAGTGAAAG 584

C.latipes           GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATATCAATTAAACTGAA 642
C.indica            ------------------------------------------------------------
C.megaloxycarpa     ------------------------------------------------------------
C.limettioides      GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATATCAATTAAACTGAA 642
C.rugulosa          GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATATCAATTAAACTGAA 642
C.aurantifolia      GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATATCAATTAAACTGAA 642
C.limon             GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATAT------------- 630
C.volkameriana      ------------------------------------------------------------
C.ichangensis       ------------------------------------------------------------
C.grandis           ------------------------------------------------------------
C.assamensis        ------------------------------------------------------------
C.jambhiri          GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATAT------------- 629
C.paradisi          GATTAAAATTGAATCCTTTTAAGAAATAAAGTTTTTGATTGAAATATCAATTAAACTGAA 643
C.aurantium         ------------------------------------------------------------
C.medica            ------------------------------------------------------------
P.trifoliata        ------------------------------------------------------------
C.reshni            ------------------------------------------------------------
C.reticulata        ------------------------------------------------------------
C.macroptera        ------------------------------------------------------------
C.nobilis           ------------------------------------------------------------
C.sinensis          ------------------------------------------------------------
```

```
C.karna             ------------------------------------------------------------
C.limonia           ------------------------------------------------------------
M.paniculata        GATTCAAATTTAATCCTTTTAAGAAATAAAGTTTTTGATTGGAATATCAATTAAACTGAA 649
C.pseudolimon       ------------------------------------------------------------
A.marmelos          GATTCAAATTGAATCCTTTTTAGAAATAAAGTTTTTGATTGGAATATCAATTAAACTGAA 644

C.latipes           GGACCCCTTAACTA 656
C.indica            --------------
C.megaloxycarpa     --------------
C.limettioides      GGACCCCTTAACTA 656
C.rugulosa          GGACCCCTTAACTA 656
C.aurantifolia      GGACCCCTTAACTA 656
C.limon             --------------
C.volkameriana      --------------
C.ichangensis       --------------
C.grandis           --------------
C.assamensis        --------------
C.jambhiri          --------------
C.paradisi          GGACCCCTTAACTA 657
C.aurantium         --------------
C.medica            --------------
P.trifoliata        --------------
C.reshni            --------------
C.reticulata        --------------
C.macroptera        --------------
C.nobilis           --------------
C.sinensis          --------------
C.karna             --------------
C.limonia           --------------
M.paniculata        GGACCCCTTAACTA 663
C.pseudolimon       --------------
A.marmelos          GGACCCCTTAACTA 658
```

Appendix 3. Aligned nucleotide sequences of the rps16 gene

```
C.medica            ---------------AATGGGCTCTTGGCTCGACATTTTTTCTTCGGTTCTATTAGAATC 45
C.indica            -------------------GGCTCTTGGCTCGACATTTTTTCTTCGGTTCTATTAGAATC 41
C.reshni            TTTCTATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCAGTTCTATTAGAATC 60
C.nobilis           TTTCTATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCAGTTCTATTAGAATC 60
C.aurantium         TTTCTATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCAGTTCTATTAGAATC 60
C.reticulata        TTTCTATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCAGTTCTATTAGAATC 60
C.sinensis          TTTCTATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCAGTTCTATTAGAATC 60
C.grandis           TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.ichangensis       ----------------ATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 44
C.latipes           TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.jambhiri          TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.paradisi          TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.megaloxycarpa     ---------------AATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 45
C.rugulosa          TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.karna             TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.limon             TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.pseudolimon       TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.limonia           TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.limettioides      TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.aurantifolia      TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.volkameriana      TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
P.trifoliata        TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.assamensis        TTTCGATTTTATATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
C.macroptera        ---------------AATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 45
M.paniculata        TTTCGATTTTCTATGAATGGGCTCTTGGCTCGACATTTTTTCTTCTGTTCTATTAGAATC 60
A.marmelos          TTTCGATTTTATATGAAAGGGCTCTTGGCTCGACATTTTTTCTTATGTTCTATTAGAATC 60
                                      ************************  **************
C.medica            CTCAAGTTTTTTT-GGGGAGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 103
C.indica            CTCAAGTTTTTTT-GGGGAGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 99
C.reshni            CTCAAGTTTTTT--GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.nobilis           CTCAAGTTTTTT--GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.aurantium         CTCAAGTTTTTT--GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.reticulata        CTCAAGTTTTTT--GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.sinensis          CTCAAGTTTTTT--GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.grandis           CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.ichangensis       CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 102
```

```
C.latipes            CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.jambhiri           CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.paradisi           CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.megaloxycarpa      CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 103
C.rugulosa           CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.karna              CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.limon              CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.pseudolimon        CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.limonia            CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.limettioides       CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.aurantifolia       CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.volkameriana       CTCAAGTTTTTTT-GGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 119
P.trifoliata         CTCAAGTTTTTTTTGGGGGGGGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 120
C.assamensis         CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
C.macroptera         CTCAAGTTTTTTT-GGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 103
M.paniculata         CTCAAGTTTTTTTTGGGGGGG-TAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 119
A.marmelos           CTCAAGTTTTTT--GGGGGGCGTAATGGAACTAGTACAGGATGGAGCTCGAGTATAAAGT 118
                     ************  **** *  ************************************

C.medica             ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 162
C.indica             ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 158
C.reshni             ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.nobilis            ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.aurantium          ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.reticulata         ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.sinensis           ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.grandis            ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.ichangensis        ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 161
C.latipes            ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.jambhiri           ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.paradisi           ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.megaloxycarpa      ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 162
C.rugulosa           ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.karna              ATTTATTCCTTTCTCAGGGGCACGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.limon              ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.pseudolimon        ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.limonia            ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.limettioides       ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.aurantifolia       ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.volkameriana       ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 178
P.trifoliata         ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 179
C.assamensis         ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
C.macroptera         ATTTATTCCTTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAACTAAGTTGGAAAA 163
M.paniculata         ATTTATTCATTTCTCAGGGGCAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 178
A.marmelos           ATTTATTCATTTCTCAGGGGTAAGGATCTAGGGTTAATACCAATCAA-TAAGTTGGAAAA 177
                     ******** *********** * ************************ ************

C.medica             AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 221
C.indica             AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 217
C.reshni             AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.nobilis            AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.aurantium          AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.reticulata         AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.sinensis           AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.grandis            AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.ichangensis        AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 220
C.latipes            AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.jambhiri           AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.paradisi           AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.megaloxycarpa      AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 221
C.rugulosa           AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.karna              AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.limon              AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.pseudolimon        AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.limonia            AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.limettioides       AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.aurantifolia       AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.volkameriana       AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 237
P.trifoliata         AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 238
C.assamensis         AACTTCGT-AAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
C.macroptera         AACTTCGTCAAGTAAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 223
M.paniculata         AACTTCGT-AAGTCAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 237
A.marmelos           AACTTCGT-AAGTCAATTTTCGATATAGAAATCGAAAGGATCTGATTCGAGCAATTTTGA 236
                     ******** **** ************************************

C.medica             AATCCAAAAGCAAGGGGAAATTTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 281
C.indica             AATCCAAAAGCAAGGGGAAATTTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 277
```

```
C.reshni              AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.nobilis             AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.aurantium           AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.reticulata          AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.sinensis            AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.grandis             AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.ichangensis         AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 280
C.latipes             AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.jambhiri            AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.paradisi            AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.megaloxycarpa       AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 281
C.rugulosa            AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.karna               AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.limon               AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.pseudolimon         AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.limonia             AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.limettioides        AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.aurantifolia        AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.volkameriana        AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 297
P.trifoliata          AATCCAAAAGCAAGGGTAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 298
C.assamensis          AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
C.macroptera          AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 283
M.paniculata          AATCCAAAAGCAAGGGGAAATTTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 297
A.marmelos            AATCCAAAAGCAAGGGGAAAATTTGTTGGAATTGGGAAAACTTTTTCTACCAAAAGTGTA 296
                      ********************************************************

C.medica              TCCTGTAGGAATCAATTGTTCCTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 341
C.indica              TCCTGTAGGAATCAATTGTTCCTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 337
C.reshni              TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.nobilis             TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.aurantium           TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.reticulata          TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.sinensis            TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.grandis             TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.ichangensis         TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 340
C.latipes             TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.jambhiri            TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.paradisi            TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.megaloxycarpa       TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGT- 340
C.rugulosa            TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.karna               TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.limon               TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 356
C.pseudolimon         TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 356
C.limonia             TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 356
C.limettioides        TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 356
C.aurantifolia        TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 356
C.volkameriana        TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 357
P.trifoliata          TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAAAAAGAAATCAAAAGGGGGTGTG 358
C.assamensis          TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 356
C.macroptera          TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 343
M.paniculata          TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGGGTGTG 357
A.marmelos            TCCTGTAGGAATCAATTGTTCGTATGATTCTTTGATAGAAAGAAATCAAAAGGGG-TGTG 355
                      ********************* *************** ***************** ***

C.medica              TTGCTGCCATTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 389
C.indica              TTGCTGCCATTTTTAAAAT------AAAAAAAAAAAACGTTAAAGATCACCGAAGTAAT 391
C.reshni              TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.nobilis             TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.aurantium           TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.reticulata          TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.sinensis            TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.grandis             TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.ichangensis         TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 388
C.latipes             TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.jambhiri            TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.paradisi            TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.megaloxycarpa       TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 388
C.rugulosa            TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.karna               TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.limon               TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.pseudolimon         TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.limonia             TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.limettioides        TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.aurantifolia        TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.volkameriana        TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 405
P.trifoliata          TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 406
```

```
C.assamensis          TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 404
C.macroptera          TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 391
M.paniculata          TTGCTGCCATTTTTTAAAAT------AAAAAA------CGTTAAAGATCACCGAAGTAAT 405
A.marmelos            TTGCTGCCATTTTTTAAAATTCAAATAAAAAA------CGTTAAAGATCACCGAAGTAAT 409
                      ************** *****      ******      ********************

C.medica              GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 449
C.indica              GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 451
C.reshni              GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.nobilis             GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.aurantium           GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.reticulata          GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.sinensis            GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.grandis             GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.ichangensis         GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 448
C.latipes             GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.jambhiri            GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.paradisi            GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.megaloxycarpa       GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 448
C.rugulosa            GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.karna               GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.limon               GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.pseudolimon         GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.limonia             GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.limettioides        GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.aurantifolia        GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.volkameriana        GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 465
P.trifoliata          GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 466
C.assamensis          GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 464
C.macroptera          GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 451
M.paniculata          GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 465
A.marmelos            GTCTAAACCTAATGATTCAAAGCAAAGATAAAGGATCCTGGAACAAGGAAATACCATTTT 469
                      ************************************************************

C.medica              TCAATTGTCTCAACAATTCAATCCAATCCAAAAATCGATTCGATTCGAAACGAGACAAAC 509
C.indica              TCAATTGT-TCAACAATTCAATCCAATCCAAAAATCGATTCGATTCGAAACGAGA-AAAC 509
C.reshni              TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.nobilis             TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.aurantium           TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.reticulata          TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.sinensis            TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.grandis             TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.ichangensis         TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 508
C.latipes             TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.jambhiri            TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.paradisi            TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.megaloxycarpa       TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 508
C.rugulosa            TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.karna               TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.limon               TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.pseudolimon         TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.limonia             TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.limettioides        TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.aurantifolia        TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.volkameriana        TCAATTGTCTCAACAATTCAATCCAATCAAAAAATAGATTCGATTCGAAACGAGACAAAC 525
P.trifoliata          TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 526
C.assamensis          TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 524
C.macroptera          TCAATTGTCTCAACAATTCAATCCAATCCAAAAATAGATTCGATTCGAAACGAGACAAAC 511
M.paniculata          TCAATTGGCTCAACAATTCAATCCAATCCAAAAATCGATTCGATTCGAAACGAGACAAAC 525
A.marmelos            TCAATTGTCTCAACAATTCAATCCAATCCAAAAATCGATTCGATTCGAAACGAGACAAAC 529
                      *******  ****************** ****** ****************** ****

C.medica              AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGT- 568
C.indica              AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGT- 568
C.reshni              AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.nobilis             AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.aurantium           AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.reticulata          AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.sinensis            AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.grandis             AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.ichangensis         AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 568
C.latipes             AAAAAAGGGGCTTGAGACCACTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.jambhiri            AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.paradisi            AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.megaloxycarpa       AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 568
C.rugulosa            AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.karna               AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
```

```
C.limon             AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.pseudolimon       AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.limonia           AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.limettioides      AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.aurantifolia      AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.volkameriana      AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 585
P.trifoliata        AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 586
C.assamensis        AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 584
C.macroptera        AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCGTT 571
M.paniculata        AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCTTT 585
A.marmelos          AAAAAAGGGGCTTGAGACCGCTCAAAAAAGGAAATGCCTAAGGATTTTCGGCTGGGCTTT 589
                    ****************** ************************************ *

C.medica            ------------------------------------------------------------
C.indica            ------------------------------------------------------------
C.reshni            GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.nobilis           GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTG------ 638
C.aurantium         GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.reticulata        GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.sinensis          GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.grandis           GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.ichangensis       GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 628
C.latipes           GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.jambhiri          GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.paradisi          GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.megaloxycarpa     GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTG------ 622
C.rugulosa          GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTG------ 638
C.karna             GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.limon             GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.pseudolimon       GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.limonia           GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.limettioides      GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.aurantifolia      GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 644
C.volkameriana      GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 645
P.trifoliata        GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTGAAAATG 646
C.assamensis        GAAACTTATCTAACTTGAGTTATGAGAGTACGAATTCTTTTTTTGTTTCTTTTG------ 638
C.macroptera        GAAACTTATCTAACTTGAGTTATG------------------------------------ 595
M.paniculata        GAAACTTATCTAACTTGAGTTATGAGATTACGAATGCTTTTTTTGTTTCTTTTGAAAATG 645
A.marmelos          GAAACTTATCTAACTTGAGTTATGAGAGTACGAATGCTTTTTTTGTTTCTTTTGAAAATG 649

C.medica            ------------------------------------------------------------
C.indica            ------------------------------------------------------------
C.reshni            ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.nobilis           ------------------------------------------------------------
C.aurantium         ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.reticulata        ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.sinensis          ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.grandis           ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.ichangensis       ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTA------- 681
C.latipes           ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.jambhiri          ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.paradisi          ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.megaloxycarpa     ------------------------------------------------------------
C.rugulosa          ------------------------------------------------------------
C.karna             ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.limon             ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.pseudolimon       ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.limonia           ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.limettioides      ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.aurantifolia      ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 704
C.volkameriana      ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 705
P.trifoliata        ACAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 706
C.assamensis        ------------------------------------------------------------
C.macroptera        ------------------------------------------------------------
M.paniculata        AAAAAGAAACAAAAAAAGAATAACTTAAATCTCTAATTGATTTGATTATTTTATAGATCT 705
A.marmelos          AAAAAGAAACAAAAAAGAATAACTTCAATTTCGAATTGATTTTATTATTTTAGAGATCTA 709
```

Appendix 4. Aligned nucleotide sequences of the ITS2 gene

```
C.macroptera          ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
P.trifoliata          ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.medica              ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.karna               ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.nobilis             AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.pseudolimon         ACCATCGAGTCTTTG-AACGCAAGGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTCT 58
A.marmelos            ACCATTGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CTGTTAGGCCGAGGGCACGTC- 56
C.reticulata          ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.rugulosa            ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.ichangensis         AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.reshni              ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTCC 57
C.limettioides        ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.limon               ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.aurantifolia        ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.megaloxycarpa       ACCATCAATTCTTTGCAACGACAAGTTGCGCCCCAAGGCCATTAGGCCGAGGGCACGTC- 59
C.aurantium           ACCATCGAT-CTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.latipes             ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.jambhiri            AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.sinensis            AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.volkameriana        AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.paradisi            AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.limonia             ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
C.indica              --CATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 54
C.grandis             AC-ATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 55
C.assamensis          ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACGTC- 56
M.paniculata          ACCATCGAGTCTTTG-AACG-CAAGTTGCGCCCCAAG-CCATTAGGCCGAGGGCACATC- 56
                         **  *  ***** ****  * ************ *  *************** **

C.macroptera          TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
P.trifoliata          TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.medica              TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.karna               TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.nobilis             TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.pseudolimon         TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCCCAAAACCAAGGCGGGGGCCC 118
A.marmelos            TGCCTGGGTGTCACGCACCGTCGCCCCACCCCACCCCCCTTCGGACCGAGGCGGGGGCCC 116
C.reticulata          TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.rugulosa            TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.ichangensis         TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.reshni              TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 115
C.limettioides        TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.limon               TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.aurantifolia        TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.megaloxycarpa       TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 117
C.aurantium           TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.latipes             TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.jambhiri            TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.sinensis            TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAAACCAAGGCGGGGGCCC 114
C.volkameriana        TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.paradisi            TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.limonia             TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
C.indica              TGCCTGGGTGTCACGCATCGTTGCCTCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 112
C.grandis             TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 113
C.assamensis          TGCCTGGGTGTCACGCATCGTTGCCCCACCCCACCCCCCC-AAA-CCAAGGCGGGGGCCC 114
M.paniculata          TGCCTGGGTGTCACGTATCGTCGCCCCTCCCCACCCCTCT-----------------TCG 99
                      *************** * *** *** * ********* *                    *

C.macroptera          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
P.trifoliata          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.medica              CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.karna               CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.nobilis             CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.pseudolimon         CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 176
A.marmelos            CGAAGGTGCGGG-TGGACATTGGCCTCCCGTGTGCTGGCCGCTCGCGGTTGGCCCAAATC 175
C.reticulata          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.rugulosa            CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.ichangensis         CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.reshni              CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 173
C.limettioides        CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.limon               CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.aurantifolia        CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.megaloxycarpa       CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 175
C.aurantium           CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.latipes             CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
```

```
C.jambhiri          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.sinensis          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
C.volkameriana      CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.paradisi          CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.limonia           TGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACTGCTCGCGGTTGGCCCAAATA 172
C.indica            CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 170
C.grandis           CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 171
C.assamensis        CGG-GGTGCGGG-CGGAGATTGGCCTCCCGTGCGCTGACCGCTCGCGGTTGGCCCAAATA 172
M.paniculata        CGG-AGTGCGGGGCGAAAAATGGCCTCCCGTGCGCAACTCGCTCGCGGTTGGCCCAAACA 158
                     *   *******  * * * ************ **     ******************

C.macroptera        TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
P.trifoliata        TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.medica            TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.karna             TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-ACCTCTCGAGC 228
C.nobilis           TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-AGCTCTCGAGC 227
C.pseudolimon       TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGT--AAACAAAAGCCTCTCGAGC 232
A.marmelos          CGAGTCCTCGGCGGCCGAAGC--CGCGGCGATTGGTGGCG-AAAGAAAAGCCTCTCGAGC 232
C.reticulata        TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAAAGCCTCTCGAGC 229
C.rugulosa          TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAAAGCCTCTCGAGC 229
C.ichangensis       TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAAAGCCTCTCGAGC 228
C.reshni            TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAAAGCCTCTCGAGC 230
C.limettioides      TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.limon             TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.aurantifolia      TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.megaloxycarpa     TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 231
C.aurantium         TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 227
C.latipes           TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.jambhiri          TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 227
C.sinensis          TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.volkameriana      TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 227
C.paradisi          TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 227
C.limonia           TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
C.indica            TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAT-GCCTCTCGAGC 226
C.grandis           TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 227
C.assamensis        TGAGTCCTCGGCGACCGAAGC--CGCGGCGATCGGTGGTG-AAACAAA-GCCTCTCGAGC 228
M.paniculata        TGAGTCCCAGGCGACCAGAGCGCCGCGACGATCGGTGGTGTGTCCTTATGCTCGTCG--- 215
                     ******  **** **  ***  **** **** *****                ***

C.macroptera        TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
P.trifoliata        TCCCGCCGCGC---CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 281
C.medica            TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.karna             TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.nobilis           TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.pseudolimon       TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 287
A.marmelos          TACCGCCACGCGC-CCGGTCTCCGCAAGG---GGACCCCATGACCCCAACGC-TCCACGC 287
C.reticulata        TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 284
C.rugulosa          TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 284
C.ichangensis       TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.reshni            TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGGCCCTGAAGC-TCCGCGC 285
C.limettioides      TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.limon             TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.aurantifolia      TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.megaloxycarpa     TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 286
C.aurantium         TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.latipes           TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.jambhiri          TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.sinensis          TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.volkameriana      TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.paradisi          TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.limonia           TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
C.indica            TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 281
C.grandis           TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 282
C.assamensis        TCCCGCCGCGCGC-CCGGTCTCC---AAGTGTGGACTCTGCGACCCTGAAGC-TCCGCGC 283
M.paniculata        -CGCGTCGCGCGCGCCGGTCGCCCTTAGGGATG--CCTCGAGACCCTTAAGCGTCCCCTC 272
                       ** * ***   ****** **   * *   *  *     * ***  * ** *** * *

C.macroptera        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
P.trifoliata        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGGGATTACCCGCCTGAGTTTAAGCA 339
C.medica            -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.karna             -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.nobilis           -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGCTTGAGTTTAA-CA 338
C.pseudolimon       -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGT-ATTACCCGC-TGAGTTTAAGCA 343
A.marmelos          -AAGGGCAGCTCGCGCCGCGACCCCAGGTCAGGCAGG-ATCACCCGC-TGAGTTTAAGCA 344
C.reticulata        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 340
C.rugulosa          -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 340
C.ichangensis       -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
```

```
C.reshni              -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 341
C.limettioides        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.limon               -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.aurantifolia        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.megaloxycarpa       -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 342
C.aurantium           -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 338
C.latipes             -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.jambhiri            -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 338
C.sinensis            -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.volkameriana        -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 338
C.paradisi            -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 338
C.limonia             -AAGGGC-GCTCGCATTGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.indica              -AAGGGC-GCTCGCATCGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 337
C.grandis             CAACGGC-GCTCGCATCGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 339
C.assamensis          CAACGGC-GCTCGCATCGCGACCCCAGGTCAGGCGGG-ATTACCCGC-TGAGTTTAAGCA 340
M.paniculata          GAA-GAC-GCTCGCATCGCGACCCCAGGTCAGGTGGG-ACTACCCGC-TGAGTTTAAGCA 328
                       ** * * ******   ***************  *  *  ****** ********* **

C.macroptera          TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
P.trifoliata          TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.medica              TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.karna               TATTCAATAA--CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGGCAGCGAAC 397
C.nobilis             TATTCAATAA--CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 396
C.pseudolimon         TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 401
A.marmelos            TAT-CA------------------------------------------------------ 349
C.reticulata          TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 398
C.rugulosa            TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 398
C.ichangensis         TAT-CAATAAGCCGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 398
C.reshni              TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 399
C.limettioides        TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.limon               TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.aurantifolia        TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.megaloxycarpa       TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 400
C.aurantium           TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 396
C.latipes             TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.jambhiri            TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 396
C.sinensis            TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.volkameriana        TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 396
C.paradisi            TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 396
C.limonia             TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.indica              TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 395
C.grandis             TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 397
C.assamensis          TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 398
M.paniculata          TAT-CAATAAG-CGGAGGAAAAGAAACTTACCAGGATTCCCTTAGTAACGGCGAGCGAAC 386
                      *** **

C.macroptera          CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
P.trifoliata          CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGC---CG----------------- 436
C.medica              CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.karna               CGGGAAAGAGCCCA-CTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.nobilis             CGGGAA--AGCCCAG-TTGAAAAT-GGGCGCCCCCGGC---------------------- 430
C.pseudolimon         CGGAAG--ACCCCAG-TTGAAAATCGGGCGCCCCCGGC---------------------- 436
A.marmelos            ------------------------------------------------------------
C.reticulata          CGGGAA--AGCCCAG-TTGAAAATCGGGCGCCCCCGGC---------------------- 433
C.rugulosa            CGGGAA--AGCCCAG-TTGAAAATCGGGCGCCCCCGGCG--------------------- 434
C.ichangensis         CGGGAA--AGCCCAG-TTGAAAATCGGGCGCCCCCGGC---------------------- 433
C.reshni              CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 458
C.limettioides        CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.limon               CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.aurantifolia        CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.megaloxycarpa       CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 459
C.aurantium           CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 455
C.latipes             CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.jambhiri            CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 455
C.sinensis            CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.volkameriana        CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 455
C.paradisi            CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 455
C.limonia             CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.indica              CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 454
C.grandis             CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 456
C.assamensis          CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCCCGGCGTCCGAATTGTAGTCTGGAGAA 457
M.paniculata          CGGGAA-GAGCCCAGCTTGAAAATCGGGCGCCCTCGGCGTCCGAATTGTAGTCTGGCG-- 443

C.macroptera          GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
P.trifoliata          ------------------------------------------------------------
C.medica              GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
```

```
C.karna             -CGTCCTCAGCG---------------------------------------------- 467
C.nobilis           ------------------------------------------------------------
C.pseudolimon       ------------------------------------------------------------
A.marmelos          ------------------------------------------------------------
C.reticulata        ------------------------------------------------------------
C.rugulosa          ------------------------------------------------------------
C.ichangensis       ------------------------------------------------------------
C.reshni            GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 518
C.limettioides      GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.limon             GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.aurantifolia      GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.megaloxycarpa     GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 519
C.aurantium         GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 515
C.latipes           GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.jambhiri          GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 515
C.sinensis          GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.volkameriana      GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 515
C.paradisi          GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGC---------------- 499
C.limonia           GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGATAGGGTGAGA 516
C.indica            GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCAGAGAGGGTGAGA 514
C.grandis           GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 516
C.assamensis        GCGTCCTCAGCGGCGGACCGGGCCCAAGTCCCCTGGAAAGGGGCGCCGGAGAGGGTGAGA 517
M.paniculata        ------------------------------------------------------------

C.macroptera        GCCCCG---------------------------------------------- 522
P.trifoliata        ------------------------------------------------
C.medica            GCCCCG---------------------------------------------- 522
C.karna             ------------------------------------------------
C.nobilis           ------------------------------------------------
C.pseudolimon       ------------------------------------------------
A.marmelos          ------------------------------------------------
C.reticulata        ------------------------------------------------
C.rugulosa          ------------------------------------------------
C.ichangensis       ------------------------------------------------
C.reshni            GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 568
C.limettioides      GCCCCGTCGCGCCCGGACCCTGTCGCACCACG------------------ 548
C.limon             GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 566
C.aurantifolia      GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 566
C.megaloxycarpa     GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 569
C.aurantium         GCCCCG---------------------------------------------- 521
C.latipes           GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 566
C.jambhiri          GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 565
C.sinensis          GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 566
C.volkameriana      GCCCCGTCGCGCCCGGACCCTGTCGCACCACG------------------ 547
C.paradisi          ------------------------------------------------
C.limonia           GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 566
C.indica            GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 564
C.grandis           GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAATC 566
C.assamensis        GCCCCGTCGCGCCCGGACCCTGTCGCACCACGAGGCGCTGTCTGCGAGTC 567
M.paniculata        ------------------------------------------------
```

Appendix 5. Brief description of the methodologies for the reconstruction of *Citrus* phylogeny

Maximum Parsimony

The maximum parsimony method is one of the most widely used sequence-based phylogeny reconstruction method. This method finds phylogenetic trees from a number of aligned sequences through minimum number of evolutionary changes. Each nucleotide characters considered as distinct characters and the topologies obtained through the smallest number of substitutions from the observed alignment. The minimum number of character changes at a site is often called the character length or site length. The sum of character lengths over all sites in the sequence is the minimum number of required changes for the entire sequence and is called the tree length, tree score, or parsimony score. The tree with the smallest tree score is the estimate of the true tree, called the maximum parsimony tree. The parsimony method was first introduced by Edwards and Cavalli-Sforza (1964) for gene frequency data and then first applied to molecular sequence data by Eck and Dayhoff (1966). This method considers that the transformation of one character state to another implies a transformation through any intervening state, as defined by the ordering relationship (Farris 1970). Which permit the reversibility of the tree, that is, transformation in character states can be in either direction between nodes. Different parsimony methods were defined and commonly used in the phylogeny reconstructions are the Fitch and Wagner parsimony and Dollo parsimony. The trees generated by these methods are unrooted and the different rootings do not cause changes in the branch lengths, as represented by the number of steps. The Fitch and Wagner parsimony criteria are based on the assumption that the probabilities of character changes are symmetrical (i.e., the probabilities of transformations from character 0 to 1 and 1 to 0 are the same). The Fitch (1971) and Hartigan (1973) algorithm are commonly used in tree reconstruction that calculates the minimum number of changes and that is implemented in PAUP program. This yields large number of trees with common tree scores and among these trees the strict consensus of all trees are considered to be the best tree. Phylogeny reconstruction through this method is free from any evolutionary processes or assumptions (Felsenstein 1978). Therefore, when substitution rates variation is less then such topologies are considered to be good estimate phylogeny. However, when substitution rates are high, sequence evolution lineage divergence is much greater than the actual divergence between

lineage splits (a tree with very long terminal branches and short internodes) known as long branch attraction (Huelsenbeck 1995). The long branches become artificially connected due to nonhomologous similarities increasing the number of number of homologous similarities in the groupings of true closest relatives (Swofford et al. 1996). Goodness of fit of the characters in the data matrix can be validated by the consistency index (Kluge and Farris 1969), retention index (Farris 1989) and rescaled consistency index (Farris 1989b). Parsimony methods do not provide any statistical support; hence the bootstrap is employed to place confidence intervals on parsimony-inferred phylogenies.

Maximum likelihood

Maximum likelihood is one of the widely used model based method for phylogeny reconstruction. In maximum likelihood approach phylogenetic inference are based on the net likelihood values through evolutionary models on the observed sequences and that yield trees with the highest likelihood scores (Felsenstein 1981). This provides the tree topology, branch length and parameters of the evolutionary model through maximizing the probability of the observed data. The likelihood is the sum of the probabilities of observing data of each possible reconstruction under a particular substitution model through Markov process. The probability of the observation that is the tree and parameters are the functions of the observed event independent of the evolutionary model. The tree with the highest log-likelihood score is the phylogeny hypothesis best supported by the observed data and finally tree branches are supported by re-sampling method, i.e., bootstrap analysis (Felsenstein 1985). This method considers that the evolutionary history with greatest probability of the observed parameter is most likely to be correct (Swofford et al. 2006). Several evolutionary model options are available in maximum likelihood analysis and that varies in assumptions on processes of nucleotide substitution. The program ModelTest (Posada and Crandal 1998) uses log likelihood scores to establish the model that best fits the data.

Bayesian Phylogenetic Inference

Bayesian a powerful method of phylogeny reconstruction through posterior probability estimate for a hypothesis using models of evolution. The Bayesian method is based on the Bayesian theorem which provide the degree to which one believes that a proposition is true depends on the a priori belief which one has in the truth of the proposition and in the

evidence collected to investigate the proposition. Bayesian posterior probability for a tree or clade is the probability that the tree or clade is true to the given data, the likelihood model and the prior. In Bayesian analysis the value of the parameter is unknown, hence probability distribution value must be specified and the distribution of the parameter before the data are analysed is called the prior distribution. This can be specified by using either an objective assessment of prior evidence or the researcher subjective views of the parameter. The objective principle take the prior to be a representation of prior objective information about the parameter and the subjective view accepts the prior to represent the researcher's subjective belief about the parameter before analysing the data. The Bayes theorem is then used to calculate the posterior distribution of the parameter, that is, the conditional distribution of the parameter given the data and inferences about the parameter are based on the posterior probabilities (Huelsenbeck et al. 2001). Posterior probability is the summation and integration over all possible combinations of tree, branch length and substitution model parameters and the Markov chain Monte Carlo (MCMC) algorithm (Metropolis et al. 1953; Hastings 1970) is used for approximating probability distributions. A variant of MCMC called Metropolis-coupled MCMC (MCMCMC) implemented in the phylogenetic analysis to approximate the posterior distribution of tree probabilities (Huelsenbeck et al. 2001). MCMC works in three different steps: first using a stochastic mechanism a new state for the Markov chain is proposed. Secondly, the probability of this new state to be correct is calculated. Thirdly, a new random variable (0, 1) is proposed. If this new values are less than the acceptance probability the new state is accepted and the state of the chain is updated. This process is repeated for either thousands or millions of times to get highest probability support values. The amount of time a single tree is visited during the course of the chain is just a valid approximation of its posterior probability (Huelsenbeck and Ronquist 2001).

Bootstrap Analysis

The bootstrap was introduced by Efron (1979) and is applied in phylogeny reconstruction as a method for obtaining confidence limits on phylogenies (Felsenstein 1985). This method is also known as 'resampling method' as it involves the generation of new data sets by random resampling of positions in the original data set. Generally, tree topologies obtained from different phylogeny reconstruction methods represent taxon relatedness in a series of nested

taxon bipartitions. Branch lengths of individual taxon bipartitions indicated the number of inferred synapomorphies supporting those relationships without confidence in the branches. Boostrapping approximate the underlying distribution of empirical data matrix that from a finite sample by random resampling with the replacement from the empirical data (Felsenstein 1985). Sufficient pseudoreplicate data matrices that were constructed through resampling undergo heuristic analysis and the optimal trees derived from heuristic searches on each bootstrap pseudoreplicate were compared across pseudoreplicates and each taxon bipartition was assigned a percentage indicating the proportion of instances it was recovered. The resulting percentages do not represent strict confidence statements about the accuracy of the taxon bipartition, but indicate the relative degree of internal consistency in the data suggesting that bipartition. The bootstrap values of 95% or greater be considered statistically significant for support for a clade and values less than 50% considered as insufficient statistical support (Felsenstein 1985). This is a is a neutral statistical process that only reflects the phylogenetic signal (or noise) without any evolutionary relationships, therefore, confidence intervals in the biased / incorrect estimate of phylogeny reconstruction are not meaningful.

Appendix 6. *Citrus* species diversity in northeast India

C. aurantifolia

C. aurantium

C. grandis

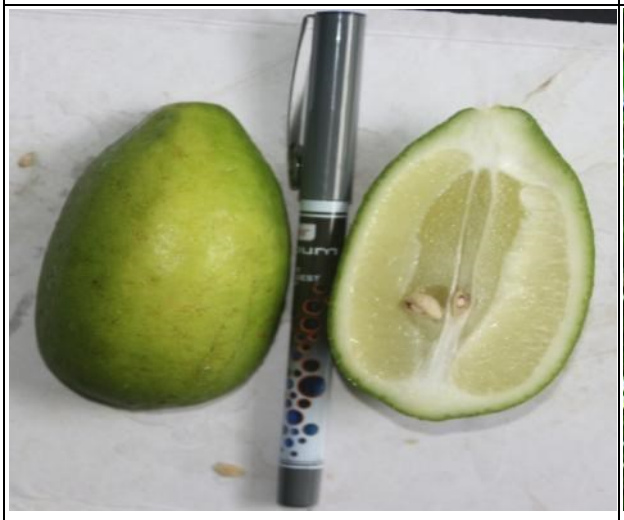
C. grandis

C. indica

C. jambhiri

C. karna

C. limonia



C. limon

C. latipes

C. macroptera

C. megaloxycarpa

C. medica

C. paradisi

C. sinensis

C. volkameriana

P. trifoliata

P. trifoliata

Appendix 7. Morphological diversity in *Citus medica* populations in northeast India

	
Population # 1	Population # 2
	
Population # 3	Population # 4
	
Population # 5	Population # 6

Population # 7

Population # 8

Population # 9

Population # 10

Population # 11

Population # 12

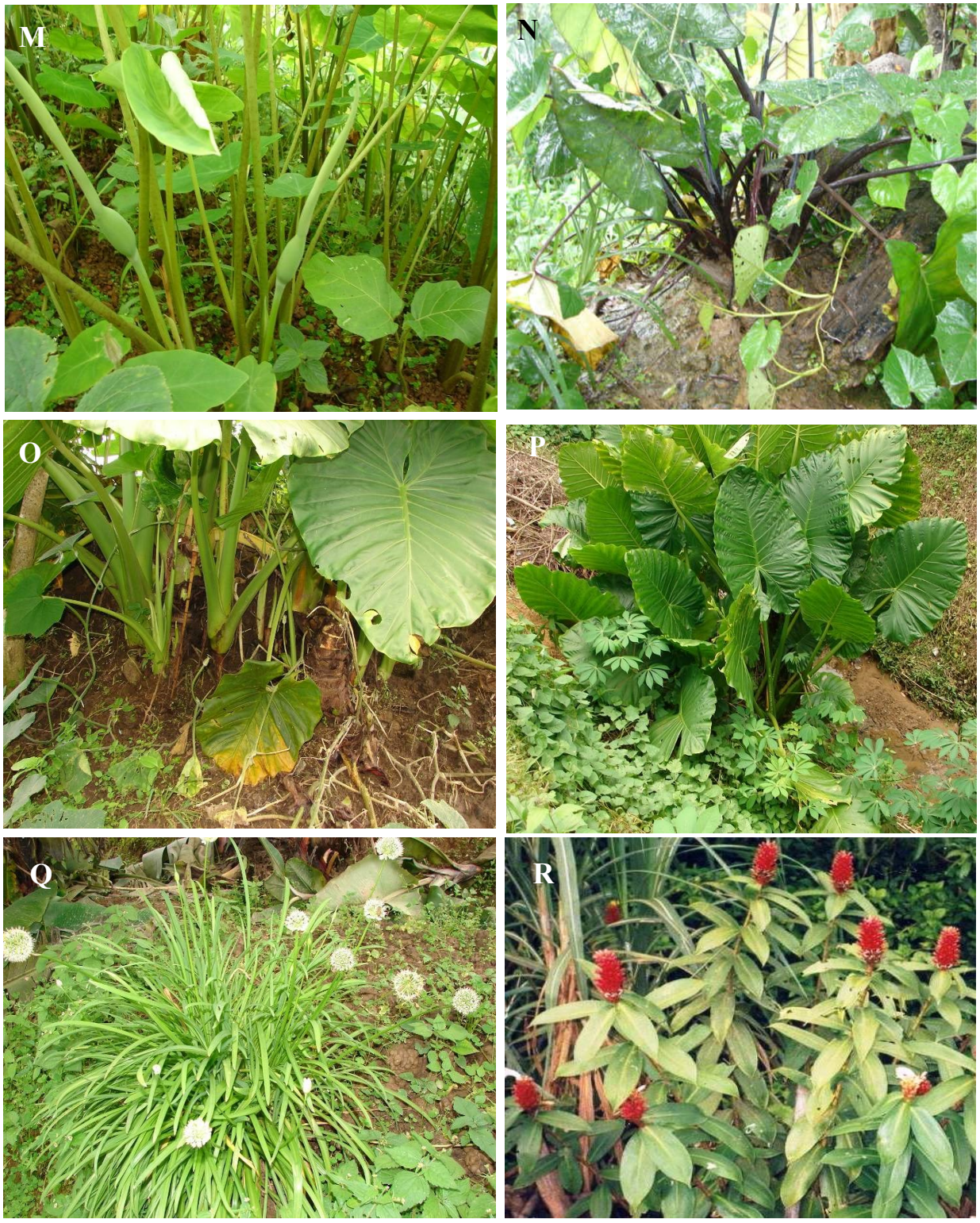
Appendix 8. Typical home gardens in (a) Sairang (b) Selesih and (c) Thingsulthliah in Aizawl district, Mizoram

Appendix 9. Typical home gardens in (a) Serchhip (b) Keitum and (c) Chhiahtlang in Serchhip district, Mizoram

Appendix 10. A few wild crop relatives and domestic plants / varieties commonly grown in the home gardens in Mizoram.

(**A-***Ficus recemosa,* **B**- *Ficus cunia,* **C**- *Ficus recemosa,* ***D***- *Mangifera sylvatica,* **E-** *Artocarpus heterophyllus,* **F**- *Oroxylum indicum).*

(**G-** *Trevesia palmata,* **H-** *Rauvolfia serpentina*, **I-** *Solanum violaceum,* **J-** *Solanum khasiana,* **K-** Wild Chenopodium, **L-** *Calamus guruba*)



(**M-** *Colocasia esculenta,* **N-** *Colocasia lihengiae,* **O-** *Colocasia macrorrhiza,* **P-** *Colocasia gigantea,* **Q-** *Allium hookerii,* **R-** *Costus speciosus).*