Evaluating DNA Extraction and Analysis of Ancient Seeds: A Study in Peach

A thesis presented to

The Faculty of Graduate Studies

of

Lakehead University

by

Felicia Joseph

In partial fulfillment of requirements
for the degree of
Master of Science in Biology

July 2015

Declaration:

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgment has been made in the text.

Felicia Joseph

Abstract

The peach (*Prunus persica)* is one of the most widely grown stone fruit in the world but its origin and domestication is still being debated. Genetic research has focused on identifying desirable genes for agricultural purposes, suitable genetic targets for species identification, the development of a genetic linkage map for the peach, and ultimately the publication of the whole genome by the International Peach Genome Initiative. However, there has been no genetic research applied to ancient peach stones. In the present study DNA was extracted from modern peach samples collected from the Zhejiang Province in China to assess the most effective methods to extract DNA from seeds. These methods were then applied to an archaeological peach stone collected from the Maoshan archaeological site in China dated to the Liangzhu Culture Period. A search for previously sequenced genetic targets from *P. persica* and its wild species *P. ferganensis*, *P. mira*, *P. kansuensis*, and *P. davidiana* was performed to search for suitable genes for genetic and phylogenetic analysis. Phylogenetic analysis was used to evaluate the genetic targets for species identification. A Type II chlorophyll *a/b*-binding protein gene (*Lhcb2*) was identified as an ideal marker for species identification for ancient peach specimens due to its small amplicon size and ability to amplify DNA from low quantities. Chloroplast DNA and microsatellite markers were assessed for their ability to distinguish between peach and its various wild species. Several chloroplast genes successfully amplified DNA in the modern peach samples (*rbcL*, *psbM-trnD*, *clpP*), but were not successful when applied to ancient samples. GC-MS analysis was performed on the ancient peach stones to determine whether miscoding lesions were present in the DNA sequence which may have led to amplification failure. Damage due to oxidation and hydrolysis are consistent with the location in which the stone was collected and suggest the possibility of exposure to water and burning. Both sources of damage may explain the lack of amplification of these archeological samples.

Lay Summary

"Faculty and students in the Department of Biology are bound together by a common interest in explaining the diversity of life, the fit between form and function, and the distribution and abundance of organisms." The peach is one of the most widely grown fruit in the world, and it is of great importance agriculturally, economically, and socially. Specifics' regarding where the domesticated peach originated and how it came to become part of everyday life is still being debated. An origin in China has been well documented but no DNA analysis has been performed on an archeological sample from this area. In this study we attempted to analyse DNA from two modern peach samples from the Zhejiang Province in China and one ancient peach sample collected from the Maoshan site in China. The purpose of this study was to assess the success of different DNA extraction and analysis methods and to attempt to provide an insight into the origins of the peach and its relationship to the fruit we see today.

Acknowledgements

Firstly, thank you to Dr. Carney Matheson for giving me the opportunity to continue my studies of ancient DNA at Lakehead University. I appreciate the guidance that you have given to me and encouraging me to never give up.

Thank you to my committee, Dr. Kam Leung and Dr. Lada Malek , and to my external reader for your support and evaluation throughout my studies.

Thank you to my friends and colleagues of CB3037 both past and present; especially Karen, Margaret, Christy, Heidi, and Jessie for all your support. The long days and nights were much more enjoyable with your presence. I cannot thank you enough for your words of encouragement and many laughs.

To the staff of the Paleo DNA Laboratory, Steven and Renee Fratpietro and Karen Maa, you were an incredible sounding board and never stopped pushing me to continue. The opportunity and experience you provided me with contributed in my success.

Mum, dad, and Sabrina; it has been a long road and I could not have made it through this without your love and support coming all the way from Hamilton and Ottawa. Sarah, Paulina, Paula and Erin, thank you for the long talks every visit home. To Nichola, you are simply the best, always the voice of reason and reminding me of my strength to keep going. Lastly, to all my Thunder Bay family, Sarah, Karen, Byron, and Steph, thank you for always pushing me to keep moving forward and being there for me through my endeavors.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Å	Angstrom
A	Adenine
aDNA	Ancient DNA
amu	Atomic mass units
AP	Apurinic or Apyrymidinic site
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Tool
BSTFA	Bis(trimethylsilyl)trifluoroacetimide
C	Cytosine
cpDNA	Chloroplast DNA
CTAB	Cetyl trimethylammonium bromide
ddNTPs	Dideoxynucleotide triphosphates
ddTTP	Dideoxythymidine triphosphate
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotides
EDTA	Ethylenediaminetetraacetic Acid
EtOH	Ethanol
ExoI	Exonuclease I
f	Forms
G	Guanine
GC-MS	Gas Chromatography-Mass Spectrometry
GDR	Genome Database for Rosaceae
GuSCN	Guanidinium Thiocyanate
LTR	Long Terminal Repeat
mtDNA	Mitochondrial DNA
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
nDNA	Nuclear DNA
PCR	Polymerase Chain Reaction
PK	Proteinase K
RAPD	Randomly Amplified Polymorphic DNA
RNA	Ribonucleic Acid
SAP	Shrimp Alkaline Phosphatase
SDS	Sodium dodecyl sulfate
SNP	Single Nucleotide Polymorphisms
SSR	Simple Sequence Repeats
T	Thymine
Taq	*Thermus aquaticus*
TMS	Trimethylsilyl
UPGMA	Unweighted Pair Group Method Average
UV	Ultraviolet
Var	Varieties

1.0 Introduction

1.1 History and Structure of DNA

Deoxyribonucleic acid (DNA) is the genetic make-up of all living things and is known as the 'blue print" of life. DNA is a large organic molecule, which encodes the hereditary information cells need to survive; the information contained is passed on from generation to generation. The early work of Franklin, Wilkins, Watson and Crick revealed the now recognisable helical structure of DNA using X-ray diffraction (Franklin and Gosling 1953a; Franklin and Gosling 1953b; Watson and Crick 1953a). The structure was identified as two helical chains wrapped around a central axis (Franklin and Gosling 1953a; Franklin and Gosling 1953b; Watson and Crick 1953a; Watson and Crick 1953b). The structure of DNA consists of complex units made up of a deoxyribose sugar and phosphate backbone. Attached to the deoxyribose sugar is one of the four nitrogenous bases: adenine (A), cytosine (C), guanine (G), or thymine (T) (Pauling and Corey 1953; Watson and Crick 1953a). The purines, guanine and adenine, pair with the pyrimidines, cytosine and thymine, respectively. The nitrogenous bases are attached to the $1^{'}$ carbon atom of the sugar through the $3^{'}$ atom nitrogen on the ring of the pyrimidines, and $9^{'}$ atom nitrogen on the ring of the purines (Pauling and Corey 1953). The two helices are held together by hydrogen bonding between the two complementary nitrogenous bases and the distance between sets of nitrogenous pairings is approximately 3.4 angstroms (Å) (Figure 1)(Watson and Crick 1953a). The phosphate-sugar backbone is held together through a phosphodiester bond joining β-D-deoxy-ribofuranose residues with $3^{'}$, $5^{'}$, linkages (Watson and Crick 1953a). The DNA molecule has a width of approximately 15-20 Å, is thousands of angstroms long, and is suspected to be relatively rigid (Watson and Crick 1953a). This rigidity is further supported by the discovery of the specific pairings of the nitrogenous bases and the strict orientations that must be achieved in order for the DNA molecule to remain stable and form the hydrogen bonds which make up the helical structure of the molecule (Watson and Crick 1953a). Understanding these restrictions and patterns, and the relatively flat nature of the bases allows them to

stack roughly on top of each other, which provides a means to determine the sequence on one strand

based on the other due to the complementary nature of opposite DNA strands (Watson and Crick 1953a).
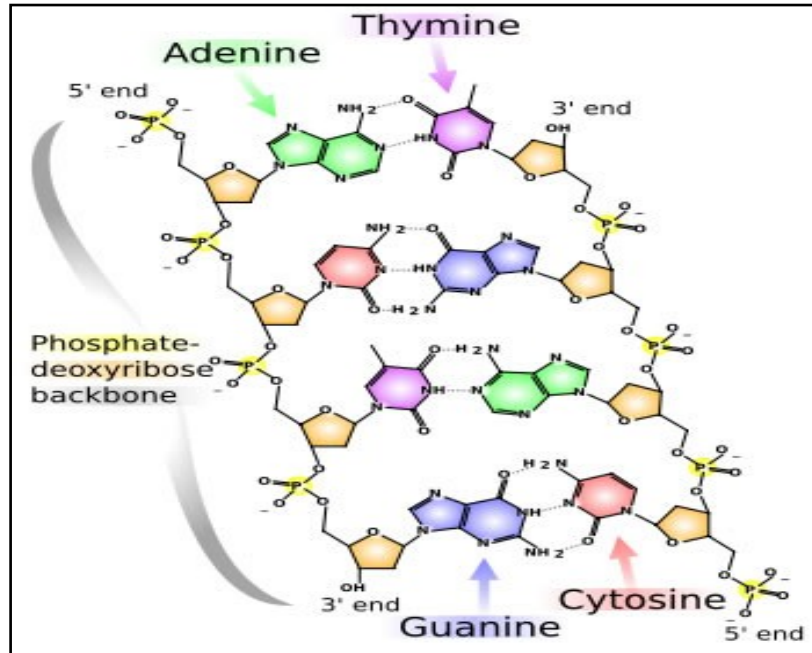
Figure 1. Basic structure of DNA molecule
http://www.sciencegeek.net/Biology/review/U5Storyboard.htm

The complementary nature was based on the work of Erwin Chargaff, who further investigated the

interactions between the nitrogenous bases believing that he could chemically demonstrate differences

between the DNA of different species (Chargaff 1950; Kresge et al. 2005). In 1950, he published that '*in*

a double-stranded DNA, the number of guanine units equals the number of cytosine units, and the number

of adenine units equals the number of thymine units' (Chargaff 1950; Kresge et al. 2005). He also stated

that '*the composition of DNA varies from one species to another*' (Chargaff 1950; Kresge et al. 2005).

There are three main different regions of the cell that DNA is found: the nucleus, the

mitochondria, and the chloroplasts in plants. Mitochondrial DNA (mtDNA) is located in the mitochondria

of the cell, where energy is produced (Budowle et al. 2003; Burger et al. 2003). In most cases, the

mtDNA is circular but in recent years, it has been found that it can exist in a linear form in some species

(Burger et al. 2003). The mtDNA is maternally inherited and exists in high copy numbers due to the multiple copies of mitochondria and their DNA in each cell (Budowle et al. 2003). In plants, there is less information available regarding mtDNA, as the focus has mainly been on the chloroplasts (Olmstead and Palmer 1994). In contrast, there are only two copies of nuclear DNA (nDNA) per cell; therefore mtDNA is very effective when conducting DNA analysis on ancient samples (Budowle et al. 2003). Like mtDNA, chloroplast DNA (cpDNA) is circular, exists in multiple copies in the cell, and is most often maternally inherited (Demesure et al. 1995; Palmer 1985). It consists of simple and stable genetic structures, is haploid, there is little to no recombination, and is generally transmitted uniparentally (Dong et al. 2012). The genome is conserved which makes it a good candidate for evolutionary studies, and therefore has been used to determine domestication of many agriculturally significant plants (Olmstead and Palmer 1994; Palmer 1985). Chloroplasts have been shown to have significantly less structural diversity than seen in the mitochondrial genomes and the rate of nucleotide substitution is low over the chloroplast genome as a whole (Burger et al. 2003; Palmer 1985). The nDNA is found in the nucleus of eukaryotic cells. Methods for analysing nDNA include sequencing, random amplified polymorphic DNA (RAPD) and the use of simple sequence repeats (SSRs) to reveal various characteristics of different plants, as well as to infer genetic relationships and phylogeny among between different species.

1.2 Ancient DNA

The study of ancient DNA (aDNA), or the analysis of DNA sequences from museum specimens, archaeological artifacts, buried remains, or other unusual finds often present a multitude of obstacles which must be overcome in order to perform analysis on various samples (Pääbo et al. 2004). The aDNA is often present in minute amounts and demonstrates some form of degradation. This is a result of various depositional conditions from which the sample has been recovered or in the way the sample has been stored. Within the field of aDNA it has become imperative to follow guidelines that have been devised to ensure the authenticity of results. Without these guidelines the risk of human DNA contamination is greatly increased, and can often present false positives (Pääbo et al. 2004). Various protocols have been

3

designed for the successful analysis of aDNA (Cooper and Poinar 2000; Gilbert et al. 2005; Handt et al. 1994; Hofreiter et al. 2001b; O'Rourke et al. 2000; Pääbo 1989; Pääbo et al. 2004). These protocols include working in a physically isolated work area; this is to avoid contamination from the beginning. By separating the DNA from areas where modern DNA is present, the risk of contamination is reduced (Cooper and Poinar 2000; Gilbert et al. 2005; Handt et al. 1994; Hofreiter et al. 2001b; O'Rourke et al. 2000; Pääbo 1989; Pääbo et al. 2004). The use of control amplifications during the extraction and amplification process allows for a detection of contamination and results should be replicated when possible to ensure the authenticity of results. Results can also be authenticated through independent replication. Sending samples to an independent laboratory can eliminate the suspicion of intra-laboratory contamination (Cooper and Poinar 2000; Gilbert et al. 2005; Handt et al. 1994; Hofreiter et al. 2001b; O'Rourke et al. 2000; Pääbo 1989; Pääbo et al. 2004). This is especially important when dealing with human remains as the risk of human contamination is much higher. Calculating the quantity of DNA allows for an understanding of the limitations of the polymerase chain reaction (PCR) and the risk of contamination. Low copy numbers can often have an effect on PCR and can increase the possibility of contamination (Cooper and Poinar 2000; Gilbert et al. 2005; Handt et al. 1994; Hofreiter et al. 2001b; O'Rourke et al. 2000; Pääbo 1989; Pääbo et al. 2004). Biochemical preservation can also play a very important role in the analysis of aDNA. The DNA molecule itself is prone to degradation and under extreme conditions this process can be expedited. These conditions may include the storage of the specimen, the extent of decomposition, and the environmental conditions in which the specimens are found (Gilbert et al. 2005; Lamers et al. 2009; O'Rourke et al. 2000; Pääbo 1989). All of these factors can result in significant damage to the DNA molecules which can inhibit PCR, therefore it is imperative to adhere to the criteria outlined above when dealing specifically with aDNA to ensure accurate results.

1.3 DNA Damage

Living cells are continuously being maintained by a variety of enzymatic repair processes; after death, these repair enzymes are no longer active and the DNA is rapidly degraded by lysosomal nucleases

(Pääbo et al. 2004). Under normal circumstances, the DNA repair mechanisms help to prevent single and double stranded breaks, oxidative and hydrolytic damage, and cross-linking (Lamers et al. 2009; Lindahl 1993; Pääbo et al. 2004). As a result of this degradation, a reduction of the DNA to approximately 100-500 base pairs occurs (Pääbo et al. 2004). Small fragmented DNA presents problems when conducting experiments such as isolation of bacterial clones carrying the same DNA sequence and blockage of PCR elongation (Pääbo et al. 2004). Areas prone to damage on the DNA molecule can be seen in Figure 2.

The degree and type of post-mortem damage is dependent on a variety of factors; the time of deposition, the taphonomy of the environment in which the remains were recovered and the conditions in which the samples were stored (Lamers et al. 2009). Taphonomy is the study of the processes that affect biological material after death. There are many biotic (bacteria, fungi, flora, and fauna) and abiotic (soil type, pH, weather, humic substances, temperature, oxygen content, etc.) factors that have major effects on the type of damage or preservation observed in recovered remains (Lamers et al. 2009). Although hydrolytic damage is often a main cause for DNA damage, DNA damage due to oxidative species should not be ignored, especially in regards to areas where hydrolytic activity is virtually absent (Dizdaroglu et al. 2002; Lamers et al. 2009). In many cases where endogenous DNA is preserved, there is no correlation that has been found between the age of a sample and the preservation of the DNA observed (Pääbo et al. 2004; Poinar and Stankiewicz 1999).

1.3.1 Hydrolytic Damage

Hydrolytic attack often occurs at the N-glycosyl bond, which links the deoxyribose sugar and the nitrogenous base; this results in an apurinic or apyrimidinic (AP) site (Lamers et al. 2009; Lindahl 1993). The most common break occurs between the N-glycolsyl bond and a purine base, with a slightly higher incidence of guanine being released (Lindahl 1993). Apurinic sites can induce strand breaks due to their weakened nature which in turn can block the PCR by blocking the polymerase activity (Friedberg et al. 2006). Hydrolytic damage can also result in deamination, where damage occurs within the nitrogenous bases at the glycosyl bonds (Hoss et al. 1996; Lamers et al. 2009; Lindahl 1993). The pyrimidines,

5

cytosine, 5-methyl cytosine and thymine are primary targets, whereas deamination of the purines, guanine and adenine occur less often (Lamers et al. 2009). The deamination of cytosine results in a conversion to uracil in DNA, and the deamination of its homolog 5-methyl cytosine results in a conversion to thymine (Lindahl 1993). Deamination of adenine results in a conversion to hypoxanthine which preferentially base pairs with cytosine (Lindahl 1993). Guanine can also undergo deamination and forms xanthine, which preferentially base pairs with cytosine (Lindahl 1993). As a result of deamination, mis-pairings between bases occurs.

1.3.2 Oxidative Damage

In addition to hydrolytic damage, oxidative damage can also cause extensive damage to both living and postmortem DNA molecules. The major causes of oxidative damage are reactive oxygen species; hydroxyl radicals, peroxide radicals, and hydrogen radicals (Lamers et al. 2009; Pääbo et al. 2004). The most common oxidative damage observed is the oxidation of guanine resulting in the formation of 8-hydroxyguanine, which preferentially base pairs with adenine (Dizdaroglu et al. 2002). It has been found that oxidative damage resulting in 8-hydroxyguanine occurs at a much higher frequency in mtDNA than in nDNA, approximately 1 per 8,000 bases (Lindahl 1993; Richter et al. 1988). In addition, DNA damage by reactive oxygen species may result in chemical changes to pyrimidines and purines, single and double stranded breaks, abasic sites, DNA-protein cross-links and modified sugars (Jenner et al. 1998). The double bonds of the nitrogenous bases are also susceptible to oxidative damage, leading to ring fragmentation (Dizdaroglu et al. 2002; Lindahl 1993; Pääbo et al. 2004). Ring fragmentation can also occur within the deoxyribose sugar molecule; damage to the sugar ring or base rings leads to blockage of DNA polymerase (Dizdaroglu et al. 2002; Lindahl 1993; Pääbo et al. 2004). High proportions of thymine and cytosine are oxidised to form hydantoins, which block DNA polymerase and thus PCR amplification (Hofreiter et al. 2001b).
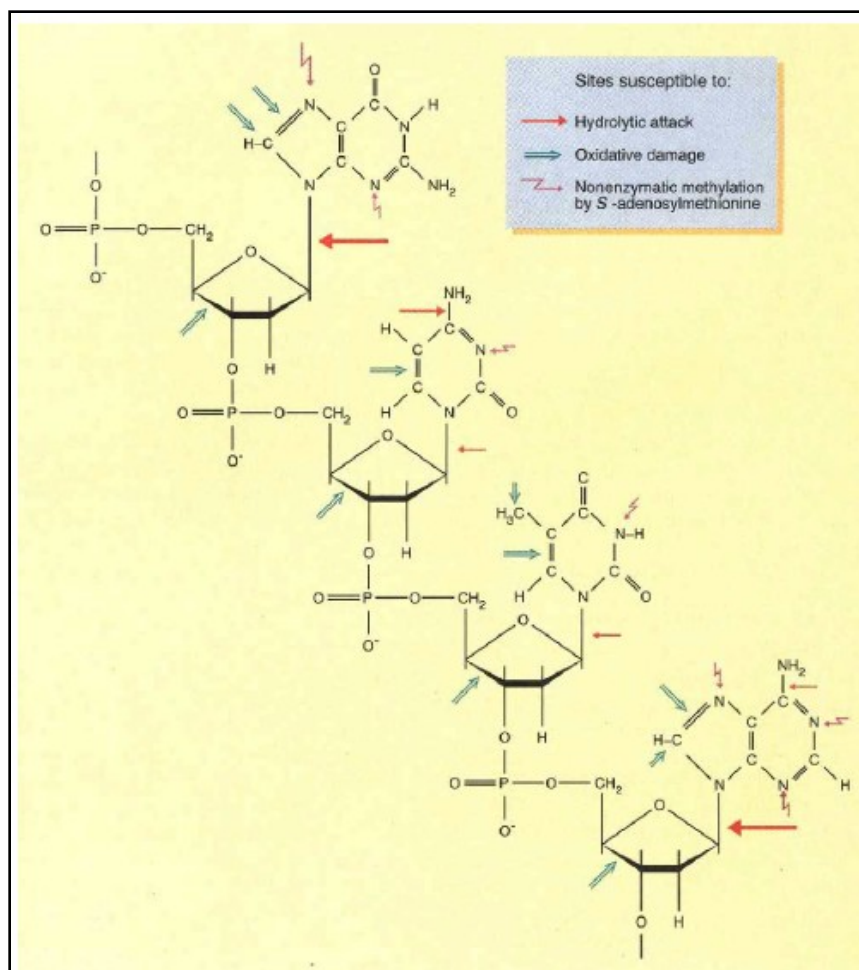
Figure 2. Hydrolytic and Oxidative Attack on a DNA molecule (Lindahl 1993)

1.4 DNA Preservation

Due to the fact that the DNA molecule contains all genetic information, it is a very important molecule to study. As a result of the advances made in the field of aDNA with regards to PCR and more recently DNA sequencing, the ability to extract DNA from extremely old samples has increased. The complete genome of a 700,000 year old horse extracted from bone fragments was accomplished in 2013, revealing a genome almost ten times older than the previous record (Millar and Lambert 2013; Orlando et al. 2013). Even with these advances, extracting DNA from ancient material can be difficult and many factors should be considered. Hofreiter et al. (2001a; 2001b) discussed DNA decay and our current understanding of the limits to DNA preservation. He described how rapid desiccation and low

7

temperatures aid in preservation and how the actions of nucleases destroy DNA, and in addition, how still slower, and relentless processes over time will continue to affect the DNA (Hofreiter et al. 2001a; Hofreiter et al. 2001b). These processes include oxidation and hydrolytic damage to the DNA. Due to the natural degradation of DNA and the possibility of being unable to recover any viable DNA, a variety of methods have been used in order to determine its preservation. Other organic macromolecules are present in the same species as the DNA and are exposed to the same environmental conditions and stress (Poinar 1999). Hofreiter et al.(2001a; 2001b) stressed the need for a rapid-screening method to determine whether or not the samples are so poorly preserved that it would be no longer possible to amplify the DNA.

1.5 Peach Background and Domestication

Prunus persica L. Batsch, commonly known as the peach (Figure 3), is an ancient fruit that has become widely grown throughout the world. Although its origins in China have been accepted, the putative country of origin was rumoured to be Persia until the 19th century, resulting in the naming of *P. persica* (Hedrick et al. 1917; Layne and Bassi 2008). This was challenged early when it was realised that if the peach had originated in Persia, its presence in Asia minor and Greece would have occurred earlier (Hedrick et al. 1917). Botanists could often begin determining the origin of a species by assessing where the plant would grow spontaneously, and it was found that the peach had been growing wild in China for a very long time (Hedrick et al. 1917). The lack of mention of the peach in the Bible or other Hebrew books also contributed reason to doubt the Persian origin (Hedrick et al. 1917).

Figure 3. *Prunus persica* **(The peach)**

Evidence of peach is present in the Chinese literature dating back as far as 1,000 years before any evidence of its presence in Europe (Hedrick et al. 1917; Layne and Bassi 2008). The peach was a symbol of immortality and various stories are present in the Chinese literature (Layne and Bassi 2008). Early evidence of the use of peach in China has been found at various sites. In a Neolithic site in Hemudu village, Yujao city, in the Zhejiang province in 1973, peach stones were discovered dating back approximately 6,000-7,000 BC, and peach stones from the Shang Dynasty (1600-1100 BC) were found at a site in the Taixi village of Goachen city in the Hebei province (Layne and Bassi 2008).

As early as 4,000 years ago, the peach has been used and cultivated extensively throughout China and is one of the most ancient domesticated fruits (Layne and Bassi 2008). The earliest Chinese texts in reference to the peach were relics of the Pre-Qin Dynasty (1100-221 BC), and were described as '*Peach growing in the garden, its fruit for eating*' (Layne and Bassi 2008). There were also descriptions revealing early cultivation, ideal soil, and its uses as an edible or landscaping plant (Layne and Bassi 2008). Throughout the Eastern and Western Han Dynasty (222 BC- AD 220), cultivar selection and technique continued to advance, leading to a new era in the domestication of the peach (Layne and Bassi 2008). The time between the Wei-Jin Dynasty and the Sui-Tang Dynasty and the Five Dynasties period (AD 221-960) established the traditional peach culture using seedling selection and domestication of wild trees to create new cultivars (Layne and Bassi 2008). Knowledge regarding the biology of peach trees also contributed to this selection process and agriculture practices (Layne and Bassi 2008). The Song,

Yuan, Mong, Qing Dynasties and Republican period AD 961-1948) showed the expansion of peach both domestically and internationally while advances in cultivation and technology continued to expand throughout the country and moved towards germplasm selection (Layne and Bassi 2008). Selection of the non-tip and honey peach most likely resulted as the progenitor varieties from which all Southern Chinese peaches were derived (Layne and Bassi 2008). Assessment of the cultivation and domestication of the peach in China was investigated using peach stone morphology of archaeological stones recovered from various sites in the Zhejiang Province (Zheng et al. 2014). Over 24 archaeological sites, mostly from the Yangzi valley have revealed peach stone remains dating as early as the Three Kingdoms Period (AD 57-668) (Zheng et al. 2014). Examining peach domestication of various peaches, the Lower Yangzi River valley was proposed as the most likely region of early peach selection and domestication as early as 7,500 years ago (Zheng et al. 2014). It is assumed that prior to this time, the peach was already being collected and thus selection for preferred traits was already occurring (Zheng et al. 2014). As early as 3,000 years ago the peach moved in the Asian continent and into Persia via the Silk Road (Byrne et al. 2012).

Exploration resulted in the movement of the peach into Europe and North America (Figure 4). One of the earliest accounts of *P. persica* in Europe is in 332 BC in Greece by Theophrastus where there was a brief mention of a 'Persian fruit' (Hedrick et al. 1917). The peach further moved into Italy where it was mentioned by Vergil in 71-19 BC , this was the first mention of the fruit in this area (Hedrick et al. 1917). A few stories of the peach, written by Columella, describing it as a poisonous gift sent of Persia to Egypt, have also been found in European history books (Hedrick et al. 1917).

As a result of the Spaniards conquest, the introduction of the peach into the Americas began. The peach cultivation began in Mexico, south of the Rio Grande, but this may not have been the first time the peach had been introduced (Hedrick et al. 1917). Missionaries who came over from Europe, also would plant gardens in the places they completed their work, and one of the fruits they introduced was the peach (Hedrick et al. 1917). Once distributed in Mexico, the peach was moved into New Mexico, Arizona, Florida and California, where ideal conditions for growth were available (Hedrick et al. 1917).
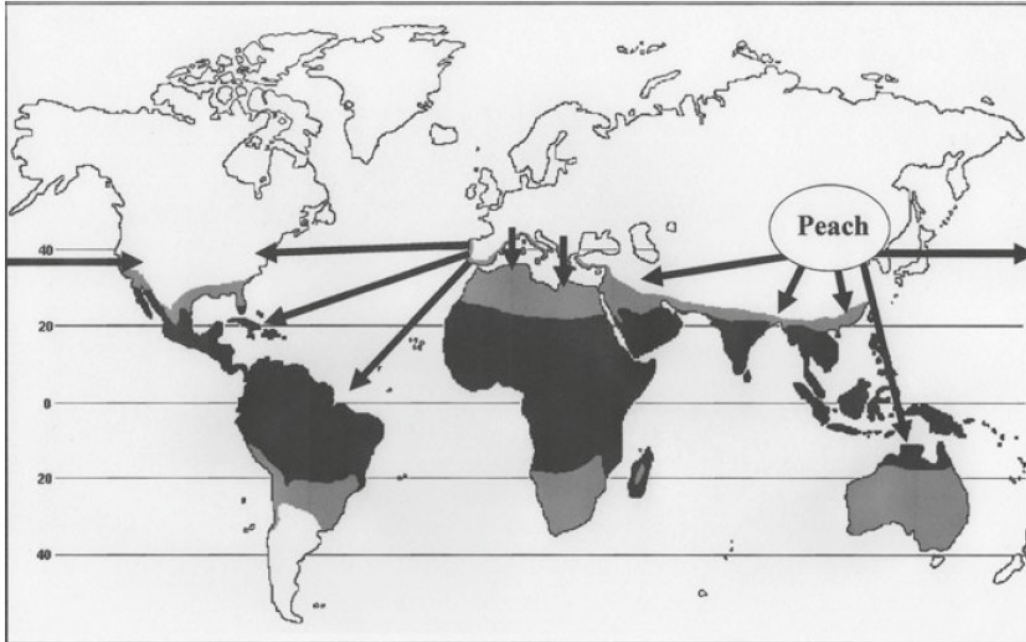
Figure 4. Early dispersal of *Prunus persica* from China approximately 3,000 years ago throughout Asia, Japan, and Persia via the Silk Road. The peach spread throughout Europe and eventually the Americas as a result of exploration (Byrne et al. 2012).

Today, the peach is the third most produced temperate fruit species in the world and is an important model fruit in the Rosaceae family (Byrne et al. 2012). The peach is able to adapt to a broad range of climates and is often a favoured study subject due to a variety of distinct features. Peaches are diploid consisting of sixteen chromosomes, and a relatively small genome of ~220-230 Mbp (International Peach Genome et al. 2013). It has a relatively short juvenile period of 2-3 years in contrast to other fruit species which can have juvenile periods up to 6-10 years (International Peach Genome et al. 2013). Peach has been extensively studied over the past few years, giving insight to a variety of genes that have led to advances in mapping the genome. Modern peach cultivars appear to have arisen as a result of only a small number of cultivars brought from China, the 'Chinese Cling' and the 'Shanghi Shumi' (Byrne et al. 2012; International Peach Genome et al. 2013). From these cultivars, genetic manipulation occurred as they attempted to manipulate different phenotypes and genotypes that were more desirable for different climates (Byrne et al. 2012).

11

1.6 Genetic Studies of Peach

In general, peach cultivars have a narrow genetic base due to the limited number of genotypes used as parents in breeding (International Peach Genome et al. 2013; Yoon et al. 2006). As a result, commercial cultivars have a more restricted level of adaptability than the native cultivars found in China (Yoon et al. 2006). The diversity of *P. persica* has been strongly reduced due to subsequent genetic bottlenecks that have occurred during domestication and movement into the New World (International Peach Genome et al. 2013). The first major bottleneck occurred in China over 4,000 years ago when the peach was first used and this was followed by vegetative propagation, the second occurred as a result of the dispersal through Persia into Europe, and the final bottleneck occurred throughout the 16th-20th century when peach was brought into the Americas (International Peach Genome et al. 2013). As breeding programs in the United States increased, nucleotide diversity decreased with the use of grafting, the process of connecting two different plants so they grow as one, in place of seed propagation occurred (International Peach Genome et al. 2013).

There are four wild species of peach, all native to China. *Prunus ferganesis* (Kostina & Rjabov) Y.Y.Yao, *Prunus mira* Koehne*, Prunus davidiana* (CarriŠre) Franch*.,* and *Punus kansuensis* Rehder. Both *P. davidiana* and *P. kansuensis* are native to Northern China and are used as seedling rootstock due to resistance to hostile conditions such as drought and disease (Layne and Bassi 2008). *P. mira* is found in far-West China and has a wide variety of fruit types, some very similar to *P. persica* (Layne and Bassi 2008)*. P. ferganensis* is found in Western China, more specifically, the Fergana Valley and is genetically almost indistinguishable from various *P. persica* accessions (International Peach Genome et al. 2013).

Early genetic studies of peach primarily focused on identifying alleles for morphological traits. These studies were used for agricultural purposes in hopes of identifying and isolating desired characteristics that would produce more presentable and desirable products. Genetic studies have been reported as early as 1975, resulting in peach being the best genetically-characterised species in the *Prunus*

genus (Chaparro et al. 1994; Hesse 1975; Monet 1989; Yamazaki et al. 1987). Although morphological traits had been identified, few single gene analysis or linkage analysis had been performed.

As of 1994, 40 morphological traits and four linkage groups had been identified (Bailey and French 1949; Monet 1989; Monet et al. 1996). In 1994, two linkage maps were published; the first looked at nine different F2 families, obtained by selfing F1 trees derived from various parental crosses to assess linkage relationships between 14 morphological and two isozyme loci (Chaparro et al. 1994). Linkage was found between weeping and white flower; double flower and pillar; and flesh colour and malate dehydrogenase (Chaparro et al. 1994). Linkage relationships were determined using pairwise combinations, and linkage chi-squared values of greater than 10.8 were calculated identifying four linkage groups from the morphological and isozyme traits used in the study (Chaparro et al. 1994). Using the F2 plants generated by selfing an F1 tree derived from a NC174RL x 'Pillar' cross, one of the first published genetic maps consisting of 15 linkage groups, 83 RAPD markers, one isozyme, and four morphological markers were developed for peach (Figure5) (Chaparro et al. 1994).
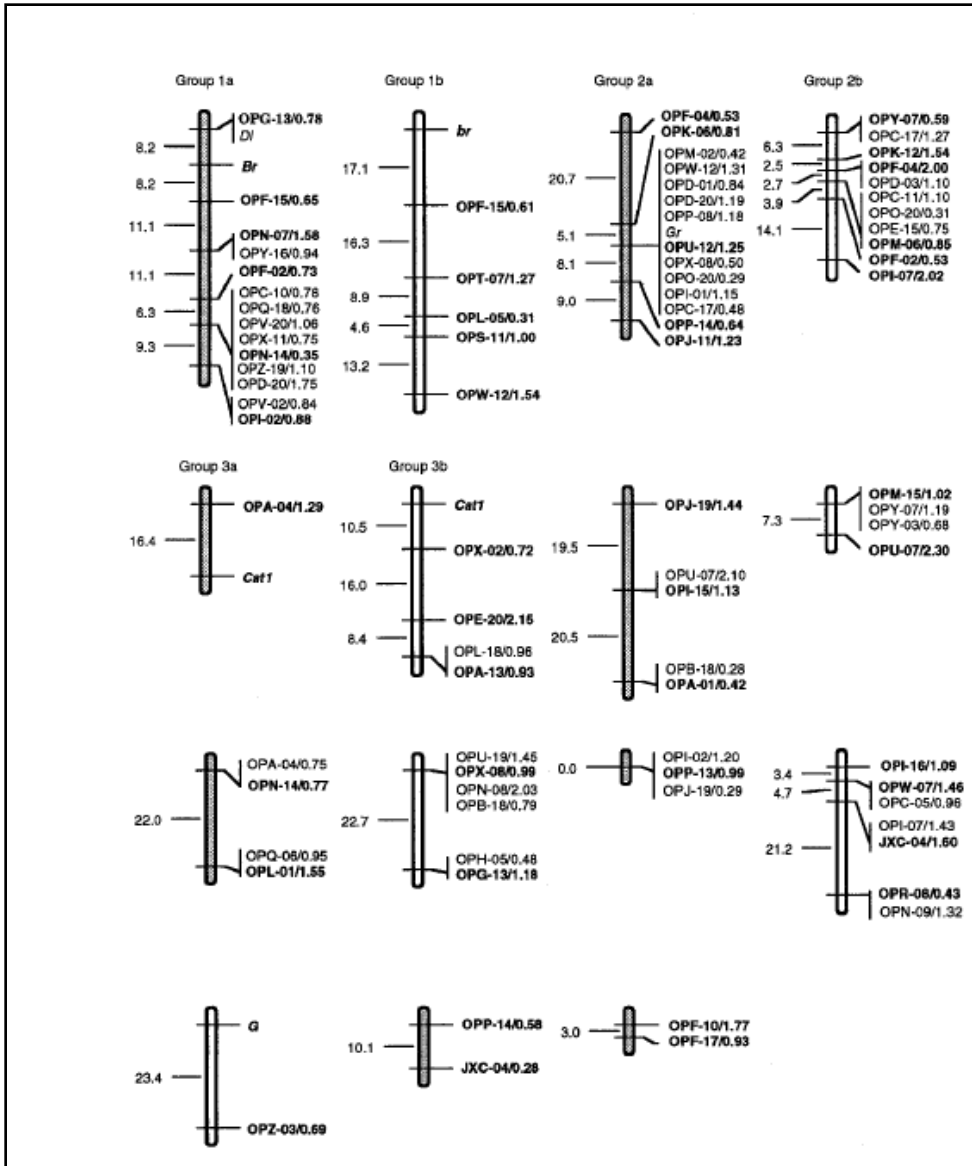
Figure 5. Genomic map of the 8 scaffolds of *Prunus persica* including morphological, isozyme, and RAPD markers in the NC174RL x 'Pillar' F2 family (Chaparro et al. 1994).

The second linkage map of the peach genome was published by Dirlewanger and Bodo (1994), who examined F2 populations to segregate several morphological characteristics and pest resistance. Like the previous study, RAPD techniques were used and eight linkage groups were established. Morphological traits were analysed such as nectarines, weeping shape and aphid resistance (Dirlewanger and Bodo 1994). Peach aphid resistance is controlled by a single dominant gene, while nectarine and weeping shape are controlled by a single recessive gene (Dirlewanger and Bodo 1994). However only the

14

weeping shape was extensively analysed, where 7% of the primers examined segregated in the F2 population and of those 38 primers, 11 gave polymorphic results (Dirlewanger and Bodo 1994).

This began a shift to the use of RAPD markers and bulk segregate analysis to identify specific markers flanking regions that were of interest. Early genetic studies of the peach genome revealed that DNA markers in peach had low polymorphic tendencies (Cipriani et al. 1999). Dirlewanger and Bodo (1994) reported that 5% of the RAPDs tested resulted in segregated bands while Chaparro et al. (1994) reported only 16%, Rajapaske et al. (1995) reported 18%, and Quarata et al. (1998) reported 38% of the RAPD primers tested showed polymorphisms. Due to these low polymorphic results, research turned to SSRs (microsatellites) approximately 1-6 nucleotide base pairs in length, in hopes of revealing more polymorphic motifs. These SSRs were desirable as they could be flanked by regions that were highly conserved in related species and could be used in cross species analysis (Cipriani et al. 1999). A study performed by Cipriani et al. (1999) identified that 88% of the 17 SSRs analysed showed polymorphic tendencies, showing 2-4 alleles (Cipriani et al. 1999). These SSR markers were also found to successfully amplify in other *Prunus* species and were integrated into the current peach linkage map (Cipriani et al. 1999).

In 2000, a genomic plasmid library (pUC8) and a cDNA library were used to help characterise SSRs that could be used in peach analysis. It was determined that CT repeats occur every 100 kb, CA repeats occur every 420 kb and AGG repeats occur every 700 kb in the peach genome (Sosinski et al. 2000). Five of the SSRs were found to segregate into intra-specific peach mapping crosses (Sosinski et al. 2000). Another benefit of using SSRs is that they are often co-dominant unlike RAPD markers that are often the dominant mode of inheritance; therefore SSRs are able to be utilised for genetic mapping (Sosinski et al. 2000). One of the difficulties that arose as a result of this direction of study was that although SSR markers were polymorphic in certain crosses, it was difficult to assign or map them to any previously studied linkage groups (Sosinski et al. 2000). Although some primer sets were unsuccessful, the marker *pchgms*1 was mapped to linkage group 1 in two of the crosses; this linkage group contained two important morphological characteristics, the double flower and pillar (Scorza et al. 2002; Sosinski et

al. 2000). The primers were also tested for their ability for cross species amplification. All of the primers produced an amplified product in apricot (*Prunus ameniaca* L.), thale cress (*Arabidopsis thaliana* (L.) Heynh.) and four sets amplified in rose (*Rosa* spp.) and sour cherry (*Prunus cerasus* L.) (Sosinski et al. 2000)**.**

Twenty six SSRs of AC and CT repeats were isolated in *P. persica* and all of the SSRs tested positive for polymorphisms of 2-8 alleles per locus (Testolin et al. 2000). In order to progress further with regards to the peach genome, the construction of a bacterial artificial chromosome (BAC) library to identify abundance and location of SSRs in peach was investigated in 2002 (Georgi et al. 2002). It was determined that SSRs were not randomly distributed within gene containing regions and were more abundant than previously believed (Georgi et al. 2002). As a result, the identification of the SSRs could be facilitated easily by hybridisation to BAC clones and sequencing of the clones (Georgi et al. 2002).

The genetic diversity and ecogeographic phylogenetic relationships among peach and nectarine cultivars were investigated using SSRs (Yoon et al. 2006). Thirty three SSRs resulted in polymorphic markers, and using unweighted pair group method average (UPGMA) cluster analysis, the genotypes were classified into six groups corresponding to their ecogeographical origins (Yoon et al. 2006). These groups included group 1, which consisted of Northern Chinese and Northwestern Chinese local cultivars, group 2 which consisted of Southern Chinese local, Japanese and North American cultivars and Group 3, 4 and 5 consisted of Chinese local ancient cultivars (Yoon et al. 2006). Previous studies of genetic relationships between peach and nectarine had been conducted but were restricted to cultivars from the US and Europe and not Asian cultivars (Dirlewanger and Bodo 1994; Sosinski et al. 2000; Testolin et al. 2000; Wang et al. 2002; Yoon et al. 2006).

As genetic studies of the peach continued and early genetic maps were being developed, studies turned to examining reference genes for gene expression and identifying unique genes that were indicative of species identification. Tong et al. (2009) set out to evaluate the ability of 11 reference genes for normalisation in peach genome expression. It was found that translation elongation factor 2 (TEF2), ubiquitin 10 (UBQ10) and RNA polymerase II (RPII) genes were suitable reference genes for peach

(Tong et al. 2009). Aranzana et al. (2010) used various SSRs that were evenly distributed along the eight

peach linkage groups to analyse the distribution, and structure of peach variation between North

American and European peaches. In addition, they also included old Spanish varieties and various

founders cultivars that were used in US peach breeding programs (Aranzana et al. 2010). The peaches

tested, separated into three clusters based on certain fruit characteristics: melting flesh peaches, melting

flesh nectarines, and non-melting peaches and nectarines (Aranzana et al. 2010). It was observed that the

founders of the US cultivars tended to cluster with the modern cultivars giving the indication that their

contribution to the modern genetics was represented (Aranzana et al. 2010).

Genetic studies of the peach continued to thrive over the years and research turned to the goal of

sequencing the whole peach genome. While advances in genomic research continued and more

information regarding the peach genome became public, an initiative was introduced to create a central

location for emerging data to be stored, deposited, and accessed. In 2003, the Genome Database for

Rosaceae (GDR) was created. Over the years, the website has become a tool for researchers to access

whole genomes of various Rosaceae species for downloads and browsing in addition to being able to

access a range of sequence alignments, primer design information, mapped genetic markers, published

polymorphisms, up to date publications, and various other tools for Rosaceae information (Jung et al.

2014; Jung et al. 2004; Jung et al. 2008).

As whole genome sequencing became the focus in the late 2000's, the International Peach

Genome Initiative sought to develop a high quality draft genome of peach, and to identify unique patterns

for genetic diversity, domestication and evolution. The draft genome was completed in 2010

(International Peach Genome et al. 2013). The cultivar 'Lovell' peach was used for the project because it

exhibited the typical phenotype of the domesticated peach and all of the alleles present were homozygous

(International Peach Genome et al. 2013). In 2010, a 224.6 Mb map of the peach genome, organised into

eight pseudomolecules was released containing 27,825 protein coding genes, as well as noncoding

ribonucleic acids (RNAs) were predicted (International Peach Genome et al. 2013). Analysis of repeat

sequences revealed that long terminal repeat (LTR) retrotransposons comprise 18.56% of the genome,

17

DNA transposons comprise 9.05%, and the total number of transposable elements sequenced comprised 29.60% of the entire genome (International Peach Genome et al. 2013). When compared to other plant species, these values were lower than expressed in apple and grape, but higher than that of *A. thaliana* (International Peach Genome et al. 2013). Further gene analysis revealed 27,852 protein coding genes and 28,689 protein coding transcripts, many of which have homologs with *A. thaliana* (International Peach Genome et al. 2013). The gene content predicted was much lower than observed in apple but similar to those observed in grape and *A. thaliana* (International Peach Genome et al. 2013). Gene density in peach was also investigated and compared to other plant species showing much higher diversity than seen in apple (1.22 genes per 10 kb vs 0.78 genes per 10 kb) but lower than that of *A. thaliana* (2.29) (International Peach Genome et al. 2013). The four wild species of peach were also aligned to 11 *P. persica* accessions including the 'Lovell' sequence. It was found that three of the four wild peach species, *P. kansuensis, P. mira* and *P. davidiana* were all very different from the reference sequence, detecting more than one million single nucleotide polymorphisms (SNPs) than observed in *P. persica* and *P. ferganensis* (International Peach Genome et al. 2013). In contrast, *P. ferganensis* appeared almost indistinguishable from all *P. persica* accessions tested (International Peach Genome et al. 2013). This was of great interest as *P. ferganensis* is native to the Fergana Valley which lies west to one of the proposed centers of origin of the peach in Northwestern China between Kunlun and Shan mountains and the Tarim basin (International Peach Genome et al. 2013). It is suggested that this wild species, possessing some undomesticated traits, may be an intermediate genome haplotype in peach (International Peach Genome et al. 2013). With the use of next generation sequencing and bioinformatics, high frequency SNPs were identified in the peach genome (Ahmad et al. 2011). Three peach cultivar genomes were sequenced and aligned to the 'Lovell' draft peach sequence, 'Dr. Davis', 'F8, 1-42', and 'Georgia Belle', which resulted in the selection of 6,654 high quality SNPs (Ahmad et al. 2011). As a result of this information being made available to the public and the continued advancements in whole genome sequencing, genomic studies in peach continued to increase.

With the use of previously published markers, the genetic diversity and phylogenetic relationships between different peach genotypes was investigated (Bakht et al. 2013). Analysis of the banding patterns revealed different levels of genetic polymorphisms between 20-80% (Bakht et al. 2013). The RAPD markers are important for breeding programs as tree growth can take up to 3-4 years, and it is important to select desirable genotypes for future breeding (Bakht et al. 2013).

Small variants such as SNPs and insertions/deletions have also been used in genetic studies to investigate genomic variants among different peach genotypes using sequence analysis (Fresnedo-Ramírez et al. 2013). By studying older cultivars such as 'Georgia Bell', an almond by peach introgression breeding line 'F, 1-42' and a model cultivar 'Dr. Davis', analysis of the genomic and phenotypic differences have helped lead to selections and protocols for future crops (Fresnedo-Ramírez et al. 2013). This is of importance as there is low genetic diversity, therefore comparisons between the various cultivars to the 'Lovell' genome have revealed small variations between different varieties (Fresnedo-Ramírez et al. 2013). Analysis of the variations between cultivars allowed for insight into variable regions across the genome, although base change rates differed between the three genotypes (Fresnedo-Ramírez et al. 2013). Between the three genotypes, it was noted that scaffold 2 exhibited the highest change rate, as a possible result due to a high number of recombination hot spots (Fresnedo-Ramírez et al. 2013). This chromosome is known for controlling traits such as ripening time, skin colour, diameter, and other important fruit traits (Fresnedo-Ramírez et al. 2013). Small variants in breeding are much more easily manipulated and their inheritance are better understood (Fresnedo-Ramírez et al. 2013). Scaffold 4 also exhibited a high change rate across the three genotypes, this was expected as scaffold 4 contains a trait loci for genes such as flesh texture (melting/non-melting) which were the main distinguishing factors between the three genotypes (Fresnedo-Ramírez et al. 2013). The variation was observed at the end of the scaffolds in which these genes were located (Fresnedo-Ramírez et al. 2013).

The cpDNA has also been investigated as a source of information for DNA barcoding of plant species due to the low evolutionary rates of chloroplast genes (Dong et al. 2012). These genes are often used for phylogenetic analysis but usually are unable to differentiate between closely related species; this

led to examining the noncoding regions which provide greater levels of variation (Cohen 2011; Dong et al. 2012; Shaw et al. 2007; Shaw and Small 2004). In the past, some regions of the chloroplast genome such as the *rbcL and rpoC1* gene have been used for DNA barcoding but alone do not provide sufficient variation to resolve plants at the species level (Dong et al. 2012; Quan and Zhou 2011). The DNA barcodes for peach were identified by Quan and Zhou (2011) using 11 DNA loci and tested using samples from 38 populations including all species of peach (Quan and Zhou 2011). The combinations of loci provided better distinction between different species of peach. Two combinations in particular, *atpB-rbcL* and *trnL-F* genetic regions and *atpB-rbcL* and *atpF-H* genetic regions, were able to distinguish between all five species studied (Quan and Zhou 2011). These five species included, *P. davidiana, P. kansuensis, P. mira, P. persica* and *P. potanini* (Quan and Zhou 2011). Using previously published complete chloroplast genomes; highly variable regions were identified across different genera, and used to determine their ability to resolve between closely related species. Forty seven highly variable loci ranged from between being present in only one of the genera being investigated to being present in five or more genus, 23 of these sites were further analysed and primers were designed to test their ability to distinguish between eight different species (Dong et al. 2012). A separate study was conducted applying nine of the chloroplast loci to the six species belonging to peach (Dong et al. 2012). The results revealed all six species were resolved using a combination of *psbM-trnD* intergenic spacer and *clpP* intron (Figure 6)(Dong et al. 2012).

20
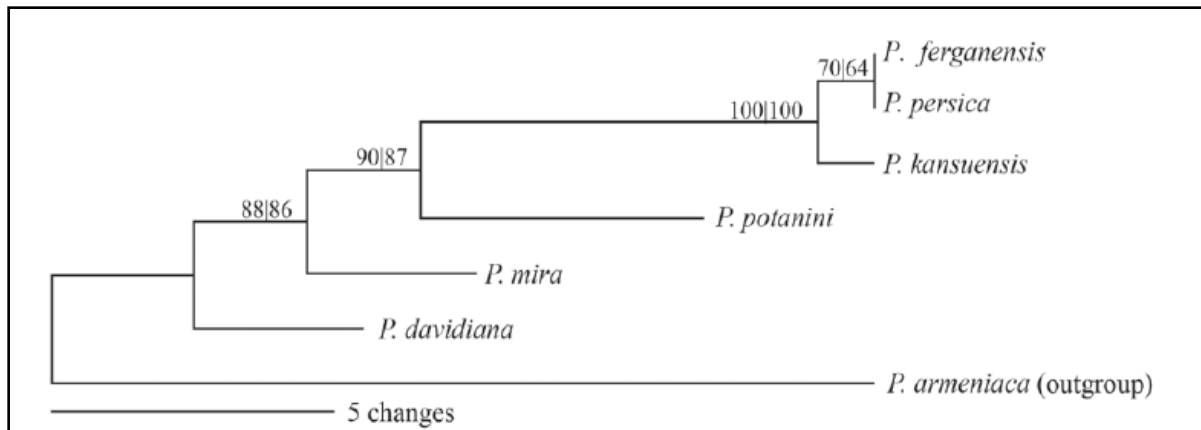
Figure 6. Maximum parsimony tree of 6 *Prunus* species were resolved using a combination of *psbM-trnD* intergenic spacer and *clpP* intron (Dong et al. 2012)

1.6.1 Characterisation of the Type II Chlorophyll *a/b*-binding Protein Gene (*Lhcb2*Pp1*) in Peach

In 1998, a Type II chlorophyll *a/b* binding protein was amplified in peach (Bassett et al. 1998). This gene is associated with photosystem II and with the light harvesting complexes of photosystems I and II specifically, expressed during the development of leaves in the absence of flowering (Bassett et al. 1998). As a result, it was determined that the highest expression of this gene was shown in leaves that were near or at full expansion at sampling time and therefore this gene could be used as a marker to indicate photosynthetic maturity under field conditions (Bassett et al. 1998). It was also determined that the *Lhcb* gene family were able to maintain their identity through concerted evolution (Bassett et al. 1998).

The use of DNA analysis in fruit has also extended beyond evolutionary exploration and gene discovery; it has also been used in detection of fruit products. Determination of an endogenous reference gene in peach for species specific identification is important in the food industry for the detection of food adulteration (Shang et al. 2014). A chlorophyll a/b binding protein (*Lhcb2*) was determined as a species specific gene for peach (Shang et al. 2014). Fruit juice adulteration occurs in the food industry and it is important to be able to identify the authenticity of foods. Food adulterations include the addition of water, sweeteners, sour agents, or fruit dreg extracts (Shang et al. 2014). Previous authentication methods have

21

included chemical and physical identification, sense organ appraisal, etc. which are used to identify main components of the juices (Shang et al. 2014). Molecular biology techniques have been applied to food identification more recently, particularly the use of species specific genetic identification to be used to effectively identify individual fruit components in pure and mixed juices and food. Endogenous genes are ideal as they have many desirable characteristics such as species specificity, they exist in low and constant copy numbers in various varieties, and have little to no variation (Shang et al. 2014). Among these benefits, the genes offer a low detection limit and can detect sequences with as little as 3 pg of DNA (Shang et al. 2014). This is very effective for aDNA analysis as there is often damage or very little intact DNA present in samples. In food, DNA can be damaged during the production phases (Shang et al. 2014). Various endogenous genes have been identified for different plants such as tomato, rice, cotton, papaya, soybeans, etc. (Shang et al. 2014). The use of the nuclear *Lhcb2* gene as a species specific indicator for peach was confirmed when the homology of the gene was compared to other closely related species of the genus *Prunus* (Shang et al. 2014). It was determined that the *Lhcb2* gene had no homology with other genes and no homology with genes belonging to closely related species (Shang et al. 2014).

While the origin of the peach is still being debated; this research was conducted in hopes of using genetic analysis applied to ancient peach seeds to compare the sequences generated to various wild species and modern peach specimens.

2.0 Methodological Background

One of the most critical steps in the retrieval of DNA from ancient samples is the removal of surface contaminants. This step is crucial in the prevention of contamination of ancient samples with modern DNA. The surface layer can be removed using a scalpel, dremel drill abrasion, UV irradiation, or soaking in a 10% sodium hypochlorite solution (bleach) (O'Rourke et al. 2000; Roberts and Ingham 2008). In order to optimise the efficiency of the extraction buffer, the sample should be ground up, or undergo a reduction in size in order to increase surface area (O'Rourke et al. 2000; Roberts and Ingham 2008). This can be achieved mechanically through the use of a grinder, mill or a mortar and pestle (O'Rourke et al. 2000; Roberts and Ingham 2008). The use of liquid nitrogen when preparing plant samples aids in grinding down the plant material into finer particles and protects the integrity of the cell without causing damage to the DNA molecule (Doyle 1991; Doyle and Dickson 1987; Varma et al. 2007).

2.1 DNA Extraction

The liberation of nucleic acids from a biological sample is perhaps the most critical step of its analysis. Selecting the most appropriate method of extraction is crucial for maximising the yield of DNA and often depends on the sample type. Plants provide a particular set of obstacles which must be taken into consideration when performing an extraction as they contain inhibitors which, if not removed can inhibit further analysis. The breaking down of the rigid cell wall and the removal of polysaccharides and polyphenols are of great importance in regards to contaminants which are present in plant samples (Guillemaut and Maréchal-Drouard 1992; Lodhi et al. 1994; Varma et al. 2007). Polysaccharides can pose a problem throughout the analysis of plant DNA as they are generally undetectable and have the ability to form complexes with DNA which can result in a sticky, viscous like consistency of the sample in the presence of an extraction buffer (Guillemaut and Maréchal-Drouard 1992; Lodhi et al. 1994; Murray and Thompson 1980; Varma et al. 2007). These complexes can also co-precipitate with DNA and interfere with enzyme activity and prevent amplification or inflate DNA quantity (Lodhi et al. 1994; Varma et al.

2007). Polyphenols are released from the vacuoles and are readily oxidised by cellular oxidases (Varma et al. 2007). They are able to interact irreversibly with the DNA molecule and cause degradation (Varma et al. 2007). The presence of polyphenols can be identified by the browning of a DNA pellet (Varma et al. 2007). In order to remove contaminants the initial extraction should be followed by a purification step to ensure the highest DNA quantity and purity.

The CTAB (Cetyl trimethylammonium bromide) method and the Proteinase K (PK) methods have been used and characterised for successful plant DNA extraction from various plant material sources including both seed and ancient samples (Aljanabi and Martinez 1997; Chunwongse et al. 1993; Doyle and Dickson 1987; Guillemaut and Maréchal-Drouard 1992; Hilz et al. 1975; Kang et al. 1998; Möller et al. 1992; Murray and Thompson 1980; Porebski et al. 1997; Rogers and Bendich 1985; Varma et al. 2007). The enzyme PK was the main proteolytic enzyme isolated from *Tritirachium album* Limber (Ebeling et al. 1974). The enzyme was desirable based on its keratin hydrolysing activity and ability to break peptide bonds adjacent to the carboxylic group of the aliphatic aromatic amino acids (Ebeling et al. 1974). The PK extraction has proven to be an effective enzymatic method of extraction, successfully breaking down proteins and removing them as potential inhibitors (Aljanabi and Martinez 1997; Chunwongse et al. 1993; Ebeling et al. 1974; Hilz et al. 1975; Kang et al. 1998). Further analysis determined that the activity of PK is increased by the presence of cells lysed by sodium dodecyl sulfate (SDS), and allowed for the inactivation of DNA degrading nucleases (Aljanabi and Martinez 1997; Ebeling et al. 1974). The SDS is an anionic surfactant which can be used as a protein denaturing agent and allows access for the protease (Hilz et al. 1975). This method has been well documented throughout the literature and has become a reliable resource for DNA extraction.

2.2 DNA Purification

2.2.1 Silica Bead Purification

A commonly used method to remove impurities is the silica based purification method (Boom et al. 1990). This method uses a guanidinium thiocyanate (GuSCN) salt solution, which aids the binding of

24

DNA to the silica beads that are placed in the solution (Price et al. 2009). Once incubated the DNA is bound to these beads and after centrifugation, form a pellet and the supernatant, containing all of the cellular debris can be removed (Price et al. 2009). A wash buffer is used to remove any unwanted particles. When these particles are removed, the DNA is eluted using a Tris/NaCl/EDTA buffer, resulting in a product that is both high in purity and quantity (Price et al. 2009). The silica based purification buffer consists of 4 M GuSCN, 0.1 M Tris HCl, 0.02 M EDTA, and 1.3% Triton X-100. High ionic strengths are needed in order to ensure the selectivity of DNA (Boom et al. 1990). The use of 4 M guanidinium thiocyanate acts as the chaotrophic salt for the extraction. Guanidinium thiocyanate has proven to be a powerful reagent which can both serve to lyse the cells, thereby releasing the DNA and all cellular components, as well as the ability to inactivate nucleases (Boom et al. 1990). The GuSCN aids in creating a hydrophobic environment which facilitates the binding of the DNA molecules to the silica bead resin (Boom et al. 1990). The phosphate backbone of DNA participates in hydrogen bonding with water, and therefore dehydration is needed to unbind water from the DNA (Boom et al. 1990). The silica beads used in the extraction and purification of DNA have a high specific surface area for DNA binding (Melzak et al. 1996). This is useful because very little to no other cellular debris are able to bind to the beads. The surface of the silica beads are positively charged at an acidic pH, which binds to the negative charge found in DNA; it is due to the GuSCN that the decrease in water activity occurs and changes the helical structure of DNA, as well as disrupts the hydrogen bonds of the DNA and exposes the negative charge of the phosphate (Boom et al. 1990; Melzak et al. 1996).

2.2.2 Micro Bio Spin Column with Bio-Gel P30

Micro Bio Spin Columns with Bio Gel P30 from Biorad can be used to further purify DNA after extraction and silica bead purification. The purification columns helps to remove agents of degradation, lytic enzymes or molecules which may further inhibit PCR. The columns are filled with a Bio-Gel P30 matrix which through size exclusion, effectively removes salts, and small molecules from proteins. The purified DNA is then suspended in a Tris based buffer and ready for amplification. The spin columns are

a fast and effective method of purification with minimal steps to lower the risk of further contamination and sample loss.

2.3 The Polymerase Chain Reaction

The PCR technique was developed by Kary Mullis in 1985 based on previous studies regarding the use of restriction endonucleases to isolate specific molecular fragments (Mullis et al. 1992; Saiki et al. 1988). The reaction involves the use of a thermostable polymerase, and the reciprocal interaction of two oligonucleotides (Mullis et al. 1992; Saiki et al. 1988). The forward and reverse primers are able to anneal to each strand of DNA and allows for their extensions to overlap. As a result of repetitive cycling, the quantity of a desired segment of DNA increases exponentially (Figure 7). These cycles consist of denaturing the DNA molecule, hybridisation of specific forward and reverse DNA primers to each strand, and finally extension by the DNA polymerase (Mullis et al. 1992; Saiki et al. 1988). With these advancements, PCR has opened new doors for the analysis of degraded, low copy, or specific gene fragments of DNA (Mullis et al. 1992; Pääbo et al. 2004; Saiki et al. 1988). Primers are chosen or designed based on the need to amplify a specific position on the DNA strand, the forward and reverse primer flank the desired region and are specific enough to only bind to that particular area (Mullis et al. 1992; Saiki et al. 1988).
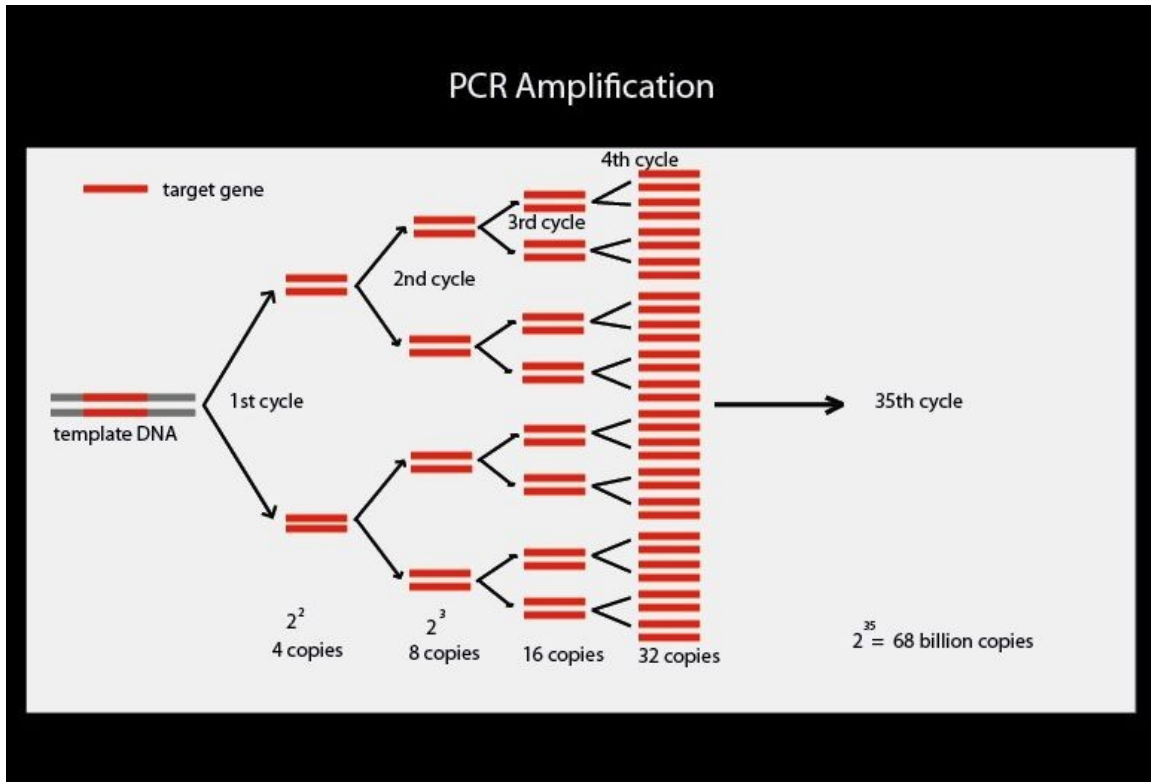
Figure 7 PCR amplification from a single template strand

A thermostable DNA polymerase was isolated from *Thermus aquaticus* (*Taq*) in 1976 with an optimum temperature of 80° C and can remain active in temperatures such as 95° C (Chien et al. 1976)**.** The polymerase requires all four deoxyribonucleotides and a divalent cation cofactor, often Mg^{2+}, for optimal activity (Chien et al. 1976). The discovery of the polymerase allowed for the simplification of PCR and allowed for the reaction to proceed at a high temperature, which previously was a problem due to the high temperatures needed to denature the DNA strand (Saiki et al. 1988). As a result, specificity, yield and sensitivity was increased, and thus allowed for fragments to be amplified from low quantity DNA samples (Saiki et al. 1988).

2.4 Sequencing

Fredrick Sanger and colleagues introduced a new method for DNA sequencing which incorporated the use of chain terminating inhibitors in 1977 (Sanger et al. 1977). The method was based on the previously published "plus and minus" method by Sanger and Coulson in 1975, which employed the use

of DNA polymerase to sequence a specific desired portion of the DNA strand (Sanger et al. 1977). Other

sequencing methods have included the use of restriction enzymes but have proven to be more laborious

(Sanger et al. 1977). Using the inhibitory effect of the 2',3'dideoxythymidine triphosphate (ddTTP) with

DNA polymerase I, it was demonstrated that extension of the nucleotide chain could be terminated, this is

due to the lack of 3' hydroxyl group (Sanger et al. 1977). With the addition of a mixture containing

deoxynucleotide triphosphates (dNTPs) and fluorescently labelled dideoxynucleotide triphosphates

(ddNTPs), to a sequencing reaction, the DNA sequence can be easily determined (Sanger et al. 1977). As

the ddNTPs are incorporated into the DNA strand, the reaction is terminated, and the result is varying

strand lengths that can be read using electrophoresis (Figure 8). The ddNTPs are also labelled with

fluorescent dyes so individual nucleotides can be detected using dye terminators. Capillary

electrophoresis applies an electric field to allow the negatively charged DNA fragments through the

capillary towards a positive electrode. The DNA fragments move through the capillary at different speeds

and pass through a detector which detects the fluorescent dyes that label the fragments.
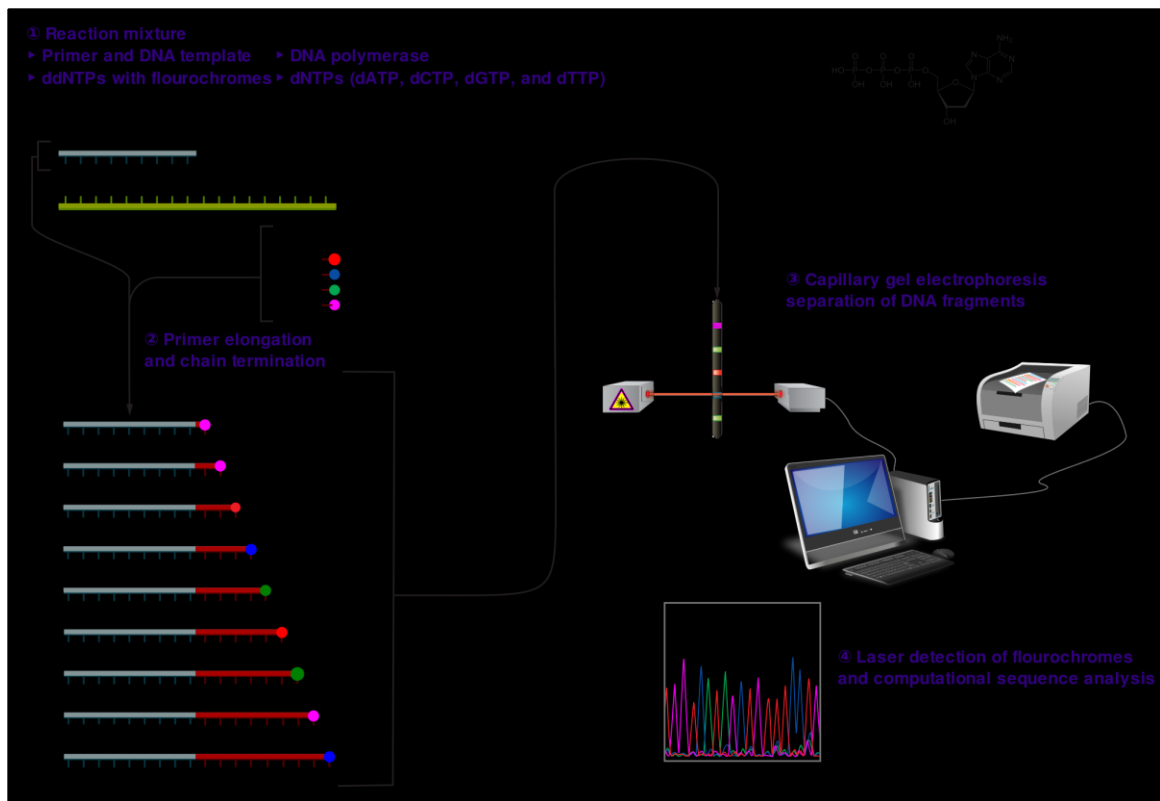
Figure 8. Sanger sequencing reaction

2.5 Phylogenetic Analysis of Sequences

Genomes are constantly evolving and accumulating mutations and the amount of differences between different genomes may indicate how recently the genomes shared a common ancestor (Baldauf 2003). Phylogenetic analysis is used to analyse sequences to infer or estimate evolutionary relationships. Phylogenetic trees depict these evolutionary relationships based on similarities and/or differences using phenotypic or genotypic traits. Groups of species or families which group together on the tree are believed to be descended from a common ancestor (Baldauf 2003).

There are a variety of different methods employed to generate phylogenetic trees including maximum likelihood and maximum parsimony. Generating a maximum likelihood tree involves finding the topology and branch lengths of the tree that will give the greatest probability of observing the DNA sequences used as data (Felsenstein 1981). This method finds the tree that maximises the probability of the genetic data given the tree. This method is often preferred as it accounts for difference in evolutionary rates in different lineages (Felsenstein 1981). In contrast maximum parsimony takes into account the identities of the ancestral and derived nucleotides and operates under the assumption that evolution follows the shortest possible route and therefore the correct tree is the one that requires the minimum number of nucleotide changes (Moret et al. 2002).

In order to assess the confidence limits of different branch points within a tree, bootstrap analysis is often performed along with individual tree analysis. New alignments of sequences are generated by taking columns at random from the original alignment and new trees are constructed (Baldauf 2003). A bootstrap value is assigned to each branch based on the number of times that branch pattern is reproduced at a particular node in each of the replicate trees (Baldauf

29

2003). A bootstrap value of approximately 0.70 or higher is considered a high degree of confidence.

2.6 Gas Chromatography Mass Spectrometry to Detect DNA Damage

There are a variety of different methods used to detect DNA damage; in this study gas chromatography- mass spectrometry (GC-MS) was used. GC-MS is very useful in identifying and quantifying the modification in DNA (Höss et al. 1996). It is able to identify multiple modified bases from all four bases in one DNA sample, products of sugar moiety, and other forms of DNA damage (Dizdaroglu et al. 2002; Halliwell and Dizdaroglu 1992). The use of GC-MS allows for the separation of molecules by gas chromatography and is conclusively identified by the structural evidence provided by the mass spectrometer (Halliwell and Dizdaroglu 1992).

For GC-MS analysis, DNA must be hydrolysed, derivatised, and injected (Beckman and Ames 1997; Dizdaroglu et al. 2002; Jenner et al. 1998). Hydrolysis of DNA involves the use of an acid to be added to the sample (Dizdaroglu et al. 2002). Hydrolysis is needed to release both modified and intact bases from the DNA (Dizdaroglu 1985; Dizdaroglu 1991; Dizdaroglu 1994). Following hydrolysis, which is usually achieved using formic acid, the DNA is derivatised, converting the DNA to a volatile compound (Dizdaroglu 1985; Dizdaroglu 1991; Dizdaroglu 1994; Halliwell and Dizdaroglu 1992; Jenner et al. 1998). The process converts the polar nucleosides/bases to volatile, thermally stable derivatives, which possess an unique mass spectra (Jenner et al. 1998). This can be accomplished by adding a trimethylsilyl (TMS) group to any free nitrogen or oxygen species catalysed by bis(trimethylsilyl)trifluoroacetimide (BSTFA). This step is carried out at high temperatures and the samples for analysis are purged with nitrogen to remove any oxygen to prevent artificial oxidation (Dizdaroglu et al. 2002; Jenner et al. 1998). The samples are separated on a fused silica capillary column in the GC instrument and the separated compounds are directed to the ion source of a mass spectrometer through an interface between the two instruments (Dizdaroglu et al. 2002). The mass spectra can be used for specific identification of the compound using the molecular peak (M), molecular ion peak (M^{+1}), and various other ionisation peaks

(eg. [M-15]$^+$) (Dizdaroglu 1984; Dizdaroglu 1985). Using the mass spectrometer for chromatographic detection also generates unique information regarding the chemical composition of the analyte (Abian 1999).

3.0 Methods and Procedures

To analyse *Prunus persica*, the first objective was to examine genetic targets to assess their ability to distinguish between *Prunus persica* and its various wild types was conducted. As a result, we hoped to identify a reliable target to identify *Prunus persica* at the species level. Finally through the use of different extraction methods, we hoped to evaluate the level of success for extraction DNA from peach seeds and finally, to apply these methods to successfully extract DNA from an ancient peach seed.

3.1 Analysis of Genes Sequenced in *Prunus persica* and its Wild Species

A search using All Databases in the National Center for Biotechnology Information (NCBI) was used to identify genes that have been sequenced in *P. persica* and its associated wild species *P. ferganensis, P. mira, P. kansuensis, P. davidiana*. Genes for analysis were chosen based on nucleotide sequences submitted in the NCBI database for at least four of the above species, and sequences were obtained for all cultivars, variaties (var), isolates, and forms (f). Eleven loci were chosen for phylogenetic analysis, *18S ribosomal* RNA gene, *maturase K* gene (*matK*), *ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit* gene (*rbcL*), *putative chloroplast RF1 protein Ycf1* gene (*ycf1*), *starch branching enzyme I* gene (*SbeI*), *NADH dehydrogenase subunit F* gene (*ndhF*), *psbA-trnH intergenic spacer*, *tRNA-Gly* gene (*trnG*), *tRNA-Leu* (*trnL*)- *tRNA-Phe* (*trnF*) *intergenic spacer*, *tRNA-Lys* gene (*trnK*) and *matK gene* partial sequence*, and *trnG-trnS intergenic spacer.* A phylogenetic analysis was performed for each locus to assess variation between *P. persica* varieties and the four wild species using MEGA 6.06. Sequences were first aligned and trimmed. A Maximum Parsimony tree was constructed using the alignment and a bootstrap value of 1000.

3.2 Sample Collection and Preparation

A sample of peach stones was first obtained from a commercial grocery store and used as a trial sample for primer assessment. Peach stone specimens were collected from the Zhejiang province of China, three stones were to be used as modern reference samples and one peach stone specimen was

collected to be used as an archaeological sample. The first stone was collected in Anjii, northwest of

Hangzhou, it was collected in the mountains but the classification of wild, feral, or cultivated could not be

determined. The shape of the stone demonstrates a more round look than typically observed in modern

cultivated/domesticated peaches. A second set of stones was collected beside the archaeological Maoshan

site. The stone is feral, the tree from which the sample was collected grew from its stone, and it was not

grown by cuttage or transplanting. A third stone was bought in a market in Shandong. The fourth stone

was an archaeological sample collected from the Maoshan site and was dated to the Liangzhu Culture

period (Figure 9 and Table 1). No additional measures were taken when collecting samples from sites

both present and archaeological. The stone recovered from the Maoshan site was preserved in

waterlogged context was reported not charred, and remained intact upon arrival (Zheng et al. 2014). All

peach stones were kept in dry isolated conditions at room temperature to prevent further damage to the

DNA and separated from all other samples to avoid contamination.

Figure 9. Location of sites where peach stones were collected in China.
A) Samples collected from the Moashan site, beside the Maoshan site, and samples collected from the Moutains in Anjii. B) Peach cultivar purchased from a market in Shandong.

Table 1. List of Sample Identification and details regarding the location and state of the individual peach stones.

Sample ID	Location of Sample	Country of Origin	Category	Condition of Peach Stone
1	Acquired from commercial grocery store	Canada	Modern	Intact
2	Anjii mountains	China	Modern	Intact
3	Beside the Maoshan site	China	Modern	Intact
4	Acquired from market in Shandong	China	Modern	Intact
5	Archaeological sample from Maoshan site	China	Ancient	Stone intact, inside appears burnt and seed black like charcoal

3.2.1 Preparation of Modern Samples

All procedures were conducted in a dedicated DNA laboratory in a designated hood at Lakehead University. The pit of the stone was opened using a Dremel drill, and the inner seed was removed. The peach seed was covered in liquid nitrogen and crushed using a mortar and pestle. Approximately 100 mg of powder was transferred into 1.5 mL microcentrifuge tubes. Samples 1, 2, 3 and 4 were processed using the above the method.

35
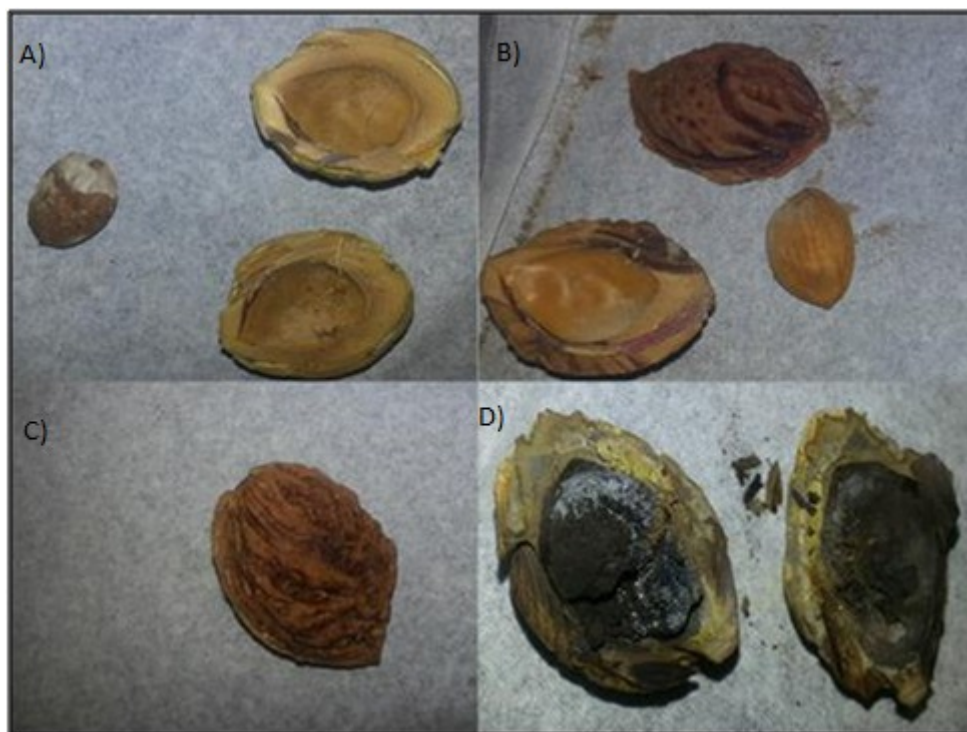
Figure 10. Peach samples collected from China.
A) Seed from the Anjii mountains, B) Feral Seed from beside Maoshan site, C) Cultivar from Shandong, D) Archeological seed from Maoshan site.

3.2.2 Preparation of Ancient Samples

All procedures were conducted on sample 5 in a dedicated clean lab at Lakehead University where no previous work on peaches had been conducted and no modern peach samples had been analysed. When performing analysis in the dedicated clean lab, all personnel are required to wear full body Tyvek suits complete with a hood, enclosed boots and sleeves. A primary pair of gloves is worn under the Tyvek suits, and a pair of secondary sleeves and gloves are worn on top of the suit. The secondary pair of gloves are changed between samples, when entering and existing the lab, and when switching between the various hoods in the lab. This is to ensure no carry over contamination occurs when preparing and prepping samples or reactions. In addition, a hair net is worn under the Tyvek suit, as well a face mask and eye goggles are worn over the suit to ensure no exposed skin or hair enters into the lab.

Prior to use, each hood is wiped down with bleach and left for 15 min. The bleach was wiped away with water, followed by ethanol, and then placed under UV light for 15 min. This protocol was also followed after the hood was used. Sample preparation, extraction, purification and PCR preparation of all ancient samples were performed in the clean lab. Once the PCR reactions were prepared, the samples were taken out of the clean lab and moved into the general analysis lab where the positive control of a modern peach sample was added and run on the thermocycler. Once the PCR reaction was complete, amplified samples were stored in this lab and gel electrophoresis was performed. Post-PCR purification and sequencing analysis was performed at the Paleo-DNA Laboratory at Lakehead University. In both general analysis areas, all work benchtops were cleaned before and after use with 70% EtOH. All equipment is regularly cleaned using a bleach wash followed by washing with EtOH. All consumables, tips and tubes are sterilized by autoclaving and UV radiation.

Preparation of the seed for the ancient sample followed the same procedure as used on the modern samples. An additional sample preparation was performed on Sample 5 using the stone of the peach. The stone was powdered using a mixer mill and approximately 200 mg was added to a 2.0 mL microcentrifuge tube.

Control samples were used throughout both modern and ancient sample analysis. Extraction negatives were used throughout the sample preparation steps, this tube was carried throughout the extraction process and tested for the presence of DNA to ensure that there was no contamination of the extraction reagents. Following the extraction and purification of samples, a PCR negative control was used throughout the remainder of analysis. This tube contains no sample and is tested for the presence of DNA to ensure that no PCR reagents are contaminated.

3.3 DNA Analysis

3.3.1 DNA Extraction
Proteinase K Extraction

A PK enzymatic extraction method was used to extract DNA from the peach seeds. The enzymatic extraction protocols have been widely used in the fields of molecular biology, and the PK

extraction method has been widely studied and validated for use in ancient DNA and forensic science. A modified PK method was used in this experiment (Hansen 1974). To the 1.5 mL microcentrifuge tube containing the peach seed powder, 290 µL of extraction buffer (10 mM Tris [hydroxylmethyl] aminomethane (Tris base), 1 mM Ethylenediaminetetraacetic acid (EDTA), 0.2 M Sodium Chloride (NaCl)), 40 µL 20% SDS, 23 µL water, and 5 µL PK enzyme (20 mg/mL) was added. The samples were vortexed and incubated at 56° C for 3 h at 500 x (g). The samples were centrifuged for 1 min at 13,000 x(g) and the supernatant was removed and placed into a new 1.5 mL microcentrifuge tube.

Silica-Spin Column Extraction Method

A silica spin column extraction method was used as it was shown a superior method for extracting plant aDNA compared to the DNeasy Plant Mini Kit and the DTAB/CTAB extraction (Moore 2011). To the sample, 1.5 mL of lysis buffer (0.5 M EDTA, 0.25% SDS, 0.5 mg/mL PK) was added and incubated at 50 °C at 500 x (g) overnight. An additional 50 µL of PK (20 mg/mL) was added to each tube and incubated for an additional 3 h at 50 °C at 500 x (g). The samples were centrifuged at 13,000 x (g) for 1 min and the supernatant removed and transferred to a new 2.0 mL tube. The supernatant was then transferred to a QIAquick® spin column, as part of the QIAquick Purification Kit (Qiagen). Washing of the samples was performed following the Qiagen QIAquick® Spin Handbook March 2008. Following the wash, the samples were eluted in 30 µL of elution buffer, and centrifuged into a new 1.5 mL microcentrifuge tube at 13,000 x (g)for 1 min.

The PK extraction method was performed on all modern samples, and used for both the ancient seed and stone samples. The silica-spin column extraction method was used for both the ancient seed and stone samples to compare different extraction methods to evaluate their ability to successfully extract DNA from different source material.

3.3.2 Purification
Guanidinium Thiocyanate (GuSCN) and Silica Bead Purification

Purification methods are employed to remove contaminants and inhibitors that may prevent amplification. A modified silica bead purification was performed based on the method by Boom et al.

(1990) and by Hoss and Paabo (1993). A volume of 1 mL of GuSCN solution (4 M GuSCN, 0.1 M Tris HCl, 0.02 M EDTA, 1.3% Triton X-100**)** and 7 µL of silica beads were added to the 1.5 mL microcentrifuge tube containing the peach seed extract. The solution was placed on ice overnight and then centrifuged for 1 m at 13,000 x (g). The supernatant was removed carefully, not disturbing the pellet and discarded. An aliquot of 500 µL of wash buffer (0.01 M Tris-HCl, 0.05 M NaCl, 1 mM EDTA, 50% Ethanol (EtOH)) was added to the silica pellet and vortexed until the pellet was dislodged from the bottom of the tube, ensuring that the silica was resuspended into the solution. The solution was centrifuge for 1 m at 13,000 x (g) and the supernatant discarded without disturbing the pellet. A volume of 200 µL of 100% EtOH was added to the pellet and vortexed until the silica pellet was dislodged from the bottom of the tube and resuspended in the solution. The solution was centrifuged for 1 m at 13,000 x (g) and the supernatant was discarded. The silica pellet was left to air dry overnight and resuspended in 75 µL of water to elute the DNA. The samples were incubated at 56° C at 300 x (g) for 1 h, the supernatant was then removed, careful not to disrupt the pellet and transferred to a new 1.5 mL microcentrifuge tube.

Micro Bio-Spin P30 ChromatographyColumn Purification

The P30 Micro Bio Spin Chromatography Columns were used to further remove impurities such as a salts and polysaccharides that may have been co-extracted or remained after the first purification by silica beads. The spin columns were inverted sharply to resuspend the gel and were placed in a 2.0 mL micro-centrifuge tube. The cap was removed from the column and the tip removed. The column was centrifuged for 2 min at 1,000 x (g) to remove the remaining packing buffer. The packing buffer was discarded and the column placed into a new 1.5 mL microcentrifuge tube. An aliquot of 75 µL of extracted DNA sample was transferred to the center of the column, and centrifuged for 4 min at 1,000x (g). The column was removed and discarded and the sample suspended in a Tris buffer. Purification with the P30 columns was performed twice, and the purified sample was stored in -20° C freezer.

3.3.3 DNA Quantification

Spectrophotometric methods were used to assess the quantity and purity of the peach samples using the Epoch spectrophotometer. DNA absorbs light at 260 nm and concentrations can be determined by measuring the absorbance at this wavelength. Proteins absorb light at 280 nm, and the purity of DNA can be determined by calculating the ratio of absorbance at 260 nm and 280 nm. A pure DNA sample will have a ratio of approximately 1.8, indicating little protein contamination. A volume of 2 µL of sample was loaded into individual wells on the Take 3 micro-well plate. The Gen5 software was used with the Epoch spectrometer, and the Take 3 experiment, using the Nucleic Acid DNA Quantification protocol to measure the absorbance at 260 nm and 280 nm of each sample loaded.

3.3.4 Assessment of Primers for DNA Analysis

All primer pairs (Table 2) were assessed before use in PCR for peach identification. Using the primer Basic Local Alignment Search Tool (BLAST) search on the National Centre for Biotechnology Information (NCBI) website, forward and reverse primer sequences were analysed. Primer pair specificity was assessed using the Genome (chromosome from all organisms) database parameters and the organism chosen was *Prunus* (taxid:3754). The top *Prunus* matches were selected and the sequences saved. Primer pairs were then compared to the primer BLAST matches using Amplify 3[©] to determine primer stability, binding sites, possibility of dimers, and amplicon size. Amplify 3[©] (Bill Engels, 2005, University of Wisconsin) is a program which simulates the PCR reaction and evaluates the primers.

Primer sequences were also analysed using the oligo analysis tool from Eurofins genomics (http://www.operon.com/). Forward and reverse primers were analysed using the analysis tool to determine the GC content and determine if there is a potential of self-hybridisation or hybridisation to the reverse primer.

Table 2. Microsatellite markers for peach initially used in this study.
Primers were assessed using Amplify 3[©] to determine the GC content, stability of the primer and the expected amplicon size.

Primer Name	Primer Sequence (Forward and reverse)	Repeat Type	GC Content (%)	Stability of Primer (%)	Expected Amplicon Size (bp)
Pchgms1	GGGTAAATATGCCCATTGTGCAATC GGATCATTGAACTACGTCAATCCTC	$(AC)_{12}(AT)_6$	44	100	194
Pchgms2	GTCAATGAGTTCAGTGTCTACACTC AATCATAACATCATTCAGCCACTGC	$(CT)_{24}$	44	100	163
Pchgms3	ACGGTATGTCCGTACACTCTCCATG CAACCTGTGATTGCTCCTATTAAAC	$(CT)_{14}$	52	100	179
Pchcms2	AGGGTCGTCTCTTTGAC CTTCGTTTCAAGGCCTG	$(CA)_8$	52.94	100	180
Pchcms4	CTCACGCTATTTCTCGG CCTCGACGAAGAGCTCG	$(CA)_9$	52.94	100	225
Pchcms5	CGCCCATGACAAACTTA GTCAAGAGGTACACCAG	$(CA)_9(TA)_8$	47	100	246

Table 3. Primers used in this study for both modern and ancient samples to identify samples at species level.
Primers were assessed using Amplify 3[©] to determine the GC content, stability of the primer and the expected amplicon size.

Primer	Primer Sequence (5'-3')	GC Content (%)	Primer Stability (%)	Expected Amplicon Size (bp)
Cob-F	GTCAGATGAGCTTTTGGGGA	50	100	187
Cob-R	GCCAGATGAAGAAGACTGGC			
Lhcb2-1F	ATCTCCCTTTCTGCCCTGCTAT	50	100	134
Lhcb2-1R	GGGTGTTTTTGTAGGCTTGAGGTA			
Lhcb2-2F	ATCTCCCTTTCTGCCCTGCTA	52	100	500
Lhcb2-2R	CAGAGGTTGCCATGACTATGC			

Table 4. Highly variable chloroplast markers used in this study for both modern and ancient samples.

Primer	Primer Sequence (5'-3')	GC Content (%)	Primer Stability (%)	Expected Amplicon Size (bp)
rcbL F	AGATTCCGCAGC CACTGCAGCCCCTGCTTC	63.3	100	200
rbcL R	ATGTCACCACAAACAGAGACTAAAGCAAGT			
psbM-F	TTTGACTGACTGTTTTTACGTA	31.8	100	500
trnD-R	CAGAGCACCGCCCTGTCAAG			
clpP-F	GCTTGGGCTTCTCTTGCTGACAT	58.8	100	500
clpP-R	TCCTAATCAACCGACTTTATCGAG			

3.3.5 Polymerase Chain Reaction

In an attempt to identify a set of peach specific primers, six sets of simple sequence repeats (Sosinski et al. 2000), an endogenous gene (*Lhcb2*) (Shang et al. 2014), and three highly variable chloroplast markers (Dong et al. 2012) were used. AccuStart[TM] *Taq* DNA Polymerase (Quanta BioSciences) was used to amplify DNA from various peach samples. Accustart[TM] *Taq* is a recombinant hot start *Taq* DNA polymerase containing monoclonal antibodies which bind to the polymerase. This causes the polymerase to remain inactive prior to PCR thermal cycling. In the initial heat activation at 95° C, the antibodies are denatured irreversibly and allow the fully activated DNA polymerase to be released. The AccuStart[TM] II PCR supermix (Quanta BioSciences) is a 2X concentrated ready-to-use PCR mastermix containing all necessary reagents except for the primers and templates used in the reaction. The supermix components include optimised concentrations of $MgCl_2$, dNTPs, reaction buffer, Accustart[TM] *Taq* DNA Polymerase, Accustart[TM] *Taq* antibodies and stabilisers. A 25 µL PCR reaction was performed in 0.2 mL PCR tubes. A PCR mastermix was made with final concentrations of 1X AccuStart[TM] II PCR Supermix, 10 µM each of forward and reverse primers (Eurofin MWG Operon) (see Tables 2, 3, 4), 5µL of DNA template, and water to bring the final reaction volume to 25 µL. Samples were vortexted for 20 s and placed in a 96 well Gradient Mastercycler (Eppendorf). The cycling parameters for the microsatellite markers (Table 2) included an initial denaturation at 94° C for 4 min followed by 32 cycles of 94° C for 30 s, 58° C for 30 s, and 72° C for 30 s, and a final extension at 72° C for 5 min. The cycling parameters for the endogenous *Lhcb2* gene included an initial denaturation at 95° C for 5 min followed by 40 cycles of 95° C for 30 s, 58° C for 30 s, and 72° C for 30 s, and a final extension at 72° C for 5 min. The cycling parameters for the chloroplast markers included an initial denaturation at 94° C for 3 min followed by 34 cycles of 94° C for 30 s, 52° C for 30 s, and 72° C for 2 min, and a final extension at 72° C for 5 min. Once the cycling was completed for all, the reaction was held at 7° C. A PCR negative control was run with all samples to ensure the sterility of all PCR reagents, along with a positive control of a modern peach sample extracted earlier.

Spiked PCR and Concentration of Extracts for final PCR Analysis

A spiked PCR was performed to determine the presence or absence of PCR inhibitors when the original PCR failed. Due to the degraded nature of aDNA, the *Lhcb2* gene was chosen for analysis for its small amplicon size. The previously described PCR method was followed but with the addition of 3 µL Sample 5 and 2 µL of either Sample 2 or 3. A final attempt to successfully amplify DNA from Sample 5 was made by combining all extracts in a 1.5 mL microcentrifuge tube and concentrated using a desiccator. The sample was resuspended in 25 µL of water and once again used for PCR analysis of the *Lhcb2* and the *rbcL* gene. A volume of 10 µL of sample was used in PCR in an attempt to increase success.

3.3.6 Gel Electrophoresis

PCR products were visualized using a 2% agarose gel electrophoresis and a 50 bp molecular marker (Fermentas). Samples were loaded with 5.0 µL of PCR product mixed with 3.0 µL of 6X Loading Dye (2.5% xylene cyanol, 2.5% bromophenol blue, 35% ficoll and 544.0 µL of water). Electrophoresis was run at 100 V for 45 m. The results were stained with ethidium bromide (BioRad) and visualized using a UV transilluminator. If the desired band size was present, the samples were then stored in -20° C until prepared for sequencing or use as a positive control for future ancient studies.

3.3.7 Post-PCR Purification

Prior to sequencing of PCR products, samples were purified to remove all unincorporated primers and dNTPs which may result in unsuccessful sequence reads. The ExoSAP-IT® protocol (Fisher Scientific) for post-PCR product clean-up was used to minimise sample loss (Bell 2008). This method includes a one-step enzymatic treatment using Exonuclease I (ExoI) and Shrimp Alkaline Phosphatase (Orlando et al.) added directly to the PCR product to degrade primers and dephosphorylate dNTPs that remain after the PCR reaction is complete. A volume of 2 µL of ExoI and 4 µL of SAP was added to each PCR reaction(Bell 2008) . The samples were placed on a thermocycler at 37° C for 15 m, followed by a 15 m incubation at 80° C to inactivate the enzyme.

3.3.8 DNA Sequencing

A modified sequencing method using chain terminating dideoxy nucleotides by Sanger et al. (1977) was used to sequence successful PCR products. In 0.2 mL PCR tubes, 0.5 μL Big Dye Terminator Cycle Sequencing Ready Reaction Mix (Life Technologies), 0.25 μL primer, 2 μL 5X buffer, 4.2 μL water and 3 μL of sample was added. Samples were briefly vortexed and run on the thermocycler using the following parameters: Initial denaturation at 96° C for 60 s 15 cycles of denaturation at 96° C for 10 s, annealing at 50° C for 5 s, and extension 60° C for 75 s, 5 cycles of denaturation at 96° C for 10 s, annealing at 50° C for 5 s, and extension 60° C for 90 s, and a final 5 cycles of denaturation at 96° C for 10 s, annealing at 50° C for 5 s, and extension 60° C for 2 min.

3.3.9 Post-sequencing Purification

An EtOH precipitation was used as a post sequencing purification. The purification was prepared in 0.5 mL microcentrifuge tubes in which 62.5 μL of cold 95% EtOH, 24.5 μL water, and 3 μL 3 M sodium acetate was added. All 10 μL of the sequencing reaction was added to the 0.5 mL tubes and left on the bench top for 30 m. Samples were then placed in the centrifuge at 13,000 x (g) for 30 m and the supernatant removed. A volume of 500 μL of 70% EtOH was added to each tube and briefly vortexted. The samples were then centrifuged at 13,000 (g) for 5 m and the supernatant removed. The samples were placed in a desiccator for 15 m in order to evaporate any remaining EtOH in preparation for sequence loading.

3.3.10 Sequence Loading

Hi-di formamide (Life Technologies) was used to maintain quality and conductivity of the solvent, 15 μL was added to each sample and vortexed for 1 m. The samples were incubated at 95° C for 3 m, and then placed on ice for 2 m. Samples were loaded onto a 364 well plate and loaded onto the ABI 3130xl genetic analyser. The instrument protocol used was Sequencing_36cmPOP4, and the analysis protocol used was 3130POP4_BDTv3_DB. A 36 cm capillary array was used and the polymer used was POP4 from Applied Biosystems.

3.3.11 Sequence Analysis

Sequences were analysed using BioEdit and MEGA 6.06. Sequences were aligned with reference sequences retrieved from BLAST searches (Altschul et al. 1990). A general BLAST for somewhat similar sequences was performed along with a search using the shotgun sequencing parameters specified for *Prunus* spp.

A phylogenetic analysis was performed using the sequencing results for each gene analysed in order to determine relationships between the ancient peach samples, modern peach cultivars, peach wild species, and other *Prunus* spp. which were determined from the BLAST analysis. Sequences for analysis were entered into MEGA 6.06 and aligned. Both a Maximum Parsimony tree and a Maximum Likelihood Tree were constructed using the alignment. A bootstrap analysis was also constructed for both maximum parsimony and the maximum likelihood trees. A maximum parsimony method for tree construction and phylogenetic analysis was chosen based on the case study by Dong et al (2012) which also used this method. All sites were used for Gaps/Missing Data Treatments with the nucleotide substitution method, and Sub-Tree Pruning and Regrafting was selected as the maximum parsimony search method. The maximum likelihood analysis was performed to compare results between the different statistical methods. All sites were used for Gaps/Missing Data Treatments, and analysis was carried out using a heuristic method of Nearest- Neighbour Exchange. Uniform rates were assumed among sites and a nucleotide substitution model was used.

3.4 Gas Chromatography Mass Spectrometry (GC-MS)

Approximately 200-500 ng of purified extract from Sample 5 was transferred into 2 separate GC vials. A volume of 500 μL of 60% formic acid was added to each GC vial and capped. Each sample was hydrolysed at 135 °C for 30 minutes on a thermomixer. The vials were removed from the heat and uncapped and left to freeze-dry overnight. Samples were derivatised by adding 400 μL of BSTFA-TMS solution and 600 μL of acetonitrile, the vials were purged with nitrogen and capped. The samples were the heated at 120° C for 30 m on a hotblock. Samples were run on the Varian model 450 gas

45

chromatograph coupled with a Varian model 300-MS quadrupole GMS mass spectrometer. The GC-MS mass spectrometer is equipped with factor four capillary column (VF-5 ms, 30 m x 0.25 mm ID, DF=0.25 um). Helium was used as the carrier gas at a flow rate of 1.0 mL/min. Samples were introduced via split mode with a one in ten split by an autosampler with the injection port at a temperature of 280° C. The oven temperature was initially 50° C for 1 minute then ramped up to 280° C at a rate of 10° C/min and then held for 6 minutes. Detection was under electrospray ionisation (EI) conditions, with ionisation energy 70 eV, ion source set at 200° C. The scan range was from 70 to 600 amu. The GC-MS interface temperature was set at 270° C. The quantitative analysis of major oxidative products was performed with GC-MS selected ion monitoring mode. The dwell time for each ion was set at 80ms. Output files were analysed using Varian MS workstation version 6 and the NIST98 Mass Spectral Database. A database and standards were used to generate a list of modified bases that were used to examine the aDNA of the archaeological peach sample (Table 5).

Table 5. Modified Bases and Molecular Weights Searched For in Study

Modified Base	Undamaged Base	Molecular Weight	Derivitised Molecular weights
8-hydroxyadenine	Adenine	FW 279.5	FW 367.7
4,6-diamino-5-formamidopyrimidine	Adenine	FW 153.1	FW 369.1
N6 methyladenine	Adenine	FW 149.2	FW 355.1
2-hydroxyadenine	Adenine	FW 151.1	FW 367.7
1-methyladenine	Adenine	FW 148.2	FW 364.2
2,6-diamino-4-hydroxy-5-formanodopyridimine	Guanine	FW 169.1	FW 457.9
7-methylguanine	Guanine	FW 171.2	FW 387.2
Oxazolone	Guanine	FW 130.1	FW 420.8
1-hydroxyguanine	Guanine	FW 167.1	FW 383.8
8-hydroxyguanine	Guanine	FW 167.1	FW 455.8
5-Formyluracil	Thymine	FW 140.1	FW 358.7
5-hydroxyhydantoin	Thymine	FW 115.1	FW 316.1
5-hydroxy-5-methyhydantoin	Thymine	FW 130.1	FW 346.6
5-Hydroxy-6-Hydrothymine	Thymine	FW 144.1	FW 360.7
5-hydroxymethyluracil	Thymine	FW 142.1	FW 358.7
Uracil	Thymine/ Cytosine	FW 113.1	FW 257.1
Thymineglycol	Thymine	FW 160.1	FW 448.9
5-OH-6-hydroperoxide	Cytosine	FW 161.1	FW 449.8
5,6-dihydrocytosine	Cytosine	FW 143.1	FW 431.8
5-hydroxy-6-hydrouracil	Cytosine	FW 130.1	FW 346.6
Uracil glycol	Cytosine	FW 146.0	FW 434.8
Cytosine glycol	Cytosine	FW 145.1	FW 433.8
5-hydroxy-6-hydrocytosine	Cytosine	FW 129.1	FW 345.7
5-hydroxycytosine	Cytosine	FW 127.1	FW 343.6
5-hydroxymethylcytosine	Cytosine	FW 141.1	FW 358.6
5,6-dihydrouracil	Cytosine	FW 114.1	FW 284.6
5-hydroxyuracil	Cytosine	FW 128.1	FW 344.6
Allantoin	purine	FW 158.1	FW 446.8
Alloxan	pyrimidine	FW 142.0	FW 432.8
4-amino-5-hydroxy-2,6(1H,5H)-pyrimidinedione	pyrimidine	FW 143.1	FW 431.8
2-amino-6, 8-dihydroxypurine	purine	FW 167.1	FW 383.6
2-aminopurine	purine	FW 135.1	FW 279.4
5,6-dihydro-3-methylcytosine	cytosine	FW 128.1	FW 272.5
5,6-dihydro-2-pyrimidinone	pyrimidine	FW 98.1	FW 170.2
5,6-dihydrothymine	thymine	FW 128.1	FW 272.4
Guanidinohydantoin	guanine	FW 155.1	FW299.5
Hydantoin	pyrimidine	FW 100.1	FW 244.4
5-hydroxyhydantoin	pyrimidine	FW 116.1	FW 332.6
2-hydroxyhypoxanthine	purine	FW 154.1	FW 370.6

8-hydroxyhypoxanthine	purine	FW 152.1	FW 296.4
5-hydroxy-8-oxoguanine	guanine	FW 183.1	FW 471.8
Hypoxanthine	purine	FW 136.1	FW 280.4
5-methylcytosine	cytosine	FW 125.1	FW 269.4
Uracil	pyrimidine	FW 112.1	FW 256.4

4.0 Results

4.1 Examination of Genetic Targets Previously Sequenced in *Prunus persica*

GenBank searches were performed to identify genes that have been sequenced in *P. persica* and its wild species. Eleven genetic targets were identified as being analysed from *P. persica* and some of the wild species. Phylogenetic analysis was performed on these 11 gene targets and phylogenetic trees were generated using MEGA 6.06 (Figures 11-16 and appendix). An assessment of these results was carried out to determine the ability of these genetic targets to distinguish between the different *Prunus* species effectively. Two maximum parsimony trees were generated for each genetic target analysed; the first included all sequences collected from GenBank (selected examples Figures 11, 13, 15, see also appendix) and the second included only published sequences (selected examples Figures 12, 14, 16, see also appendix).

The phylogeny of the *18S* (Figure 11 and 12), *matK* (Figure 13 and 14), *rbcL* (Figure 15 and 16) and the other eight genetic targets (Appendix) show little or no resolution between *P. persica* and the various wild peach species. No visible patterns or clustering of the same species were observed, these results were found in both trees generated using all sequences collected from GenBank and trees generated using only published sequences. No consistent patterns were found between trees generated using different genes. Similar results were found for all genes investigated including the *ndhF* gene (Appendix: Figures 33 and 34), *trnH-psbA* intergenic spacer (Appendix: Figures 35 and 36), *SbeI* gene (Appendix: Figure 37), *trnG-trnS* intergenic spacer (Appendix: Figure 38), and *trnL* gene (Appendix: Figure 39).
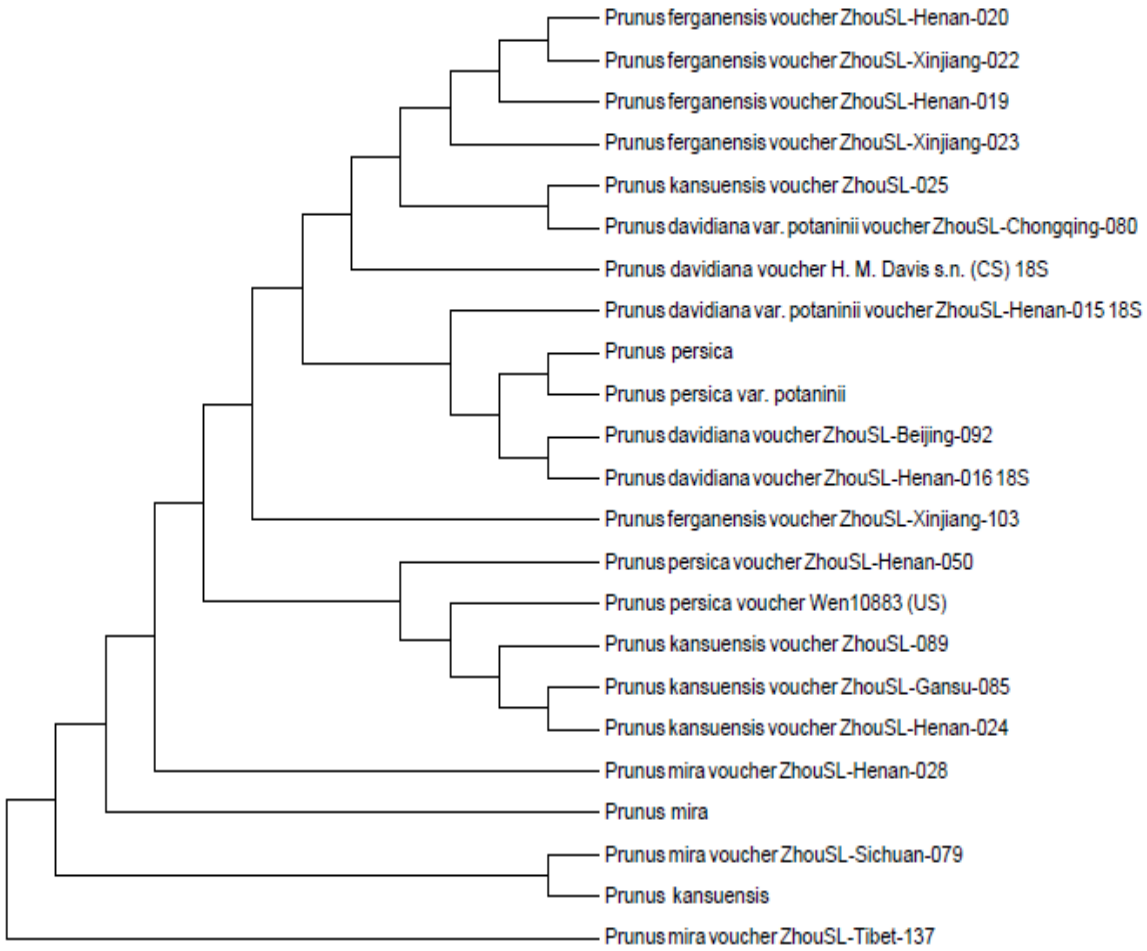
Figure 11. Phylogenetic analysis of *18S* ribosomal gene sequences
Maximum parsimony tree of various *Prunus persica* and wild species based on *18S* ribosomal gene (*18S*)
sequences. A bootstrap analysis of 1000 replicates was performed.
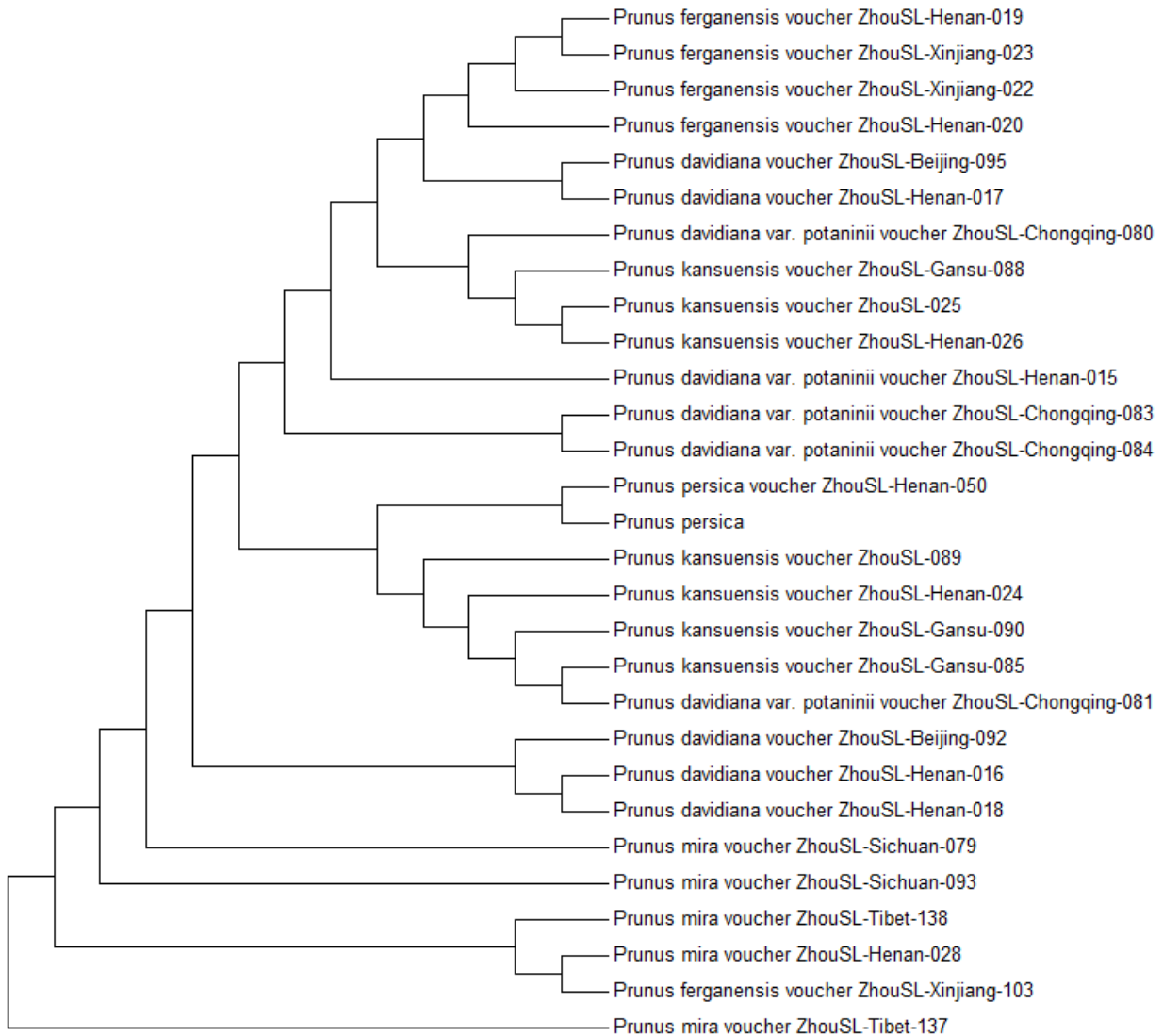
Figure 12. Phylogenetic analysis of published *18S* ribosomal gene sequences
Maximum parsimony tree of various published *Prunus persica* and wild species based on *18S* ribosomal gene (*18S*) sequences. A bootstrap analysis of 1000 replicates was performed.
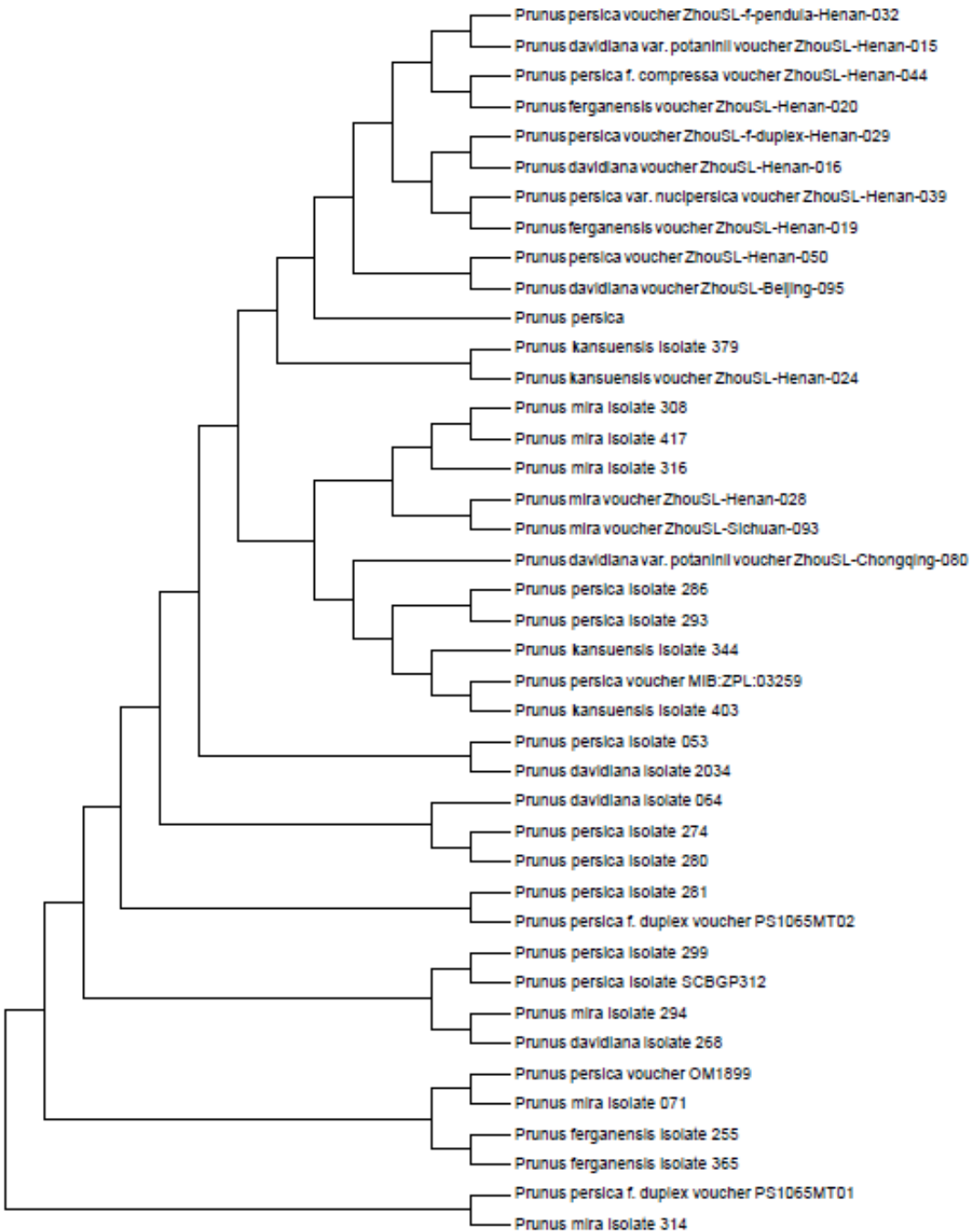
Figure 13. Phylogenetic analysis of *matK* gene sequences
Maximum parsimony tree of various *Prunus persica* and wild species based on the maturase K gene
(*matK*) . A bootstrap analysis of 1000 replicates was performed.
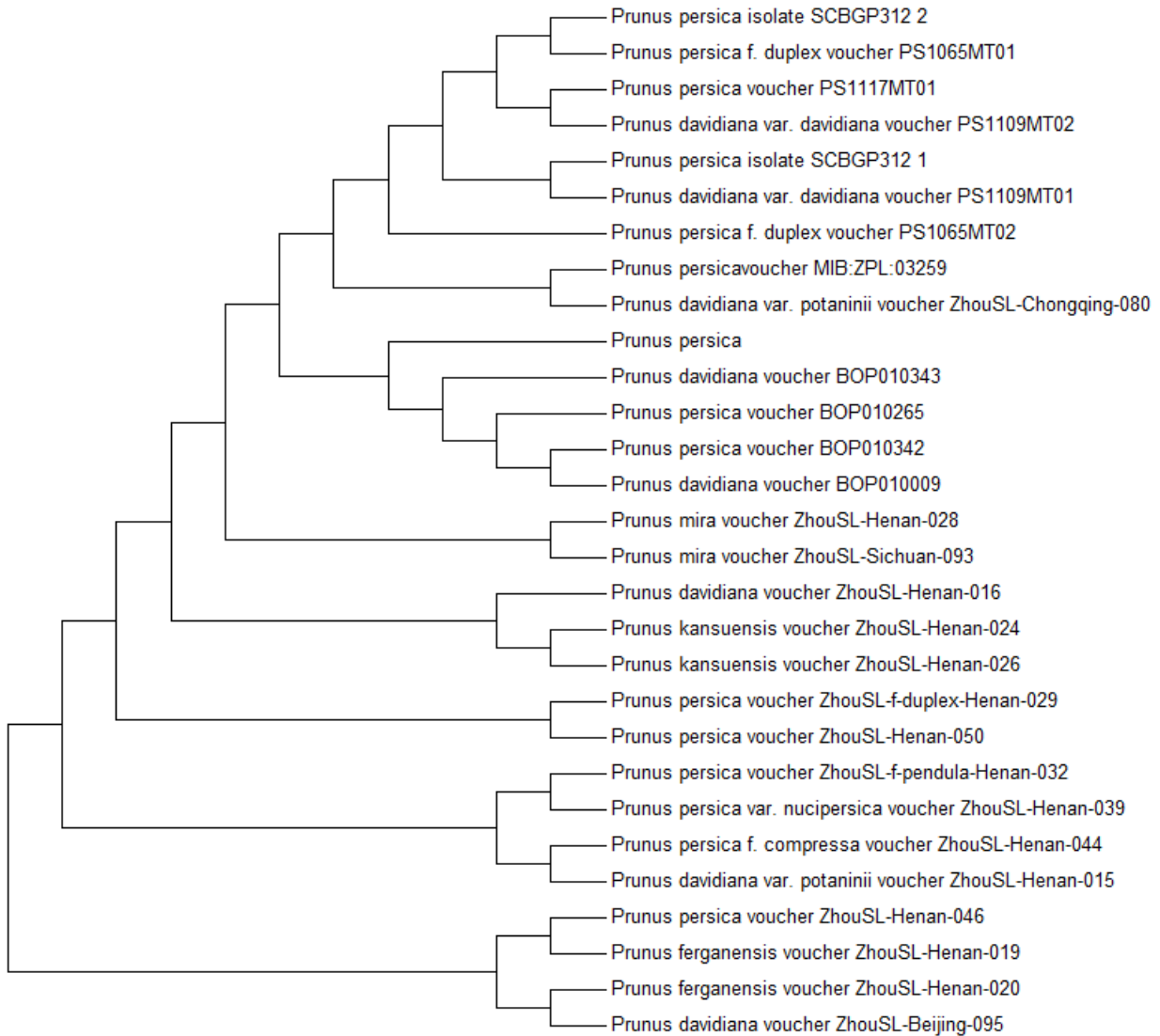
Figure 14. Phylogenetic analysis of published *matK* gene sequences
Maximum parsimony tree of various published *Prunus persica* and wild species based on maturase K gene (*matK*) sequences . A bootstrap analysis of 1000 replicates was performed**.**
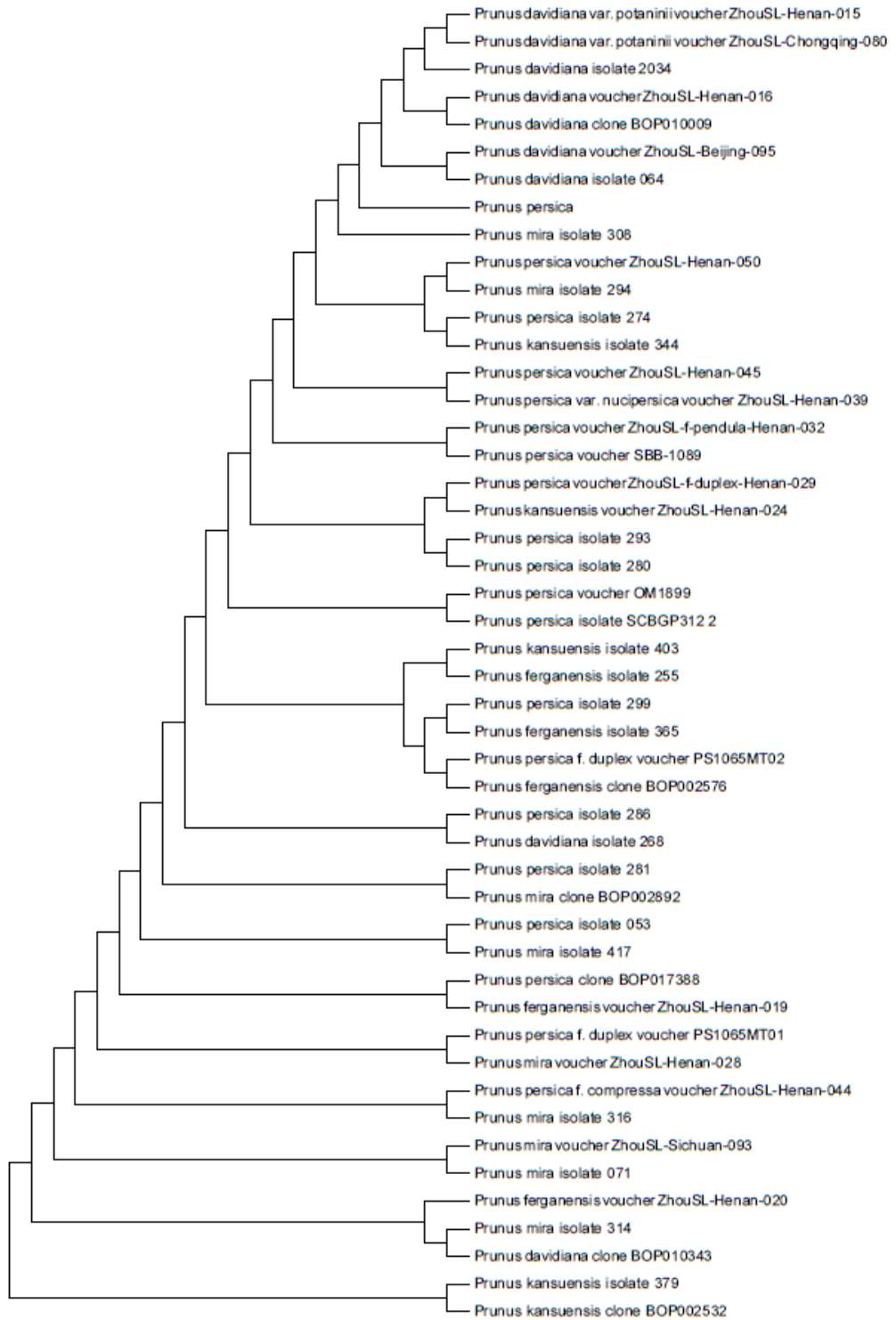
Figure 15. Phylogenetic analysis of *rbcL* gene sequences
Maximum parsimony tree of various *Prunus persica* and wild species based on ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap analysis of 1000 replicates was performed.
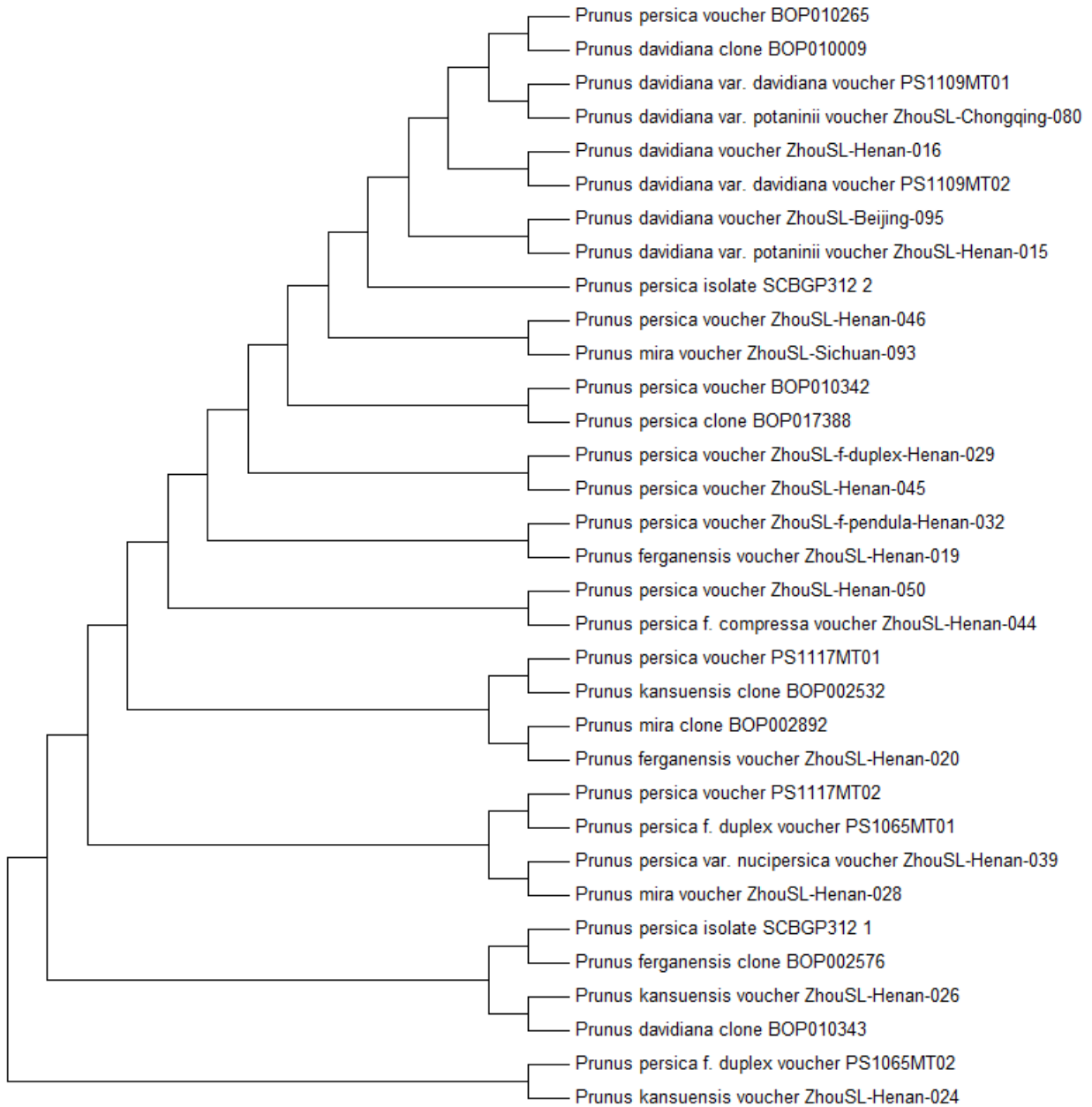
Figure 16. Phylogenetic analysis of published *rbcL* gene sequences
Maximum parsimony tree of various published *Prunus persica* and wild species based on ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap analysis of 1000 replicates was performed.

4.2 Analysis of Peach Samples

4.2.1 Assessment of Primers

Primers were assessed for their use for ancient DNA analysis. Optimisation of primers was performed on modern peach samples. Nuclear and chloroplast DNA markers were analysed using the

Oligo Analysis Tool from Eurofins Genomics (Figure 17) to determined the number and length of primer dimers that may be formed between primer pairs which may have an effect on their ability to successfully bind to the target DNA. Primer dimer complements of 4 nucleotides were found for one of the Lhcb2 primer set, and primer dimer complements of 5 nucleotides was found for both the clpP and psbM-trnD primer sets. Although observed, the degree of hydridisation between the primer sets was low enough to deem the primers acceptable for analysis.

A)

Longest primer-dimer complement: 0

B)

Longest primer-dimer complement: 4

C)

Longest primer-dimer complement: 5

D)

Longest primer-dimer complement: 5

E)

Figure 17. Oligo sequence analysis of the various primer pairs used in this study.
Primer sequences are compared against itself and another primer sequence to test for areas of hybridisation potential. Areas of hybridisation are marked with an 'X' and areas of 4 or more continuous 'X's are highlighted in yellow. A) Lhcb2-1 primer set, B) Lhcb2-2 primer set, C) rcbL primer set, D) psbM-trnD primer set, E) clpP primer set

4.2.2 Quantification of Extracted Peach Samples

All samples were found to have sufficient DNA concentrations to proceed with PCR analysis. Table 6 details the purity and DNA concentrations found using two different extraction methods, the Proteinase K enzymatic method used for all modern and ancient samples, and the silica-spin column method used for the ancient sample. The purities varied between samples, and were found to be both greater than and lower than the accepted 1.8 purity reading for DNA indicating that the samples contained some form of contamination. The extraction negatives resulted in very low concentrations of DNA or were too low to be detected but were included in the PCR analysis to ensure there was no DNA contamination from other sources. High DNA concentrations for the Proteinase K extraction method of modern peach seed samples, commercial peach (Sample 1), sample from the mountains of Zhejiang (Sample 2), the feral stone from beside the Maoshan site (Sample 3), and the stone purchased from a market in Shandong (Sample 4), were found to be sufficient for DNA analysis. The Proteinase K method resulted in much higher yields of DNA for both the ancient peach seed and stone in comparison to the silica-spin column in which no detection of DNA was found (Sample 5). Neither method resulted in DNA purity that were desirable for genetic analysis.

Table 6. Purity and concentration of Peach samples and extraction negatives.
Absorbance readings were read at 260 nm and 280 nm of 2 µL of sample using the EPOCH
spectrophotometer.

Sample	Extraction Method	Source of DNA	Purity (A260/A280)	DNA Concentration (ng/µL)
Sample 1	Proteinase K	Seed	0.846	47.186
Sample 2 extraction negative	Proteinase K	Water	ND	ND
Sample 2	Proteinase K	Seed	1.426	1106.015
Sample 3 extraction negative	Proteinase K	Water	1.9915	ND
Sample 3	Proteinase K	Seed	1.4745	751.453
Sample 4 extraction negative	Proteinase K	Water	1.327	ND
Sample 4	Proteinase K	Seed	1.0425	196.681
Sample 5 extraction negative	Proteinase K	Water	2.104	ND
Sample 5	Proteinase K	Seed	1.225	156.4735
Sample 5 extraction negative	Silica-Spin column	Water	ND	ND
Sample 5	Silica-Spin column	Seed	2.25	ND
Sample 5	Silica-Spin column	Stone	2.051	ND
Sample 5 extraction negative	Proteinase K	Water	1.267	9.908
Sample 5	Proteinase K	Stone	1.177	17.415

*ND- not detected

4.2.3 Amplification Results and Gel Electrophoresis

Successful amplification of the *rbcL* and *clpP* chloroplast genes was found only from modern

peach seed samples and amplification was not successful in both the ancient peach stone and seed (Figure

18). Lanes 3 and 8 contained bands at approximately 200 bp and 1,000 bp respectively, corresponding

with the expected band sizes of the genetic targets. Amplification of the *Lhcb2* gene was successful in the

modern sample but not the ancient sample (Figure 20). Lanes 3 and 7 contained bands at approximately

less than 200 bp. In order to rule out inhibition in Sample 5, the archaeological sample from the Maoshan

site, a spiked PCR was performed using 2 µL of a modern peach sample added to 5 µL of Sample 5. The

spiked PCR analysis resulted in successful amplification, indicating a lack of inhibition of the ancient

samples (Figures 19 and 21). A complete list of all primer sets amplified and the observed amplicon size

and sequence length for all samples can be found in Table 7.
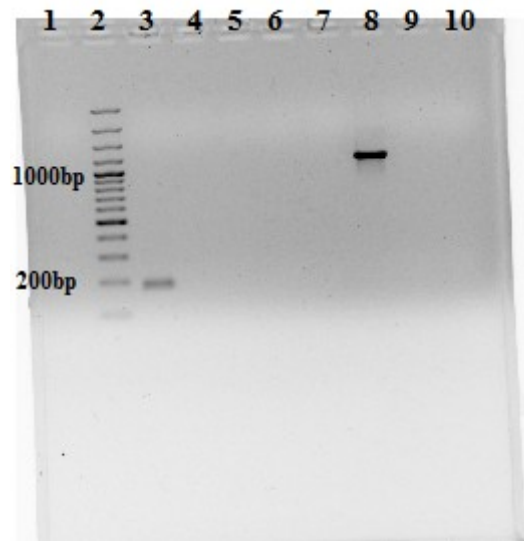
Figure 18. Amplification of the *rbcL* gene and *clpP* intron from silica spin column extractions of ancient peach seeds and stone
A 1% agarose gel showing amplification of peach DNA from various samples. 5 µL of PCR product was loaded into each well. Lane 2 contains 3 µL of GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific). Lanes 1, 3,4, 5 and 6 contain PCR products of the *rbcL* gene. Lane 1. PCR negative control. Lane 3 Peach positive control extract from peach seed from the mountains in Zhejiang Province. Lane 4 Silica spin column extraction negative. Lane 5 Silica spin column extract from seed collected from the Maoshan site in China. Lane 6 Silica spin column extract from stone collected from the Maoshan site in China. Lanes 7, 8, 9 and 10 contain PCR products of the *clpP* intron. Lane 7 PCR negative control. Lane 8 Peach positive control extract from peach seed from the mountains in Zhejiang Province. Lane 9 Silica spin column extraction negative. Lane 10 Silica spin column extract from seed collected from the Maoshan site in China.
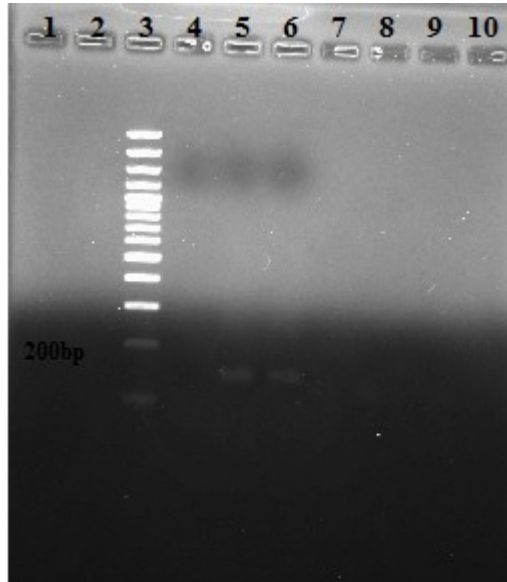
Figure 19. Spiked PCR products of the *Lhcb2* gene from silica-spin column extraction
A 1% agarose gel showing amplification of a spiked PCR of a ancient peach seed and stone. 5 µL of PCR product was loaded into each well. Lanes 1 and 2 contain no DNA. Lane 3 contains 3 µL of GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific). Lane 4. PCR negative control. Lane 5 Spiked PCR product containing 5 µL of Silica spin column extract from seed collected from the Maoshan site in China and 2 µL of extract from peach seed from the mountains in Zhejiang Province. Lane 6 Spiked PCR product containing 5 µL of Silica spin column extract from the stone collected from the Maoshan site in China and 2 µL of extract from peach seed from the mountains in Zhejiang Province.
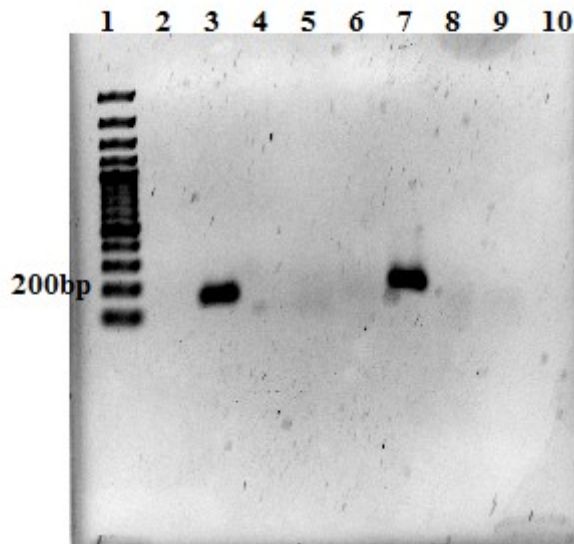
Figure 20. Amplification of the *Lhcb2* gene and *rbcL* gene from Proteinase K extractions of an ancient peach stone

A 1% agarose gel showing amplification of peach DNA from various samples. 5 µL of PCR product was loaded into each well. Lane 1 contains 3 µL of GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific). Lanes 2, 3, 4 and 5 contain PCR products of the *Lhcb2* gene. Lane 2. PCR negative control. Lane 3 Peach positive control extract from peach seed from a feral peach collected from the Zhejiang province. Lane 4 PK extraction negative. Lane 5 PK extract from peach stone collected from the Maoshan site in China. Lanes 6, 7, 8 and 9 contain PCR products of the *rbcL* gene. Lane 6. PCR negative control. Lane 7 Peach positive control extract from peach seed from a feral peach collected from the Zhejiang province beside the Maoshan site. Lane 8 PK extraction negative. Lane 9 PK extract from peach stone collected from the Maoshan site in China.
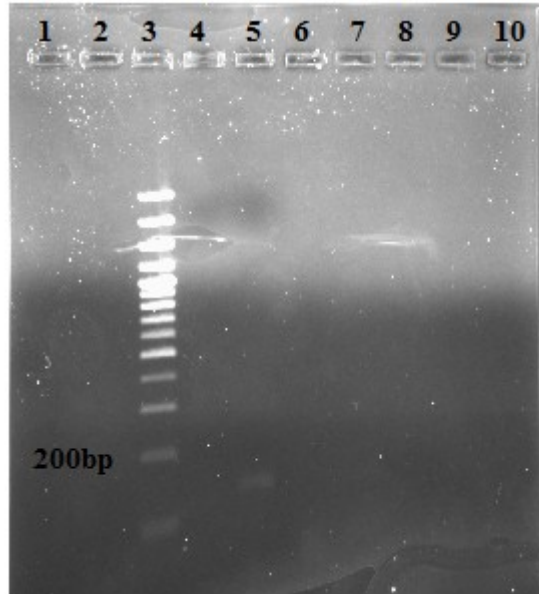
Figure 21. Spiked PCR products of the *Lhcb2* gene from Proteinase K Extraction
A 1% agarose gel showing amplification of a spiked PCR of a ancient peach seed and stone. 5 µL of PCR product was loaded into each well. Lanes 1 and 2 contain no DNA. Lane 3 contains 3 µL of GeneRuler 100 bp Plus DNA Ladder (Thermo Scientific). Lane 4. PCR negative control. Lane 5 Spiked PCR product containing 5 µL of PK extract from a peach stone collected from the Maoshan site in China and 2 µL of extract from a feral peach seed collected in Zhejiang Province beside the Maoshan site.

Table 7. Amplification results of Modern and Ancient Peach Samples

Primer	Sample 2		Sample 3		Sample 4		Sample 5	
	Observed Band Size (bp)	Sequence Length (bp)	Observed Band Size (bp)	Sequence Length (bp)	Observed Band Size (bp)	Sequence Length (bp)	Observed Band Size (bp)	Sequence Length (bp)
Pchgms1	200	151	200	150	200	142	ND	ND
Pchgms2	ND	ND	ND	ND	ND	ND	ND	ND
Pchgms3	175	131	175	125	ND	ND	ND	ND
Pchcms2	200	132	200	141	ND	ND	ND	ND
Pchcms4	ND	ND	ND	ND	ND	ND	ND	ND
Pchcms5	ND	ND	ND	ND	ND	ND	ND	ND
Cob	200	156	200	162	ND	ND	200	158
Lhcb2-1	150	107	150	110	ND	ND	ND	ND
Lhcb2-2	500	460	500	450	ND	ND	ND	ND
rbcL	200	121	200	121	ND	ND	ND	ND
psbM-trnD	1000	488	1000	411	ND	ND	ND	ND
clpP	1000	509	1000	488	ND	ND	ND	ND

**ND- not detected

The microsatellite repeat sequences from Samples 1, 2 and 3 were compared to sequences from four different peach cultivars, 'Dr. Davis', 'Lovell', 'F8-1', and 'Georgia Belle'. The number of repeats for each microsatellite marker can be found in Table 8. The microsatellite loci *pchgms1* $(AC)_{12}(AT)_6$, revealed polymorphisms in samples 1, 4 and all four cultivars. The microsatellite *pchgms3* $(CT)_{14}$, revealed polymorphisms in sample 3, 'Dr. Davis' and 'Lovell' cultivars. Finally polymorphisms were observed in the *pchcms2* microsatellite $(CA)_8$ in samples 2 and 3 and all four peach cultivars. Polymorphisms differed only by one repeat except in the case of *gms3* where $(CT)_{11}$ was observed in sample 3 and $(CT)_{26}$ was observed in 'Dr.Davis' and 'Lovell' peach cultivars.

Table 8. Microsatellites observed in this study for Modern and Ancient Peach samples and modern Peach cultivars collected from BLAST

	Number of Repeats								
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Peach 'Dr. Davis' cultivar	Peach 'Lovell' cultivar	Peach 'F8-1' cultivar	Peach 'Georgia Belle'
Repeat Type									
$(AC)_{12}(AT)_6$ (gms1)	$(AC)_{11}$ $(AT)_6$	$(AC)_{12}$ $(AT)_6$	$(AC)_{12}$ $(AT)_6$	$(AC)_{12}$ $(AT)_5$	ND	$(AC)_{11}$ $(AT)_6$	$(AC)_{11}$ $(AT)_6$	$(AC)_{11}$ $(AT)_6$	$(AC)_{11}$ $(AT)_6$
$(CT)_{14}$ (gms3)	$(CT)_{14}$	$(CT)_{14}$	$(CT)_{11}$	ND	ND	$(CT)_{26}$	$(CT)_{26}$	$(CT)_{14}$	$(CT)_{14}$
$(CA)_8$ (cms2)	ND	$(CA)_9$	$(CA)_9$	ND	ND	$(CA)_9$	$(CA)_9$	$(CA)_9$	$(CA)_9$

**ND- not detected

4.2.3 Phylogenetic Analysis of Samples

In an attempt to identify a genetic target or genes suitable for peach species specific identification, sequences generated using a nDNA target (*Lhcb2*) and cpDNA targets (*clpP, psbM-trnD, rbcL*) were aligned to other *Prunus* species to determine their effectiveness. Sequences generated using the chlorophyll a/b-binding protein (*Lhcb2*) endogenous reference gene from samples 1, 2, and 3 were aligned to various *Prunus persica* cultivars and other *Prunus* species (Figures 40 and 41) using MEGA 6.06 and a maximum parsimony tree and a maximum likelihood tree was generated (Figures 22 and 23). Sequence alignments can be found in the Appendix. All *P. persica* grouped together with a bootstrap value of 100 and formed a clade separate from the other *Prunus* species analysed indicating that the *Lhcb2* gene is a good candidate for peach species identification.

Figure 22. Phylogenetic analysis of the *Lhcb2* gene sequences in peach samples
Maximum parsimony tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus armerniaca* and *Prunus mume* using the chlorophyll a/b binding protein (*Lhcb2*) gene. A bootstrap analysis of 1000 replicates was performed.

Figure 23. Second Phylogenetic analysis of the *Lhcb2* gene sequences in peach samples
Maximum likelihood tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus armerniaca* and *Prunus mume* using the chlorophyll a/b binding protein (*Lhcb2)* gene. A bootstrap analysis of 1000 replicates was performed.

The use of highly variable chloroplast markers to distinguish between *P.persica* and other wild peach species was evaluated using phylogenetic analysis. The maximum parsimony tree generated using *clpP* intron sequences did not resolve between the different peach samples 2 and 3 analysed in this study and other wild *Prunus* species (Figure 24). The maximum likelihood tree did provide some resolution between the peach samples analysed but other *Prunus persica* cultivars were still distributed throughout the tree (Figure 25). It was determined that the *clpP* gene would not be sufficient as an identifying gene. Similar results were found for both the *psbM-trnDI* intergenic spacer (Figures 26 and 27) and the *rbcL* gene (Figures 28 and 29).
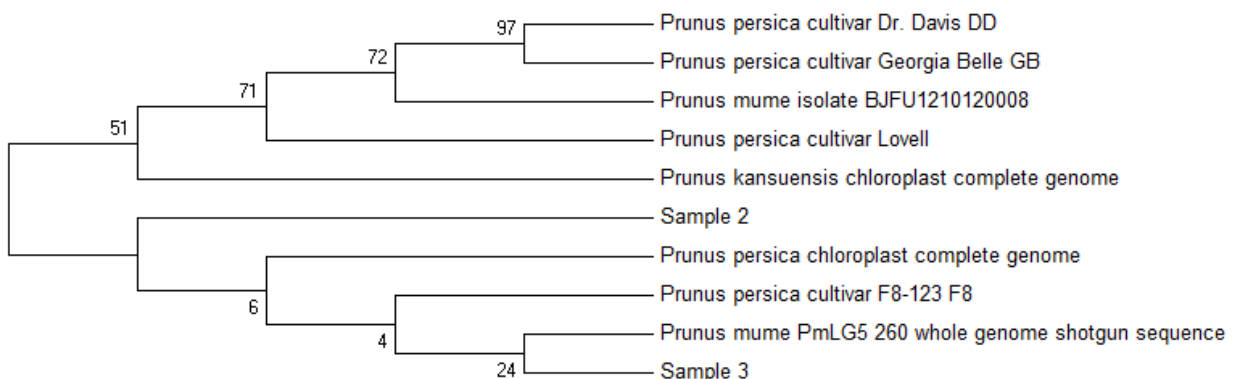
Figure 24. Phylogenetic analysis of *clpP* intron sequences in peach samples
Maximum parsimony tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild tspecies, and *Prunus mume* using *clpP* intron sequences. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.

66
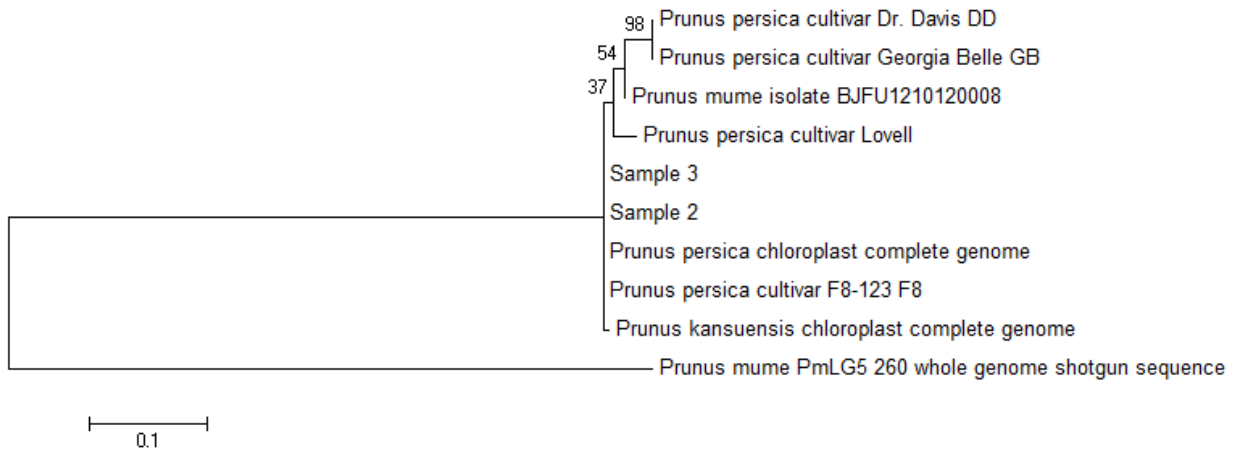
Figure 25. Second phylogenetic analysis of *clpP* intron sequences in peach samples
Maximum likelihood tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild species, and *Prunus mume* using *clpP* intron sequences. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.
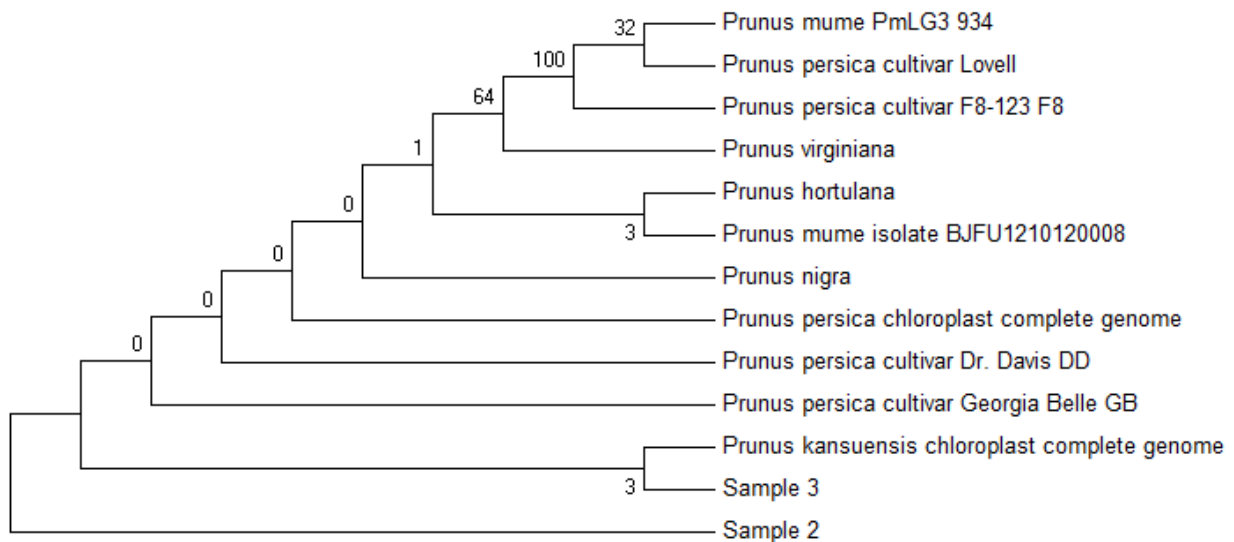
Figure 26. Phylogenetic analysis of *psbM-trnD* intergenic spacer sequences in peach samples
Maximum parsimony tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild species, and other *Prunus* species using *psbM-trnD* intergenic spacer seqeunces. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.

Figure 27. Second phylogenetic analysis of *psbM-trnD* intergenic spacer sequences in peach samples
Maximum likelihood tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild species, and other *Prunus* species using *psbM-trnD* intergenic spacer sequences. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.
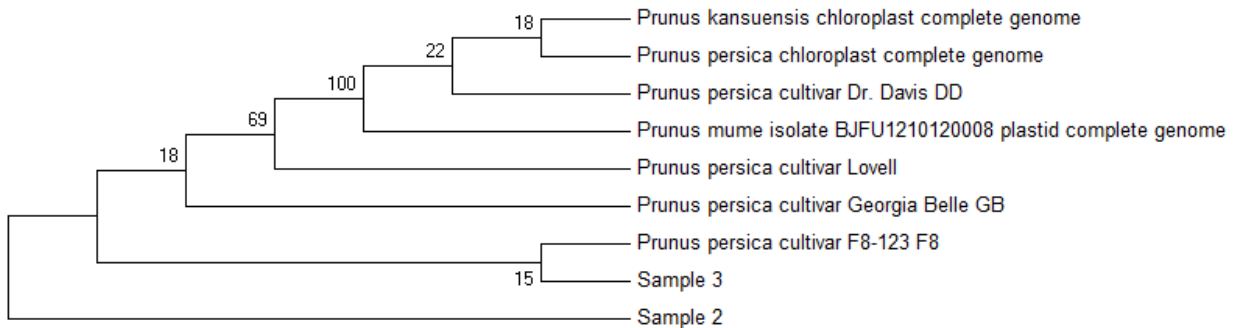
Figure 28. Phylogenetic analysis of *rbcL* gene sequences of peach samples
Maximum parsimony tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild species, and *Prunus mume* using ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.

68
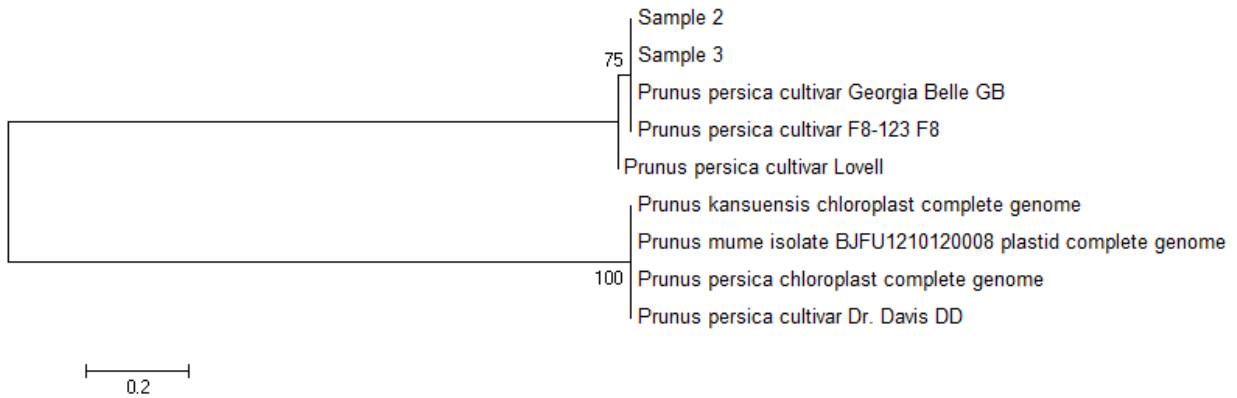
Figure 29. Second phylogenetic analysis of *rbcL* gene sequences of peach samples
Maximum likelihood tree generated of Samples 1, 2 and 3 with various *Prunus persica* cultivars, *Prunus* wild species, and *Prunus mume* using ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap analysis of 1000 replicates was performed. All sequences were chosen based on matches using BLAST.

The *rbcL* sequence data from the Chinese peach stones was aligned with sequence data collected from GenBank used previously for all peach wild species (Figure 30). The maximum parsimony tree did resolve some of the wild species, distinct branches of *P. davidiana* and *P. ferganensis* are evident but with these branches are sequences from *P. kansuensis* and *P. davidiana.* Sample 2 and 3, the Chinese peach stones, group together but still separate from other wild peach species and *P. persica* samples. Similar results were found when all the published sequences, Chinese peach stone samples, and BLAST matches were aligned (Figure 31).

69
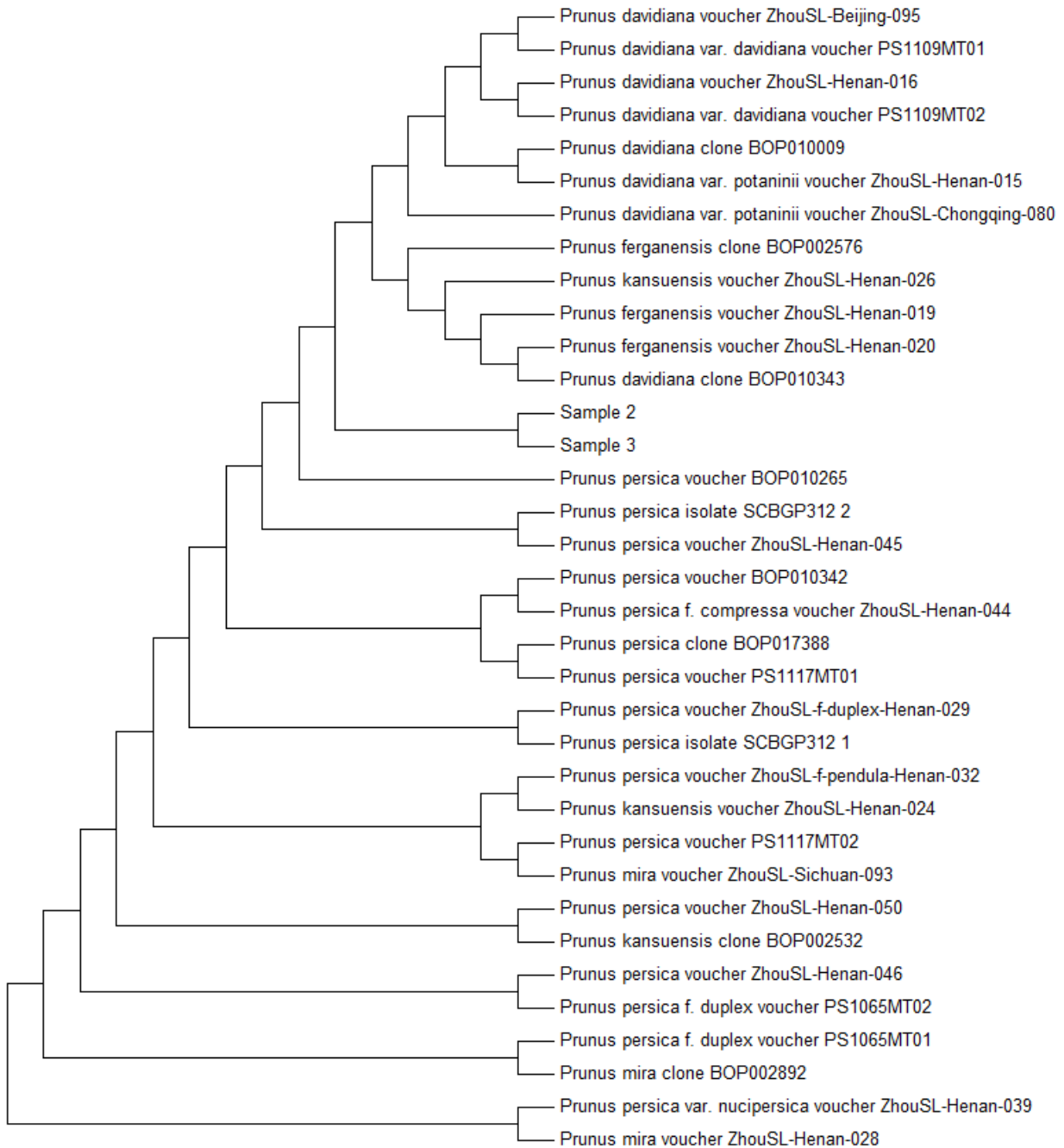
Figure 30. Phylogenetic analysis of *rbcL* gene sequences of multiple *Prunus sp* and peach samples
Maximum parsimony tree generated of Samples 1, 2 and 3, published *Prunus persica* and wild species
using ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap
analysis of 1000 replicates was performed.
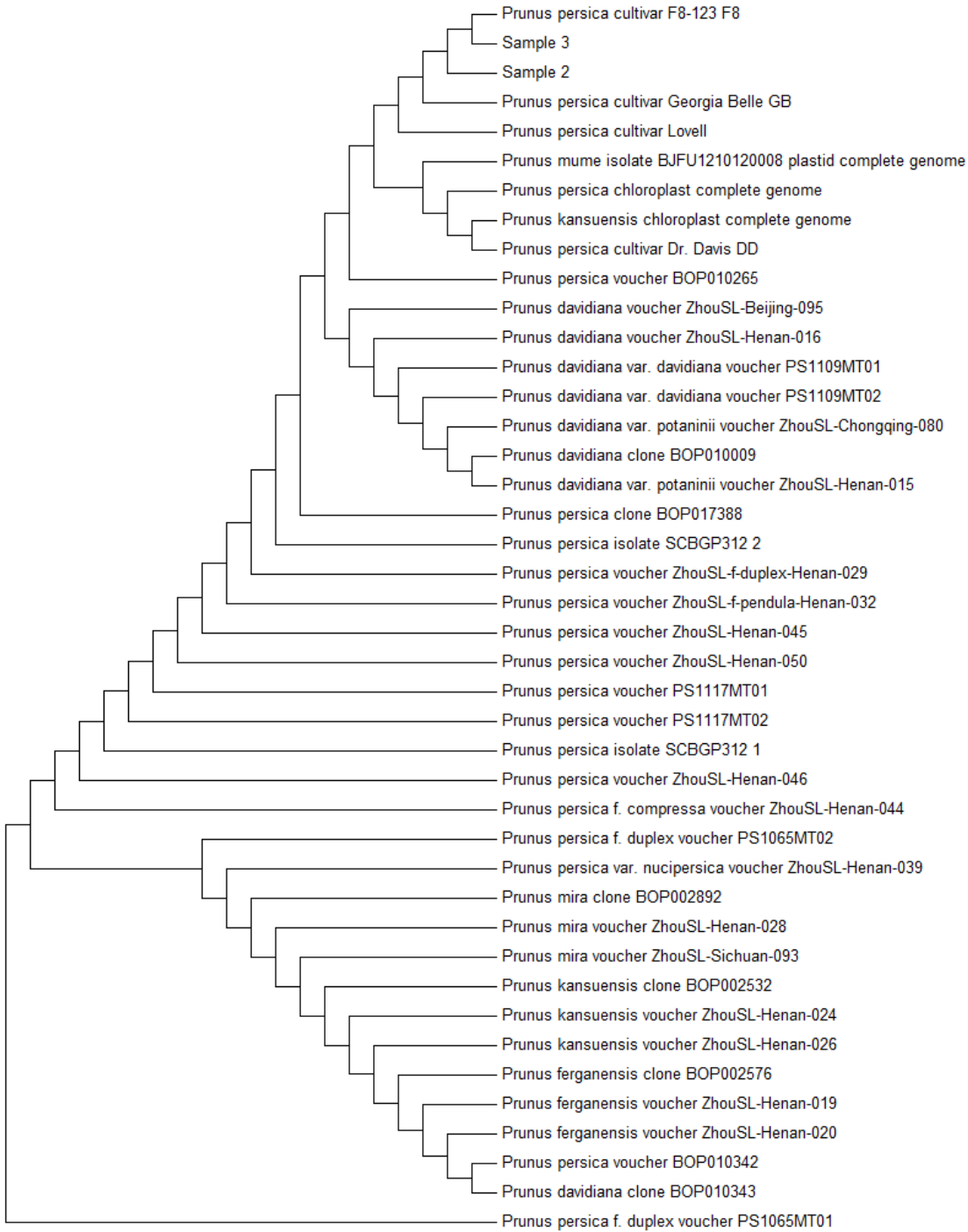
Figure 31. Phylogenetic analysis of published *rbcL* gene sequences of multiple *Prunus* sp. and peach samples

Maximum parsimony tree generated of Samples 1, 2 and 3, published *Prunus persica* and wild species, and sequences obtained from a BLAST search of samples using ribulose-1,5-biphosphate carboxylase/oxygenase large subunit gene (*rbcL*) sequences. A bootstrap analysis of 1000 replicates was performed.

4.2.4 GC-MS Analysis of aDNA Samples

The use of GC-MS analysis was applied to Sample 5 DNA extracts to determine possible reasons that may have caused a lack of amplification in the ancient sample. A number of modified bases were identified in the aDNA samples using a list that was generated using a database and standards that can be found in Table 5. Table 9 details the modified bases found in the peach seed extracts (1-1 and 1-2) and the type of damage that is associated with them. Damage caused by oxidation, hydrolysis, methylation and reduction were all observed in both extracts, all of which can lead to failure of amplification of the DNA..

Table 9. Modified Bases detected in the archaeological samples

Modified Base	1-1	1-2	Damage
8-hydroxyadenine	Y*	N	Oxidation
4,6-diamino-5-formamidopyrimidine	Y	Y	Oxidation
N6 methyladenine	Y	Y	Methylation
2-hydroxyadenine	Y*	N	Oxidation
1-methyladenine	N	N	Methylation
2,6-diamino-4-hydroxy-5-formanodopyridimine	Y	Y	Oxidation
7-methylguanine	Y	N	Methylation
Oxazolone	N	N	Oxidation
1-hydroxyguanine	N	N	Oxidation
8-hydroxyguanine	Y	Y	Oxidation
5-Formyluracil	N	N	Oxidation
5-hydroxyhydantoin	N	N	Oxidation
5-hydroxy-5-methyhydantoin	N	N	Methylation, Oxidation
5-Hydroxy-6-Hydrothymine	N	N	Oxidation
5-hydroxymethyluracil	N	N	Oxidation
Thymineglycol	Y	Y	Oxidation, Hydrolysis
5-OH-6-hydroperoxide	N	N	Oxidation
5,6-dihydrocytosine	N	N	Reduction
5-hydroxy-6-hydrouracil	N	N	Oxidation, Hydrolysis
Uracil glycol	N	N	Oxidation, Hydrolysis
Cytosine glycol	N	N	Oxidation
5-hydroxy-6-hydrocytosine	N	N	Oxidation
5-hydroxycytosine	N	N	Oxidation
5-hydroxymethylcytosine	N	Y	Oxidation, Methylation
5,6-dihydrouracil	N	N	Reduction
5-hydroxyuracil	N	N	Oxidation, Hydrolysis
Allantoin	Y	Y	Oxidation
Alloxan	N	N	Oxidation
4-amino-5-hydroxy-2,6(1H,5H)-	N	N	Oxidation

pyrimidinedione			
2-amino-6, 8-dihydroxypurine	N	N	Hydrolysis, Oxidation
2-aminopurine	N	N	Hydrolysis, Reduction, Oxidation
5,6-dihydro-3-methylcytosine	N	N	Methylation, Oxidation
5,6-dihydro-2-pyrimidinone	Y	Y	Hydrolysis, Reduction
5,6-dihydrothymine	N	N	Reduction
Guanidinohydantoin	Y	Y	Oxidation
Hydantoin	N	N	Oxidation
5-hydroxyhydantoin	Y	Y	Oxidation
2-hydroxyhypoxanthine	Y	N	Hydrolysis
8-hydroxyhypoxanthine	N	N	Hydrolysis, Oxidation
5-hydroxy-8-oxoguanine	N	N	Oxidation
Hypoxanthine	N	N	Hydrolysis
5-methylcytosine	N	N	Methylation
Uracil	N	Y	Hydrolysis

* Both modified bases share the same ion peaks and are very difficult to distinguish between. It is possible that only one of these is present.

5.0 Discussion

5.1 Analysis of Genes Previously Sequenced in *Prunus persica*

A search for various genes that have been sequenced in *P. persica* and its wild species *P. ferganensis, P. davidiana, P. mira* and *P. kansuensis* was performed using GenBank. In total, eight genes were found sequenced in all the *Prunus* species investigated, *ndhF*, *trnH-psbA*, *SbeI*, *trnG-trnS*, *18S*, *matK*, *rbcL*, and *trnL*. The genes were mostly cpDNA and have been used for evaluation of candidates for barcodes used for plant DNA analysis. The *18S* gene is a ribosomal RNA gene which has commonly been used in evolutionary studies as it is present across the Eukaryotic genome (Figure 11 and 12). The NADH dehydrogenase F (*ndhF)* gene is found in the chloroplast in plants and is a long sequence which has been found to provide a higher degree of phylogenetic information (Figures 34 and 35) (Kim and Jansen 1995). The *trnH-psbA* (Figures 35 and 36) and *trnG-trnS* (Figure 38) loci are intergenic spacers located in the chloroplast genome and are highly variable. The maturase K *(matK)* gene (Figure 13 and 14) and the large subunit of the ribulose-bisphosphate carboxylase (*rbcL*) gene (Figure 15 and 16) have extensively been used for phylogenetic studies for cpDNA (Dong et al. 2012). The *trnL* gene is a tRNA gene that has been used for barcode analysis of cpDNA (Figure 39). The starch-branching enzyme I (*SbeI*) is a nuclear gene which is still being studied for its use to differentiate between different species using nDNA (Figure 37).

Phylogenetic analysis was performed to generate a Maximum Parsimony tree for each gene using all the sequences collected from BLAST. Previous studies investigating these genes have shown using phylogenetics that species can be resolved in *Prunus* (Dong et al. 2012; Quan and Zhou 2011). A general trend expected to be seen was each individual species would group together on individual branches with *P. persica* at the top of the tree. Based on the results from the International Peach Genome Initiative, it was expected to see no differentiation between *P. persica* and *P. ferganensis* as they have been determined to be indistinguishable from one other (International Peach Genome et al. 2013). Two trees were generated where possible for each gene investigated; the first included most sequences collected

from GenBank with no discrimination, the second included only sequences that had been published in a scientific journal for comparison.

For all genes investigated the results were not as expected. Overall, we were not able to successfully group together individual *Prunus* sp. There was no consistency between the different genes, *P. persica* was found dispersed throughout and has grouped together with all the wild species. We also did not see the expected grouping of *P. persica* and *P. ferganensis* which is in contrast to the results of the two species being indistinguishable in sequence. Trees generated using only published sequences still did not serve to resolve individual *Prunus* species. No visible pattern was observed, and no expected trends or groupings were found in any of the trees. When comparing these trees to the supplementary material published by Dong et al (2012), none of the trees demonstrate the same branching patterns (Figure 32). The *trnH-psbM* intergenic spacer unpublished sequences did produce two major clusters of species, the first consisting of *P. persica,* and *P. kansuensis* sequences, and the second consisting of *P. mira, P. ferganensis,* and *P. davidiana* species (Figure 35). Yet, once again, we do not see the expected separation of individual species that has been found in other publications.

The methods for obtaining and analyzing the sequences differed from Dong et al (2012) as they used PAUP to generate the maximum parsimonious trees using the same setting as Quan and Zhou (2011). In all analyses, Clustal was used to align sequences, but previous studies reported gene lengths of approximately 600 bp, in this study gene sequences were approximately 100-300 bp in length. This may have contributed to differences in results. All characters were equally rated and multiple trees were analysed in all cases. The consensus trees were used in this particular study but all trees were analysed by Quan and Zhou (2011). Through the use of different analyses methods, this may have contributed once again to the discrepancies observed between previously analysed genes and the results found in this study.

These results demonstrated the obstacles that can arise when attempting to perform phylogenetic analysis. A high degree of variation was seen within sequences from the same species as was seen between species, leading to difficulty distinguishing between the cultivated *P. persica* sequences and the various wild species. The use of these highly variable genes has been shown to resolve most *Prunus* sp.

75

and when in combination, can be used to resolve all five species. No data could be found to compare the use of these genes to determine the diversity within species. It is possible that variation within species may be reduced if we were to look at a single population from a small geographic region. In this study, sequences from Beijing Botanical Gardens, Gansu, Xiahe; Tibet, Lhasa; Norbulingka; Henan Xinxiang, and other regions in China were chosen from published articles but are not within the same geographic region (Quan and Zhou 2011). This may contribute to the greater variation seen in the trees, as these genes are highly variable, and as cultivation and spread of the species occurred, the selection of favourable traits may have caused this. Variation among the wild species may have occurred due to different selective environmental pressures based on the different locations they are located in. A previous study investigating genetic distances between cultivars from China, Japan, Korea, and North America using SSR markers classified the cultivars into six groups, which were based on geographical regions and cultivation. Cultivars with similar ecotypes tended to group together and cultivars with favourable traits which were chosen to bring over to North America also grouped together (Yoon et al. 2006).

For future work regarding the use of genes studied in *P. persica* for phylogenetic studies, it is important to take into consideration that variation within a species across a wide geographic region can be as great as variation between species. This can cause discrepancies when creating a phylogenetic tree that can appear to give no distinct trends. One possible way to correct this would be to choose sequences from similar geographic regions or ecotypes in order to distinguish between *P. persica* and its wild species. It is also important when choosing sequences for analysis, to ensure that sequences have been published in order to confirm that the sequences are accurate and have been verified.
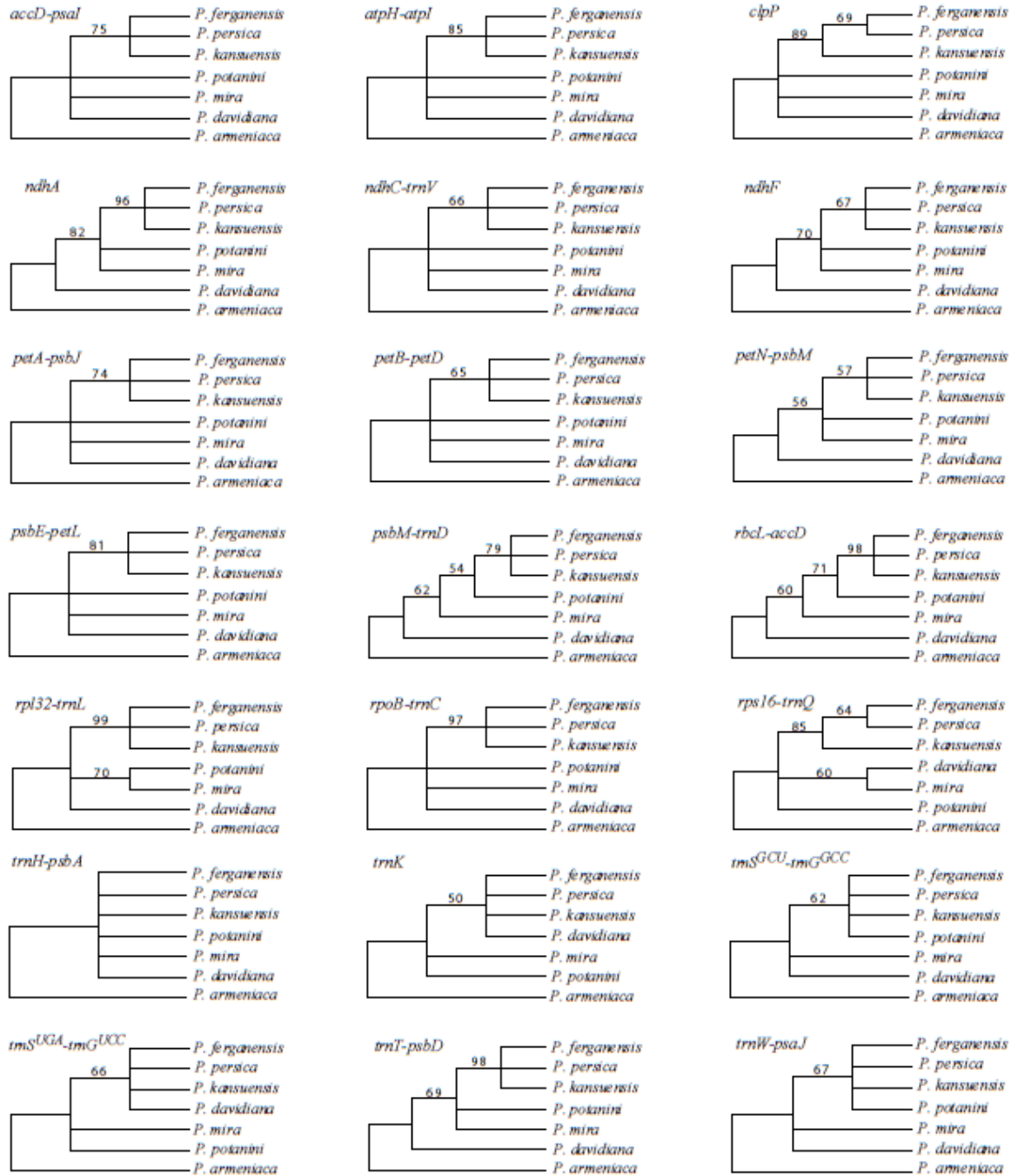
Figure 32. Maximum parsimony trees of six peach species based on 21 chloroplast loci.
The figure shows the resolution of the six species with bootstrap values included for the clades (Dong et al. 2012).

5.2 Primer Assessment

The use of a variety of different primer sets for peach identification was assessed using previously published primers. Microsatellites, nDNA markers, and cpDNA markers were first assessed using Amplify3$^{©}$ and the Oligo Analysis Tool from Eurofins Genomics (Figure 17). The GC content (%) can be found in Tables 2, 3, and 4 and were all within an acceptable range, and the primer stability for all primers sets was calculated to be 100%. The expected amplicon size was calculated and compared to those expected in the literature. Overall, the expected amplicon sizes were very close to the published values and were acceptable for further analysis. The use of most of the primers was desirable as most of amplicon sizes were favourable for aDNA analysis (approximately less than 200 bp), where there is a high likelihood for degraded DNA. Primer pairs *pchcms4*, *pchcms5*, *Lhcb2-2*, *psbM-trnD*, and *clpP* all had expected amplicon sizes of 500 bp or greater which would be unfavourable for aDNA analysis but were still used for modern analysis on peaches collected from China.

5.3 DNA Analysis on Modern Commercial Peach

A peach was purchased from a commercial grocery store for use in preliminary DNA work to assess the quality of primers chosen and to investigate the efficiency of the PK extraction method and silica bead purification. The seed was chosen as a source of DNA and was ground using liquid nitrogen. Approximately 100 ng of seed powder was used for analysis and a modified PK extraction method was performed. This method proved successful and resulted in a high yield, 47.186 ng/µL, for further analysis (Table 6). Of the six microsatellite primer pairs used for PCR analysis, only two were successful in amplifying the peach DNA and producing viable sequences for analysis, *pchgms1* and *pchgms3*. Successful amplification of DNA was achieved using *cob*, *Lhcb2-1*, *Lhcb2-2*, *rbcL*, *psbM-trnD*, and *clpP* primer pair sets, and the expected amplicon size was confirmed through gel electrophoresis.

5.4 DNA Analysis on Modern Peach Samples from the Zhejiang Province, China

Peach stones were collected from various parts of the Zhejiang Province of China and used for DNA analysis. The seeds were removed from the stone and the modified PK extraction and silica bead purification was performed as with the modern commercial sample. As seen in Table 6, all controls were negative for the presence of DNA, and samples from the mountains of Zhejiang (Sample 2), a feral stone from beside the Maoshan site (Sample 3), and a stone purchased from a market in Shandong (Sample 4), all resulted in very high yields of DNA, 1106.015 ng/μL, 751.453 ng/μL, and 196.981 ng/μL, respectively (Table 6). The purity of samples ranged between 1.0 and 1.4 which is not ideal, but did not have an effect on downstream analysis in most cases. One concern when using the EPOCH spectrometer is the inability to distinguish between the different types of DNA. Therefore, the purity and also the DNA concentration can be inflated or misrepresented by the presence of human DNA, plant DNA, microbial DNA etc. PCR analysis using the various microsatellites did not result in a high success rate (Table 7). Only 3 of the 6 primer sets used resulted in successful amplification, *pchgms1*, *pchgms3*, and *pchcms2*. No amplification was observed using the primer sets *pchgms2*, *pchcms4*, and *pchcms5*. Due to amplification observed with other microsatellite markers, inhibition was ruled out as a possibility. Successful amplification from the extraction of the peach stone purchased from a market in Shandong occurred only with the *pchgms1* primer set, and none of the other primer sets used in this study. It is possible that no chloroplast DNA was extracted in this sample or that the DNA was damaged during the extraction process. Product loss may have also occurred throughout the purification process which resulted in no amplification. The entirety of the sample was used in the analysis and therefore further experiments could not be conducted. Successful amplification of the *Lhcb2 chlorophyll a/b binding protein* gene was observed in Samples 2 and 3, both the 500 bp amplicon and 150 bp amplicon was observed (Figure 20). These results were of particular significance as this is a desirable amplicon size for ancient DNA analysis. The three chloroplast barcode genes, *rbcL*, *psbM-trnD* intergenic spacer, and *clpP* were all successfully amplified in Samples 2 and 3 (Figures 18 and 20). Successful amplification was

confirmed through the use of agarose gel electrophoresis and samples were then sequenced for further analysis.

5.5 Microsatellite Analysis

Of the three microsatellite loci which were successfully amplified, none matched for *P. persica* when input into BLAST using the nucleotide collection database. When the whole genome shotgun contigs database was chosen and the organism *P. persica* was specified, different cultivar matches could be found. The microsatellite loci sequences for four different cultivars were located, 'Dr. Davis', 'Lovell', 'F8-1', and 'Georgia Belle'. The number of repeats for each locus can be found in Table 8. Unfortunately there was not enough information from this study to accurately be used for phylogenetic studies. We were unable to locate frequencies for the observed polymorphisms as these primers were originally used to create linkage groups when early studies to map the peach chromosomes were being designed. The original purpose was to choose primers that could be used to successfully identify peach DNA through PCR. Based on the BLAST results, these primers may be used for *Prunus* analysis but are not exclusive to peach, and may also amplify other organisms. It would be useful to conduct further studies using these microsatellite primer sets on a greater sample size to assess which polymorphisms are observed and determine the frequencies for further use.

5.6 Assessment of a Peach Specific Gene *Lhcb2* as a Definitive Identifier for Peach Analysis

A search to find a genetic marker or gene which was specific to *P. persica* was undertaken in hopes to find a suitable identifier which could be used for ancient DNA analysis. The goal was to find a small target region that is unique to *P. persica* and would be able to distinguish peach from other *Prunus* species. The chlorophyll a/b-binding protein (*Lhcb2*) endogenous reference gene was identified as a species specific gene for peach by Shang et al (2014) and was investigated for the use of identification of the peach seed samples. Successful amplification of both the 500 bp product and the 134 bp product were confirmed using gel electrophoresis (Figure 20) and produced clean sequences. A BLAST analysis using

the nucleotide collection database confirmed that the sequences matched 100% identity the *P. persica* chlorophyll a/b-binding protein (*Lhcb2*) gene, promoter region and partial cds. There were few other matches, but the query cover was significantly lower, as was the identity and max scores. A search for other sequences of the *Lhcb2* gene in other *Prunus* species and peach cultivars was conducted and included in a phylogenetic analysis of samples 1, 2, and 3. Sequences from 'Lovell' and 'F8-1' peach cultivars, and *P. armeniaca* (apricot), *P. avium* (sweet cherry), *P. mume* (Chinese plum/ Japanese apricot) were collected and aligned to the samples and original BLAST match sequences (Figures 40 and 41). Due to the assumption that endogenous genes will remain unchanged within a species, we expected to see very few, if any mutations between the peach samples and reference sequence. Once aligned, we did not find any mutations between the reference gene, Samples 1, 2 and 3, or any of the peach cultivars. There was a significant difference in sequences between all peach samples and other *Prunus* species. A maximum likelihood tree was generated with the aligned sequences and all *P. persica* cultivars and samples grouped together on a single branch (Figure 23). *P. avium* and *P. armeniaca* grouped together on a single branch and *P. mume* was alone on a single branch at the bottom of the tree. Similar branching patterns was observed when a maximum parsimony tree was generated using the same alignment (Figure 22).

The *Lhcb2* endogenous gene was determined to be a good candidate for species identification of the peach. We were able to distinguish all of the peach sequences from other *Prunus* species, as well as accurately identifying the DNA samples as peach. One question that will be important to answer in future work is one of the *Lhcb2* gene and its presence in wild peach species. This gene has not been characterised in peach wild species and therefore sequences were not available for comparison. It would be useful to be able assess the presence of this gene in the wild species which may reveal an insight as to where the peach may have originated from.

5.7 Assessment of the Use of Highly Variable Chloroplast Markers for Peach Phylogeny

The *clpP*, *psbM-trnD*, and the *rbcL* chloroplast genes were used to evaluate their efficiency in distinguishing between *P. persica* and various wild peach species. Successful amplification of PCR

product was confirmed using gel electrophoresis (Figures 18 and 20) and produced clean sequences for all loci. A BLAST analysis was performed on the peach stone sequences in which the top matches for all loci were *P. persica*. The top *Prunus* species BLAST matches were collected and aligned to the Chinese peach stones. *P. kansuensis* was the only peach wild species that was able to be located using BLAST analysis and through GenBank. All other wild species are said to be uploaded to GenBank but could not be located for this analysis. A maximum parsimony and maximum likelihood tree was generated using alignments of peach stone samples and BLAST matches in order to determine whether these chloroplast markers are sufficient in resolving between different *Prunus* species. The *clpP* gene did not provide enough resolution between the different *Prunus* species to be sufficient as an identifying gene (Figures 24 and 25). *P. mume* and various *P. persica* sequences grouped sporadically with no visible pattern. Similar results were observed for the *psbM-trnD* intergenic spacer (Figures 26 and 27). A more diverse selection of *Prunus* species sequences were analysed based on BLAST matches but both phylogenetic trees did not provide enough differentiation between species. *P. persica* samples grouped together with different cultivars and multiple *Prunus* species. Phylogenetic analysis of the *rbcL* gene sequences and BLAST matches once again did not resolve species as effectively as expected (Figures 28 and 29). The maximum likelihood analysis separated into two distinct branches the first consisting *P. persica* cultivars and the Chinese stone samples, the second branch consisted of *P. persica* cultivars, *P. kansuensis*, and *P. mume*.

When the *rbcL* sequence data from the Chinese peach stones was aligned with sequence data collected from GenBank used previously for all peach wild species, the maximum parsimony tree did resolve some of the wild species but there was still no distinct branches that could be used for species identification (Figure 30). Similar results were found when all the published sequences, Chinese peach stone samples, and BLAST matches were aligned (Figure 31).

Based on these results, we once again see that variation within a species is great, and attempting to find one genetic marker to be able to distinguish between species is very difficult. It may be possible to create a phylogenetic tree using two or more markers but this is out of the scope of this thesis. The results have shown a close relationship between the two Chinese peach stones which supports the idea that

phytogenetic analysis is effective for assessing relationships between samples collected from a small geographical region. Due to the lack of mutations present between the two Chinese peach stones, we may also suspect that the stone collected from the Anjii Mountains may have been feral like the stone collected from beside the Maoshan site. It is possible that if we had been successful in extracting DNA from the stone purchased from a market in Shandong, it may have provided further support of this if it showed variation from the samples collected from the Zhejiang province.

5.8 Ancient DNA Case Study; Analysis of Peach Stones Collected from the Maoshan Site in China

Following the success of DNA extraction and amplification from modern peach samples, an attempt to extract DNA from an ancient peach seed from the Maoshan site in the Zhejiang province in China was made. For the initial extraction, the PK method was originally used on the peach seed followed by silica bead and p30 spin column purifications. This method was chosen based on the results from the modern sample experiments. Unfortunately following PCR analysis and gel electrophoresis no amplification was observed (Figure 20). Multiple attempts to amplify the *Lhcb2* gene were made including a dilution series and concentrating the samples but none were successful. As evident in Table 6, the PK extraction resulted in a DNA concentration of 156.47 ng/μL, but as stated earlier, there is no guarantee that this is all peach DNA. A low concentration of DNA was detected in the negative extraction of the peach stone using the PK method. This may be contributed to human contamination or another source due to the fact that the negative control did not amplify any peach DNA.

A study conducted by Moore (2011) proposed that the silica-spin column method was the most effective method for ancient DNA extraction from plant material. This method was applied to the Maoshan sample for both the stone and the seed. The use of a different source of DNA was in attempt to choose an area that may have less damage and have a higher success for amplification. In addition, the stone visually presented less pronounced damage than the seed which appeared burnt. Unfortunately, amplification was once again unsuccessful (Figure 18). DNA concentration was not detected Table 6) for both the peach stone extract and for the peach seed, and this may account for the lack of amplification.

The silica-spin column is very useful in that it uses a commercial kit which limits the risk of contamination. Although it is possible to run a sample through a spin column multiple times, when dealing with ancient DNA, there is still a high chance of product loss. As a result of the large volume of extraction buffer utilized for this method, it was necessary to run the sample through the column multiple times, and though this may have successfully removed contaminants, it may have in fact removed all peach DNA.

A final attempt was made to extract DNA from the ancient Maoshan peach, using the successful PK method on the ground peach stone powder. A relatively low quantity of DNA was extracted (17.41 ng/µL), this was much lower than the PK method on the peach seed, but higher than the silica-spin column method. Once again, amplification was unsuccessful.

In order to rule out inhibition, a spiked PCR was performed for all three ancient extractions. The *Lhcb2* gene was used for PCR analysis, where 2 µL of a modern peach sample was added to 5 µL of ancient material. In all reactions, amplification of the modern sample was successful and therefore inhibition was ruled out as a possibility (Figures 19 and 21).

As seen in Figure 10, the seed appeared to be almost burnt. The outside of the stone appeared untouched, but once opened, the seed was not whole as the modern samples, but like black ash. It is possible that all DNA had been destroyed. Two stones were collected from the site and both seeds were found in the same condition. Environmental conditions may have had an impact on the preservation of the seed. The peach stone did contain some cracks which may have allowed outside environmental influences to enter into the stone and cause deterioration to the seed. In order to assess some of the potential damage, GC-MS analysis was performed on the seed extracts to attempt to identify miscoding lesions which may give an insight as to whether there has been any hydrolytic or oxidative damage to the nucleotide bases.

The GC-MS analysis of modified bases within the ancient DNA of the archaeological peach seed indicates chemical modification in the form of methylation, oxidation, hydrolysis and reduction. In many cases the methylation is generated *in vivo* on cytosine nucleotides as seen with one of the two methylated modified bases 5-hydroxymethylcytosine in the sample. The other modified base exhibiting methylation

is N6-methyladenine, this can be generated through a range of methylating agents and can be generated in combination with hydrocarbon adduct in the process of burning. Ten of the modified bases identified are generated in combination with oxidation. Damage generated through oxidation is generally exposure to oxygen but it can be catalysed by the presence of transition metals in the soil, radiation and exposure to fire. Some of the modified bases produced through oxidation and identified in this research will generate blocking lesions, these include 4,6-diamino-5-formamidopyrimidine, 2,6-diamino-4-hydroxy-5-formanodopyridimine, guanidinohydantoin, 5-hydroxymethylcytosine, allantoin and 5-hydroxyhydantoin. The remaining modified bases generated through oxidation will produce miscoding lesions. The most frequent form of hydrolytic damage is the production of abasic sites. These are difficult to identify through GC-MS as the analysis focuses on detection of modified bases. However some modified bases produced in combination with hydrolysis were observed, specifically uracil which has been a focus of ancient DNA studies recently. Damage through hydrolysis would indicate exposure to water which in a maritime subtropical environment such as the Lower Yangzi basin where the samples were excavated would be plentiful. The recorded recovery of the archaeological sample was in close proximity to a small river which would also provide the appropriate environment for a high degree of water leading to hydrolysis. All of the damage to the bases produced through hydrolysis would lead to miscoding lesions. Reduction is rarely reported as a damaging form of ancient DNA due to inability to detect such damage. Using GC-MS for the detection of modified bases it is possible to identify modified bases that may have undergone reduction in their formation. Reduction is part of the redox reaction that accompanies oxidation and thus it would be expected in the oxidation of DNA, however DNA being reduced requires a reducing agent in the environment and these can be present with earth metals and simple acids like formic acids which could be found in the depositional environment. Methylation and some of the modified bases produced by oxidation will cause blocking lesions. These blocking lesions will prevent the ancient DNA from being amplified. Blocking lesions cause template inhibition or the inability for the DNA polymerase to amplify the DNA. These types of lesions do not form enzymatic inhibition of the DNA polymerase otherwise the spiked PCR results would be negative due to this type of inhibition.

Of the peach stones analysed in this study, only the archeological sample from the Maoshan site was unsuccessful for amplification. The late Neolithic peach stone is from the Liangzhu culture (3,300 BC-2, 300 BC) in which the economy was mainly agricultural (Zheng et al. 2014). During this period evidence of burning being used as an ecological management tool in sites in Eastern China have been found, and it is suggested that this was not an exclusive practice to this area. As seen in Figure 10, the state of the stone recovered from the Maoshan site suggests that seed may have been damaged through burning or water exposure. Charred peach stones have been reported in 3 different sites in Northern China, and it has been noted that peach stones preserve well after charring and generally are fire tolerant (Zheng et al. 2014). The state of the seed after burning has not been reported. The modified bases found through GC-MS analysis are consistent with the visual damage in Figure 10. The modified bases support the possibility that the damage to the seed may have been a result of water exposure or the possibility of damage from fire exposure, which would be consistent with the location the stone was recovered and environmental management methods that may have been in place at the time. The GC-MS results are also consistent with the inability to amplify DNA from the archaeological sample, which may have occurred due to water or fire exposure.

6.0 Conclusion

        Various DNA analysis methods were applied to modern and ancient peach samples collected from the Zhejiang Province of China. A PK method and a silica spin column method were used on both the seed and stone of an archeological sample collected from the Moashan site and successfully extracted DNA. An assessment of previously analysed genes was performed using the NCBI database and resulted in the identification of 11 loci sequenced in *P. persica* and its wild species. Maximum parsimony trees were generated for each locus, but the species did not separate as expected. It was determined that the variation within individual *P. persica* species is as great as the variation between species.

        Of the six microsatellite loci analysed, none produced sequences which matched *P.persica* and were not successful amplifying DNA in the ancient peach sample. Of the three highly variable chloroplast genes analysed, no amplification was successful for the ancient sample. Of the sequences generated from the modern samples, phylogenetic analysis did not reveal any species separation with other *Prunus* species. An endogenous reference gene, *Lhcb2*, a Type II chlorophyll *a/b* binding protein, was identified as an excellent gene for ancient DNA analysis of the peach. Although amplification of the ancient sample was not successful, phylogenetic analysis separated the modern *P.persica* samples from other *Prunus* species. Inhibition of the ancient peach samples was ruled out using a spiked PCR method.

        GC-MS analysis of the archeological peach seed revealed chemical damage in the form of methylation, oxidation, hydrolysis and reduction. The modified bases are consistent with the visual damage observed in the peach seed. The damage caused by oxidation and hydrolysis are consistent with the recovery of the stone from a small river, and documentation of burning for ecological management in other sites in China. The miscoding lesions were identified as blocking lesions which can be used to explain the unsuccessful amplification.

        In conclusion, in this study we have identified the *Lhcb2* gene as a useful gene for peach species identification for ancient DNA analysis. Although amplification from the peach stone collected from the Maoshan site in China was unsuccessful, we were able to successfully amplify DNA from modern peach stones from the Zhejiang Province. We have also successfully used GC-MS analysis to identify

miscoding lesions in archeological samples to assess damage which can be useful for overcoming future difficulties when analysing DNA from ancient material.

7.0 Future Considerations

In future studies, consideration must be given to more genetic research of *P. persica* and itswild speciesto allow for a better understanding of the domestication of the peach. Using the data available through the NCBI database, too much variation was observed within species and not all species investigated were represented in all genes used in this study. The *Lhcb2* gene was successful for identification of peach at a species level, but further investigation of this gene and the wild speciesof peach may reveal further insight into the domestication of the peach. It would be useful to compare sequences of this endogenous gene to the archeological record and the wild and domesticated peach.

Selection of samples must also be considered when performing DNA analysis on archaeological specimens. Success of amplification may increase if the samples collected were more protected to obvious exposure to water. In this study, the stones were collected from a small river, and this may have contributed to the damage caused by hydrolysis and inhibition of PCR. In addition, although evidence has shown that the peach demonstrates resilience to burning and the peach stone preserves well, this may not be the case for the seed as found in this study.

Application of GC-MS to archeological samples has proven beneficial when assessing damage to the DNA molecule. The use of GC-MS can be used to further understand the importance of damage when difficulties arise throughout the extraction and amplification of DNA. The identification of the cause of damage is also of importance for future use when analysing specimens from subtropical environments. Following the characterisation of miscoding lesions, application of repair techniques from the literature and the Matheson laboratory may be applied to future archeological samples to aid in recovery of viable DNA for amplification and sequence analysis.

8.0 References

Abian J (1999) The coupling of gas and liquid chromatography with mass spectrometry. Journal of mass spectrometry 34:157

Ahmad R, Parfitt DE, Fass J, Ogundiwin E, Dhingra A, Gradziel TM, Lin D, Joshi NA, Martinez-Garcia PJ, Crisosto CH (2011) Whole genome sequencing of peach (Prunus persica L.) for SNP identification and selection. BMC Genomics 12:569

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research 25:4692

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215:403

Aranzana MJ, Abbassi el K, Howad W, Arus P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11:69

Bailey JS, French AP (1949) The inheritance of certain fruit and foliage characters in the peach. University of Massachusetts

Bakht J, Jamshed A, Shafi M (2013) Genetic diversity and phylogenetic relationship among different peach genotypes through rapd markers. Pak. J. Bot 45:1241

Baldauf SL (2003) Phylogeny for the faint of heart: a tutorial. TRENDS in Genetics 19:345

Bassett C, Callahan A, Dunn L (1998) Characterization of a Type II Chlorophyll a/b-binding Protein Gene (Lhcb2* Pp1) in Peach: I. Isolation, Identification, and Abundance in Developing Leaves of MatureLoring'Trees in the Absence of Flowering. Journal of the American Society for Horticultural Science 123:486

Beckman KB, Ames BN (1997) Oxidative decay of DNA. Journal of Biological Chemistry 272:19633

Bell JR (2008) A Simple way to Treat PCR Products Prior to sequencing using Exo-SAP-IT. BioTechniques 44:1

Boom R, Sol C, Salimans M, Jansen C, Wertheim-van Dillen P, Van der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. Journal of clinical microbiology 28:495

Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and Mitochondrial DNA: Applications, Debates, and Foundations*. Annual review of genomics and human genetics 4:119

Burger G, Gray MW, Franz Lang B (2003) Mitochondrial genomes: anything goes. Trends in genetics 19:709

Byrne DH, Raseira MB, Bassi D, Piagnani MC, Gasic K, Reighard GL, Moreno MA, Pérez S (2012) Peach.505

Chaparro J, Werner D, O'malley D, Sederoff R (1994) Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theoretical and Applied Genetics 87:805

Chargaff E (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Cellular and Molecular Life Sciences 6:201

Chien A, Edgar DB, Trela JM (1976) Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. Journal of Bacteriology 127:1550

Chunwongse J, Martin G, Tanksley S (1993) Pre-germination genotypic screening using PCR amplification of half-seeds. Theoretical and Applied Genetics 86:694

Cipriani G, Lot G, Huang W-G, Marrazzo M, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [Prunus persica (L) Batsch]: isolation, characterisation and cross-species amplification in Prunus. Theoretical and Applied Genetics 99:65

Cohen IM (2011) Testing the utility of the consortium for the barcoding of life's two'agreed upon'plant DNA barcodes, matK and rbcL.

Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. Science 289:1139

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

Dirlewanger E, Bodo C (1994) Molecular genetic mapping of peach. Euphytica 77:101

Dizdaroglu M (1984) The use of capillary gas chromatography—mass spectrometry for identification of radiation-induced DNA base damage and DNA base—amino acid cross-links. Journal of Chromatography A 295:103

Dizdaroglu M (1985) Application of capillary gas chromatography-mass spectrometry to chemical characterization of radiation-induced base damage of DNA: implications for assessing DNA repair processes. Analytical biochemistry 144:593

Dizdaroglu M (1991) Chemical determination of free radical-induced damage to DNA. Free Radical Biology and Medicine 10:225

Dizdaroglu M (1994) Chemical determination of oxidative DNA damage by gas chromatography-mass spectrometry. Methods in Enzymology 234:3

Dizdaroglu M, Jaruga P, Birincioglu M, Rodriguez H (2002) Free radical-induced damage to DNA: mechanisms and measurement. Free Radic Biol Med 32:1102

Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7:e35071

Doyle J (1991) DNA protocols for plants Molecular techniques in taxonomy. Springer, pp. 283-293

Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. Taxon:715

Ebeling W, Hennrich N, Klockow M, Metz H, Orth HD, Lang H (1974) Proteinase K from Tritirachium album limber. European Journal of Biochemistry 47:91

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of molecular evolution 17:368

Franklin RE, Gosling RG (1953a) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. Nature 172:156

Franklin RE, Gosling RG (1953b) Molecular configuration in sodium thymonucleate. Nature 171:740

Fresnedo-Ramírez J, Martínez-García PJ, Parfitt DE, Crisosto CH, Gradziel TM (2013) Heterogeneity in the entire genome for three genotypes of peach [Prunus persica (L.) Batsch] as distinguished from sequence analysis of genomic variants. BMC genomics 14:750

Friedberg EC, Aguilera A, Gellert M, Hanawalt PC, Hays JB, Lehmann AR, Lindahl T, Lowndes N, Sarasin A, Wood RD (2006) DNA repair: from molecular mechanism to human disease. DNA Repair (Amst) 5:986

Georgi L, Wang Y, Yvergniaux D, Ormsbee T, Inigo M, Reighard G, Abbott G (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach [Prunus persica (L.) Batsch]. Theor Appl Genet 105:1151

Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. Trends in Ecology & Evolution 20:541

Guillemaut P, Maréchal-Drouard L (1992) Isolation of plant DNA: A fast, inexpensive, and reliable method. Plant Molecular Biology Reporter 10:60

Halliwell B, Dizdaroglu M (1992) The measurement of oxidative damage to DNA by HPLC and GC/MS techniques. Free Radic Res Commun 16:75

Handt O, Höss M, Krings M, Pääbo S (1994) Ancient DNA: methodological challenges. Experientia 50:524

Hansen JN (1974) Isolation of higher molecular weight DNA from Bacillus Cereus T using proteinase K. Preparative biochemistry 4:473

Hedrick UP, Howe GH, Taylor OM, Tubergen CB (1917) The peaches of New York, by U.P. Hedrick, assisted by G.H. Howe, O.M. Taylor [and] C.B. Turbergen. J. B. Lyon Company, printers, Albany

Hesse C (1975) Peaches. Advances in fruit breeding. Purdue University Press. West Lafayette, Indiana:285

Hilz H, Wiegers U, Adamietz P (1975) Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of 'masked' proteins. European Journal of Biochemistry 56:103

Hofreiter M, Jaenicke V, Serre D, Haeseler Av A, Pääbo S (2001a) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research 29:4793

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001b) Ancient DNA. Nature Reviews Genetics 2:353

Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic acids research 24:1304

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Research 24:1304

Höss M, Pääbo S (1993) DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Research 21:3913

International Peach Genome I, Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel LA, Decroocq V, Sosinski B, Prochnik S, Mitros T, Policriti A, Cipriani G, Dondini L, Ficklin S, Goodstein DM, Xuan P, Del Fabbro C, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B, Bielenberg D, Pirona R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arus P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45:487

Jenner A, England T, Aruoma O, Halliwell B (1998) Measurement of oxidative DNA damage by gas chromatography–mass spectrometry: ethanethiol prevents artifactual generation of oxidized DNA bases. Biochem. J 331:365

Jung S, Ficklin SP, Lee T, Cheng C-H, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S (2014) The Genome Database for Rosaceae (GDR): year 10 update. Nucleic acids research 42:D1237

Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbott A, Tomkins J, Main D (2004) GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BMC bioinformatics 5:130

Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic acids research 36:D1034

Kang HW, Cho YG, Yoon UH, Eun MY (1998) A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. Plant Molecular Biology Reporter 16:90

Kim KJ, Jansen RK (1995) ndhF sequence evolution and the major clades in the sunflower family. Proceedings of the National Academy of Sciences 92:10379

Kresge N, Simoni RD, Hill RL (2005) Chargaff's Rules: the Work of Erwin Chargaff. Journal of Biological Chemistry 280:e21

Lamers R, Hayter S, Matheson CD (2009) Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. Journal of molecular evolution 68:40

Layne DR, Bassi D (2008) The peach: botany, production and uses. CABI

Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709

Lodhi MA, Ye G-N, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. Plant Molecular Biology Reporter 12:6

Melzak KA, Sherwood CS, Turner RF, Haynes CA (1996) Driving forces for DNA adsorption to silica in perchlorate solutions. Journal of colloid and interface science 181:635

Millar CD, Lambert DM (2013) Ancient DNA: towards a million-year-old genome. Nature 499:34

Möller E, Bahnweg G, Sandermann H, Geiger H (1992) A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic Acids Research 20:6115

Monet R (1989) Peach genetics: past, present and future. Acta Hort 254:49

Monet R, Guye A, Roy M, Dachary N (1996) Peach Mendelian genetics: a short review and new results. Agronomie 16:321

Moore JN (2011) Comparative study of ancient DNA extraction methods for archaeological plant remains. Environment: Department of Archaeology

Moret BM, Roshan U, Warnow T (2002) Sequence-length requirements for phylogenetic methods Algorithms in Bioinformatics. Springer, pp. 343-356

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1992) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Biotechnology Series:17

Murray M, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic acids research 8:4321

O'Rourke DH, Hayes MG, Carlyle SW (2000) Ancient DNA Studies in Physical Anthropology. Annual Review of Anthropology 29:217

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. American journal of botany:1205

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PLF, Fumagalli M, Vilstrup JT, Raghavan M,

Korneliussen T, Malaspinas A-S, Vogt J, Szklarczyk D, Kelstrup CD, Vinther J, Dolocan A, Stenderup J, Velazquez AMV, Cahill J, Rasmussen M, Wang X, Min J, Zazula GD, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson JF, Weinstock J, Gregersen K, Roed KH, Eisenmann V, Rubin CJ, Miller DC, Antczak DF, Bertelsen MF, Brunak S, Al-Rasheid KAS, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert MTP, Kjaer K, Sicheritz-Ponten T, Jensen LJ, Olsen JV, Hofreiter M, Nielsen R, Shapiro B, Wang J, Willerslev E (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499:74

Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences 86:1939

Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38:645

Palmer JD (1985) Chloroplast DNA and molecular phylogeny. Bioessays 2:263

Pauling L, Corey RB (1953) A Proposed Structure For The Nucleic Acids. Proc Natl Acad Sci U S A 39:84

Poinar GJ (1999) Ancient DNA. American Scientist 87:446

Poinar HN, Stankiewicz BA (1999) Protein preservation and DNA retrieval from ancient tissues. Proceedings of the National Academy of Sciences of the United States of America 96:8426

Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant molecular biology reporter 15:8

Price CW, Leslie DC, Landers JP (2009) Nucleic acid extraction techniques and application to the microchip. Lab on a Chip 9:2484

Quan X, Zhou SL (2011) Molecular identification of species in Prunus sect. Persica (Rosaceae), with emphasis on evaluation of candidate barcodes for plants. Journal of Systematics and Evolution 49:138

Quarta R, Dettori M, Sartori A, Verde I (1998) Genetic linkage map and QTL analysis in peach XXV International Horticultural Congress, Part 11: Application of Biotechnology and Molecular Biology and Breeding-Gene 521: 233-242

Rajapakse S, Belthoff L, He G, Estager A, Scorza R, Verde I, Ballard R, Baird W, Callahan A, Monet R (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. Theoretical and Applied Genetics 90:503

Richter C, Park JW, Ames BN (1988) Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proceedings of the National Academy of Sciences of the United States of America 85:6465

Roberts C, Ingham S (2008) Using ancient DNA analysis in palaeopathology: a critical analysis of published papers, with recommendations for future work. International Journal of Osteoarchaeology 18:600

Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant molecular biology 5:69

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74:5463

Scorza R, Melnicenco L, Dang P, Abbott AG (2002) Testing a microsatellite marker for selection of columnar growth habit in peach [Prunus persica (L.) Batsch]. Acta horticulturae

Shang Y, Zhu P, Huang K, Liu W, Tian W, Luo Y, Xu W (2014) A peach (Prunus persica)-specific gene, Lhcb2, used as an endogenous reference gene for qualitative and real-time quantitative PCR to detect fruit products. LWT - Food Science and Technology 55:218

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American journal of botany 94:275

Shaw J, Small RL (2004) Addressing the "hardest puzzle in American pomology:" Phylogeny of Prunus sect. Prunocerasus (Rosaceae) based on seven noncoding chloroplast DNA regions. American Journal of Botany 91:985

Sosinski B, Gannavarapu M, Hager L, Beck L, King GJ, Ryder C, Rajapakse S, Baird W, Ballard R, Abbott A (2000) Characterization of microsatellite markers in peach [Prunus persica (L.) Batsch]. Theoretical and Applied Genetics 101:421

Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S (2000) Microsatellite DNA in peach (Prunus persica L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. Genome 43:512

Tong Z, Gao Z, Wang F, Zhou J, Zhang Z (2009) Selection of reliable reference genes for gene expression studies in peach using real-time PCR. BMC Mol Biol 10:71

Varma A, Padh H, Shrivastava N (2007) Plant genomic DNA isolation: an art or a science. Biotechnol J 2:386

Wang Y, Georgi L, Reighard G, Scorza R, Abbott A (2002) Genetic mapping of the evergrowing gene in peach [Prunus persica (L.) Batsch]. Journal of Heredity 93:352

Watson JD, Crick FH (1953a) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737

Watson JD, Crick FHC (1953b) The Structure of DNA. Cold Spring Harbor Symposia on Quantitative Biology 18:123

Yamazaki K, Okabe M, Takahashi E (1987) Inheritance of some characteristics and breeding of new hybrids in flowering peaches. Bulletin of the Kanagawa Horticultural Experiment Station (Japan)

Yoon J, Liu D, Song W, Liu W, Zhang A, Li S (2006) Genetic diversity and ecogeographical phylogenetic relationships among peach and nectarine cultivars based on simple sequence repeat (SSR) markers. Journal of the American Society for Horticultural Science 131:513

Zheng Y, Crawford GW, Chen X (2014) Archaeological Evidence for Peach (Prunus persica) Cultivation and Domestication in China. PloS one 9:e106595

9.0 Appendix

Figure 33. Phylogenetic analysis of *ndhF* gene sequences
Maximum parsimony tree of various *P. persica* and wild species based on the NADH dehydrogenase
subunit F gene (*ndhF*) . A bootstrap analysis of 1000 replicates was performed.
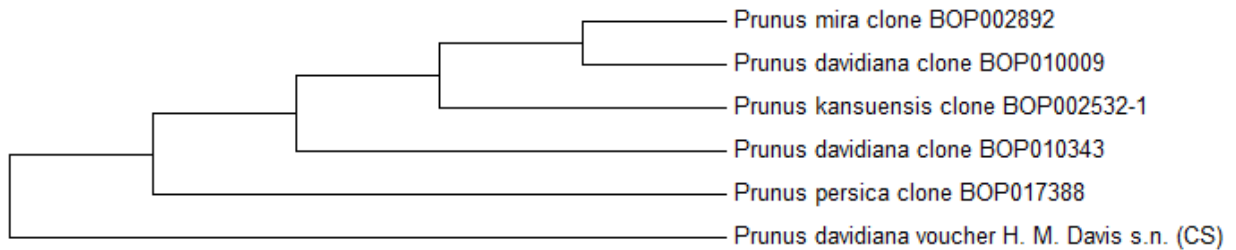
Figure 34. Phylogenetic analysis of published *ndhF* gene sequences
Maximum parsimony tree of various published *P. persica* and wild species based on the NADH
dehydrogenase subunit F gene (*ndhF*) . A bootstrap analysis of 1000 replicates was performed.
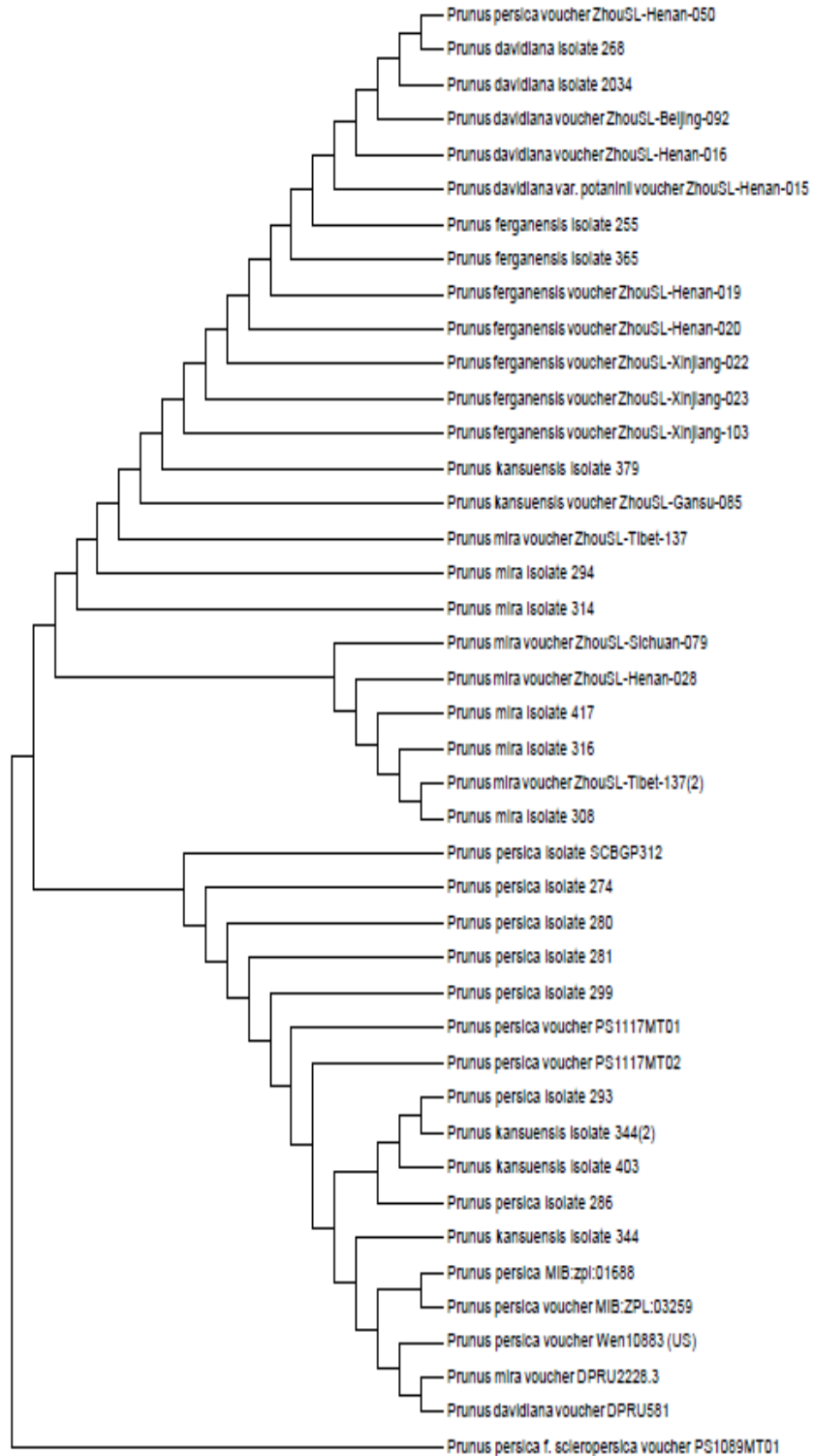
Figure 35. Phylogenetic analysis of *trnH-psbA* intergenic spacer sequences
Maximum parsimony tree of various *P. persica* and wild species based on trnH-psbA intergenic spacer
(*trnH-psbA*) . A bootstrap analysis of 1000 replicates was performed.
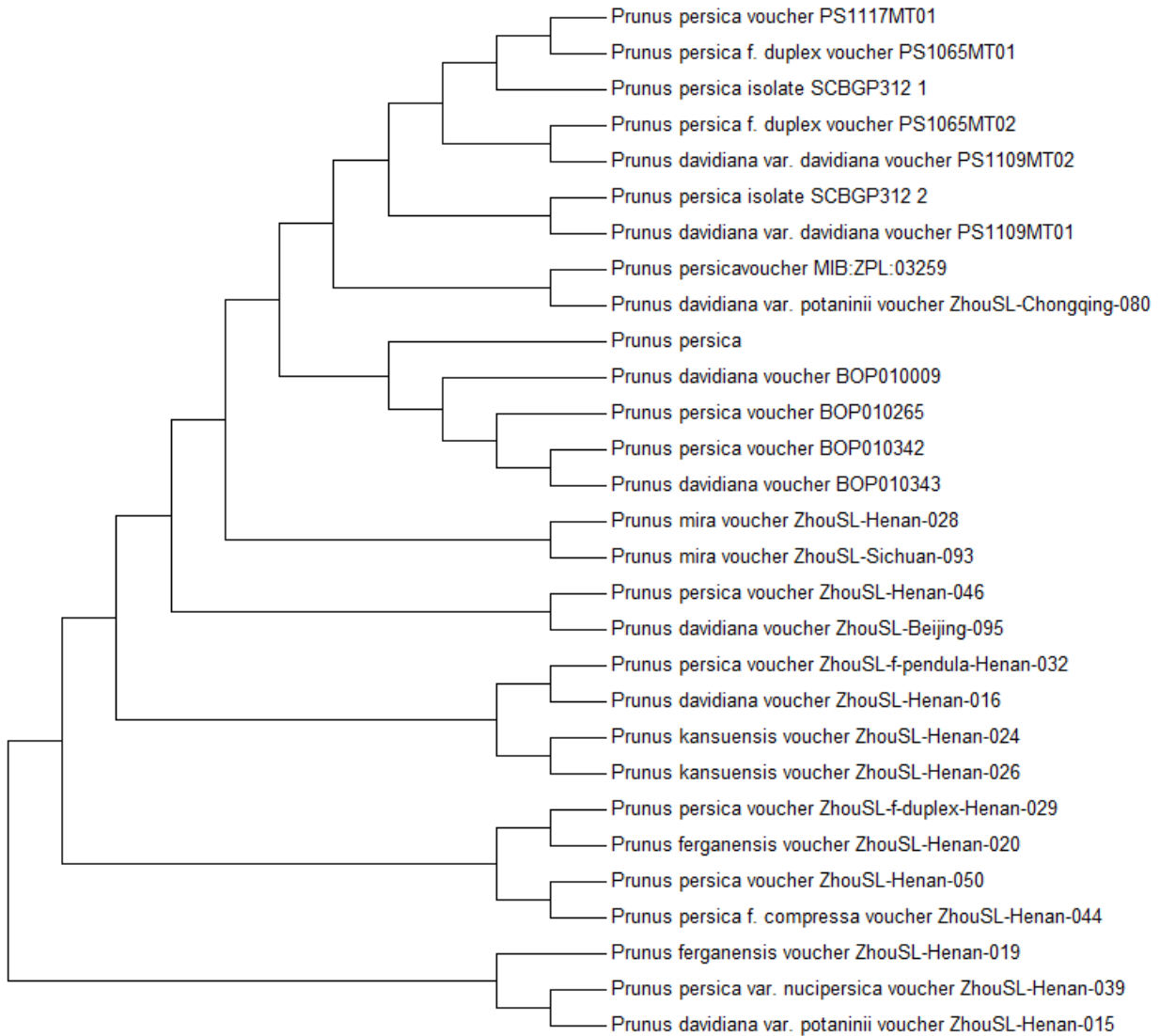
Figure 36. Phylogenetic analysis of published *trnH-psbA* intergenic spacer sequences
Maximum parsimony tree of various published *P. persica* and wild species based on trnH-psbA
intergenic spacer (*trnH-psbA*) . A bootstrap analysis of 1000 replicates was performed.
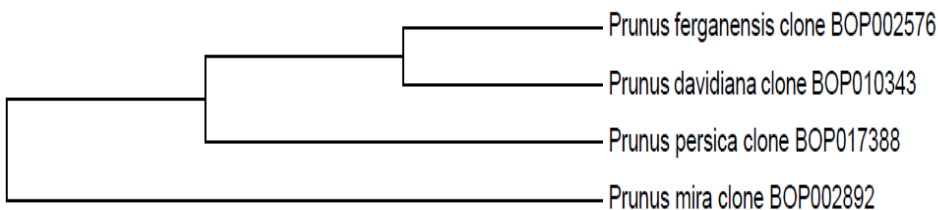
Figure 37. Phylogenetic analysis of *SbeI* gene sequences
Maximum parsimony tree of various *P. persica* and wild species based on Starch branching enzyme gene
(*SbeI*). A bootstrap analysis of 1000 replicates was performed.

98
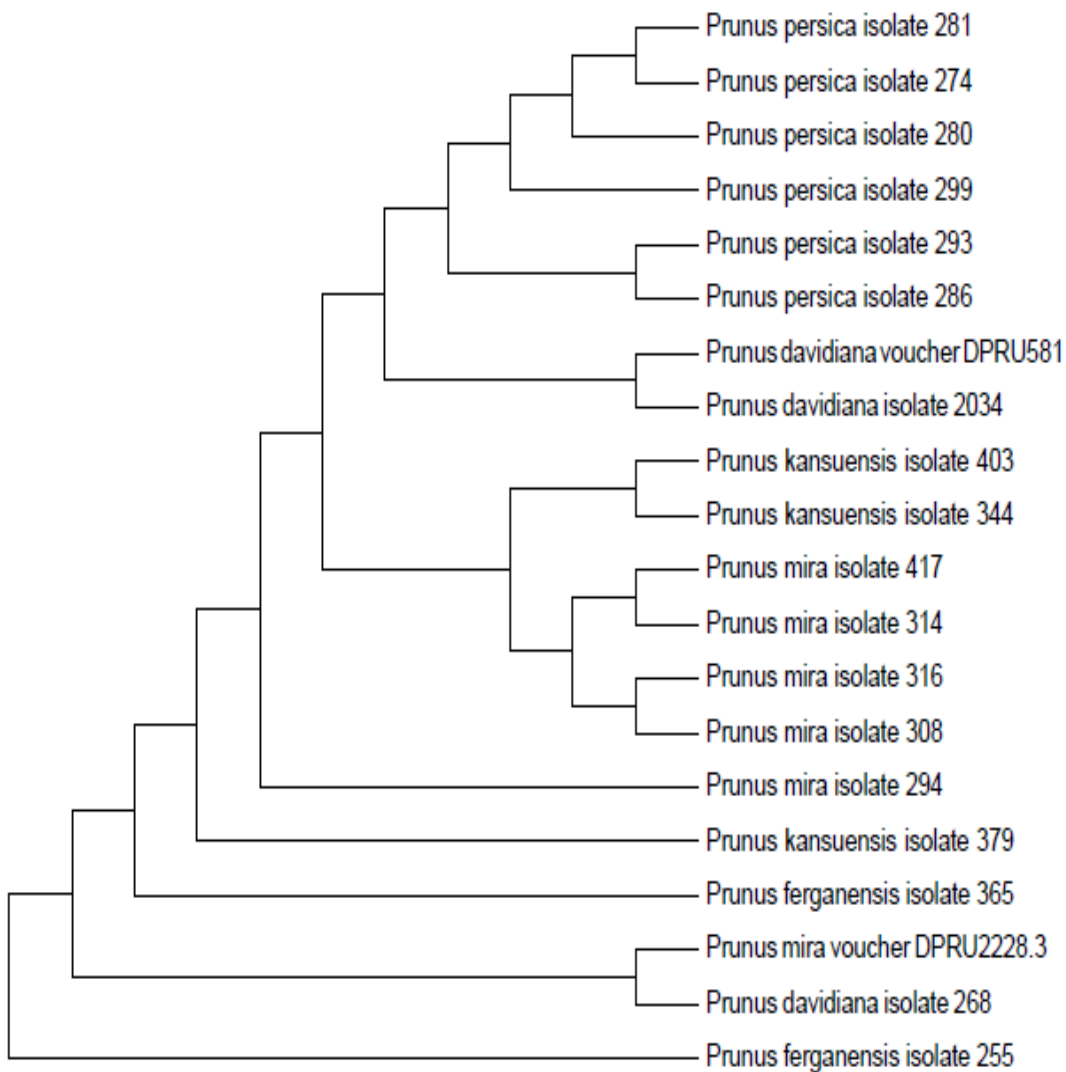
Figure 38. Phylogenetic analysis of *trnG-trnS* intergenic spacer sequences
Maximum parsimony tree of various *P. persica* and wild species based on *trnG-trnS* intergenic spacer (*trnG-trnS*) . A bootstrap analysis of 1000 replicates was performed.
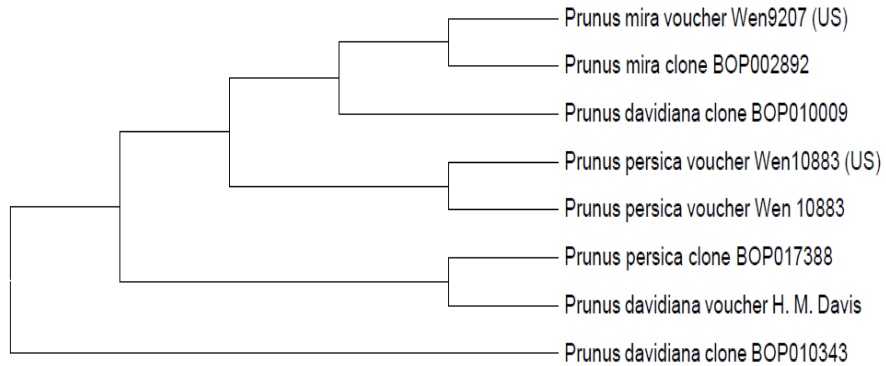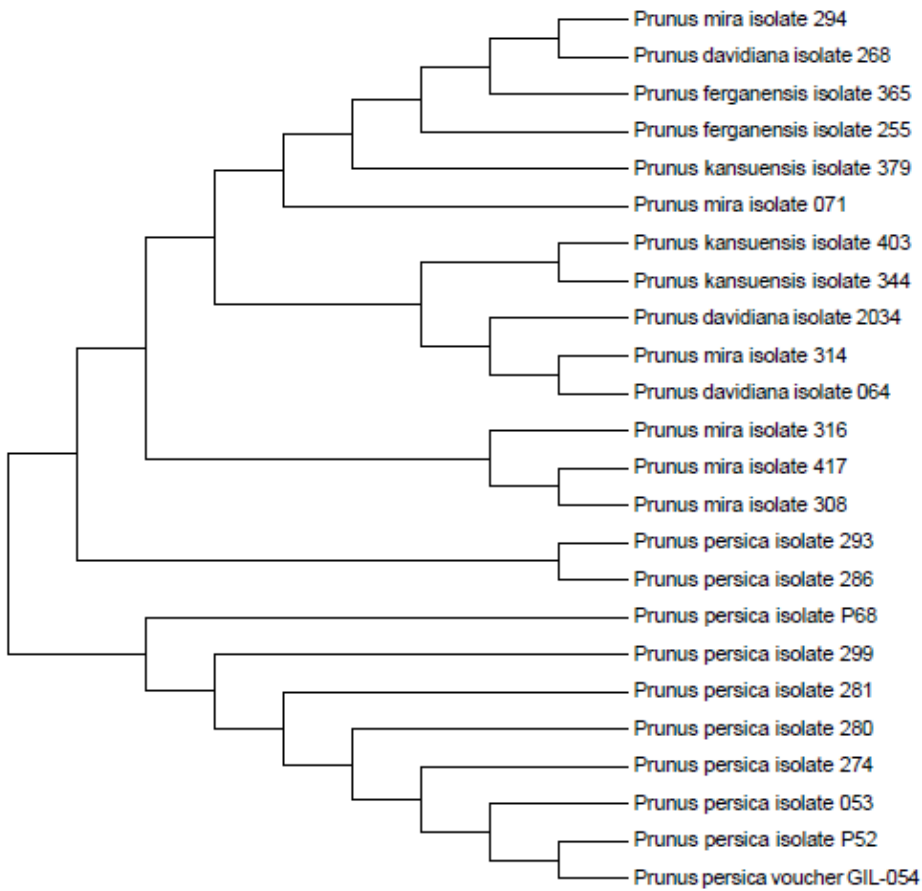
Figure 39. Phylogenetic analysis of *trnL* gene sequences
Maximum parsimony tree of various *P. persica* and wild species based on tRNA-Leu gene (*trnL*). A
bootstrap analysis of 1000 replicates was performed.

```
Prunus persica Lhcb2 gene      AAATTAAGGATTAGTTACATACTTCTAATCTCGACTGTTGGATTGCATCTAAGAGAATACGAACACATTAAATGAATGCAAGCCAACAGC
Modern Peach                   ------------------------------------------------------------------------------------------
Prunus persica (Feral Zhejiang ------------------------------------------------------------------------------------------
Prunus persica (Mountain Zheji ------------------------------------------------------------------------------------------
Prunus persica Dr. Davis       ------------------------------------------------------------------------------------------
Prunus persica Lovell          ------------------------------------------------------------------------------------------
Prunus persica F8-1            ------------------------------------------------------------------------------------------

                                       100       110       120       130       140       150       160       170       180
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene      TAAGATTAGAAGTATGTAACCAATCCTTAAAATACTTTTTGGAGAATCTCCCTTTCTGCCCTGCTATGGTTTGCTGTTGTTTTGTCAATA
Modern Peach                   ----------------------------------------------------------------------------------------..
Prunus persica (Feral Zhejiang ---------------------------------------------------------------------------------------C...
Prunus persica (Mountain Zheji ----------------------------------------------------------------------------------------..
Prunus persica Dr. Davis       -----------------------------------------------------------------------------------------..
Prunus persica Lovell          -----------------------------------------------------------------------------------------...
Prunus persica F8-1            ------------------------------------------------------------------------------------------..

                                       190       200       210       220       230       240       250       260       270
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene      CTGCGGCAAATTAACCTTTTCACAGCAGCCAAGGAACAGTTCTAATTTTAGAGGTTGGCGTTTTTTACCTCAAGCCTACAAAAACACCCA
Modern Peach                   ..........................................................................................
Prunus persica (Feral Zhejiang ..........................................................................................
Prunus persica (Mountain Zheji ..........................................................................................
Prunus persica Dr. Davis       ..........................................................................................
Prunus persica Lovell          ..........................................................................................
Prunus persica F8-1            ..........................................................................................

                                       280       290       300       310       320       330       340       350       360
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene      AGGAAAGTTTATGCTTTCATGATTTCATCAATGGGGGCCTATTAAATATTTGTCTTCTGGTTGGAGCTGGCATGAACTTCTACAGTACAA
Modern Peach                   ..........................................................................................
Prunus persica (Feral Zhejiang ..........................................................................................
Prunus persica (Mountain Zheji ..........................................................................................
Prunus persica Dr. Davis       ..........................................................................................
Prunus persica Lovell          ..........................................................................................
Prunus persica F8-1            ..........................................................................................

                                       370       380       390       400       410       420       430       440       450
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene      CTAATCTTGTGGCCTGCGTTCACCAACCACAAAAATTGCACCAATTTGCCAGCCACGTGGCTAAGAGATAACCTCTCTCTCACTAGAACA
Modern Peach                   ..........................................................................................
Prunus persica (Feral Zhejiang ..........................................................................................
Prunus persica (Mountain Zheji ..........................................................................................
Prunus persica Dr. Davis       ..........................................................................................
Prunus persica Lovell          ..........................................................................................
Prunus persica F8-1            ..........................................................................................

                                       460       470       480       490       500       510       520       530       540
                               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene      ATCAGTTTGGAAATCTAAATCCAATTGAGAACCAATAGAATGTTGCTGATGTGTAGATATCCTAATCCACAACATCTTCACATTCCCGTT
Modern Peach                   ..........................................................................................
Prunus persica (Feral Zhejiang ..........................................................................................
Prunus persica (Mountain Zheji ..........................................................................................
```

101

```
Prunus persica Dr. Davis          ...................................................................................
Prunus persica Lovell             ...................................................................................
Prunus persica F8-1               ...................................................................................

                                          550       560       570       580       590       600       610       620       630
                                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene         ATAACATCCCAGCGCTCTAGTTGTCCAAGGCATCTCATTAGGCCTCACAACACCACCCCACTTCGAACAGGAGAGCATAGTCATGGCAAC
Modern Peach                      ...........................................................................------------------
Prunus persica (Feral Zhejiang    .........................................................................--------------------
Prunus persica (Mountain Zheji    ......................................................................-----------
Prunus persica Dr. Davis          ........................................................................----------------
Prunus persica Lovell             .........................................................................------------------
Prunus persica F8-1               ........................................................................----------------
```

Figure 40. Alignment of *Lhcb2* gene of various *P. persica* cultivars and of DNA extracts from the seed of peach samples collected from the Zhejiang province in China.

102

```
                          1000      1010      1020      1030      1040      1050      1060      1070      1080
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene ------AAATTAAGGATTAGTTACATACTTCTAATCTCGACTGTTGGATTGCATCTAAGAGAATACGAACACATTAAATGAATGCAAG
Modern Peach              ------------------------------------------------------------------------------------------
Prunus persica (Feral Zhejiang ----------------------------------------------------------------------------------------
Prunus persica (Mountain Zheji ----------------------------------------------------------------------------------------
Prunus persica Dr. Davis ------------------------------------------------------------------------------------------
Prunus persica Lovell    ------------------------------------------------------------------------------------------
Prunus persica F8-1      ------------------------------------------------------------------------------------------
Prunus armeniaca         CCCCCCAA.GCA..T..TA.G.CCCAGAC.GC.C..AA.ACTTG.GACC...CTC.GAGCA.ACTCCATC.T..T.G.CTG.TGAATTCC
Prunus avium             CCCCCCAA.GCA..T..TA.G.CCCAGAC.GC.C..AA.ACTTG.GACC...CTC.GAGCA.ACTCCATC.T..T.G.CTG.TGAATTCC
Prunus mume              ------------------------------------------------------------------------------------------

                          1090      1100      1110      1120      1130      1140      1150      1160      1170
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene CCAACAGCTAAGATTAGAAGTATGTAACCAATCCTTAAAATACTTTTTGGAGAATCTCCCTTTCTGCCCTGCTATGGTTTGCTGTTGTTT
Modern Peach              ------------------------------------------------------------------------------------------
Prunus persica (Feral Zhejiang ----------------------------------------------------------------------------------------
Prunus persica (Mountain Zheji ----------------------------------------------------------------------------------------
Prunus persica Dr. Davis ------------------------------------------------------------------------------------------
Prunus persica Lovell    ------------------------------------------------------------------------------------------
Prunus persica F8-1      ------------------------------------------------------------------------------------------
Prunus armeniaca         .TGGTGA...C.GA.G.G.CAC..CTGGACTAT.CGC.G.CC..GAGACATTTGC.AAAAACCG..AG..TGAGGT.A.CCACAGCCGA.
Prunus avium             .TGGTGA...C.GA.G.G.CAC..CTGGACTAT.CGC.G.CC..GAGACATTTGC.AAAAACCG..AG..TGAGGT.A.CCACAGCCGA.
Prunus mume              ------------------------------------------------------------------------------------------

                          1180      1190      1200      1210      1220      1230      1240      1250      1260
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene TGTCAATACTGCG-GCA----AATTAACCTTTTCACAG---CAGCCAAGGAACAGTTCTAATTTTAG--AG-GTTGGCGTTTTTTACCTC
Modern Peach             ------........------....T.---............................N--..--..............
Prunus persica (Feral Zhejiang ----C........-....----....T.---.............................--.--..............
Prunus persica (Mountain Zheji ------........-....----....T.---.............................--.--..............
Prunus persica Dr. Davis ------........-....----........---.............................--.--..............
Prunus persica Lovell    ------........-....----........---.............................--.--..............
Prunus persica F8-1      ------........-....----........---.............................--.--..............
Prunus armeniaca         G.G.C..G..TG.T...CTGGG..GTGT...CC..G.AATC.T.T...A...TG..GTC..G..CG.CG..-.C..TTTGG..CA.GGCT
Prunus avium             G.G.C..G..TG.T...CTGGG..GTGT...CC..G.AATC.T.T...A...TG..GTC..G..CG.CG..-.C..TTTGG..CA.GGCT
Prunus mume              -------.C...TTC.A.GTGGGG.G-GTG..G.G.GGC------.T..T.GAT.CCT.GGACAAG.AG..C.C...GA.G...A..A.GG

                          1270      1280      1290      1300      1310      1320      1330      1340      1350
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene AAGCCTACAAAAACACCCAAGGAAAGTTTATGCTTTCATG-------ATTTCATCAATG---GGGGCCTATTAAATATTTG--TCTTC
Modern Peach             .......................................--------............--.............--....
Prunus persica (Feral Zhejiang .......................................--------............--.............--....
Prunus persica (Mountain Zheji .......................................--------............--.............--....
Prunus persica Dr. Davis .......................................--------............--.............--....
Prunus persica Lovell    .......................................--------............--.............--....
Prunus persica F8-1      .......................................--------............--.............--....
Prunus armeniaca         GGAT..CA..TCTTCT..G...GTG.CC.TGA..A..T..GCAACCCAA.CC.TG..C...CCCA.A..A.T..GGCA..C..GGCTG..
Prunus avium             GGAT..CA..TCTTCT..G...GTG.CC.TGA..A..T..GCAACCCAA.CC.TG..C...CCCA.A..A.C..GGCA..C..GGCTG..
Prunus mume              G.ATG.GA.G.TGTTGTGG.TT.GGA.A.C.A.AGATCA.-----CAAC...CT..TGG.TCTCAATTGGAT...G..T.CCAAACTGAT

                          1360      1370      1380      1390      1400      1410      1420      1430      1440
```
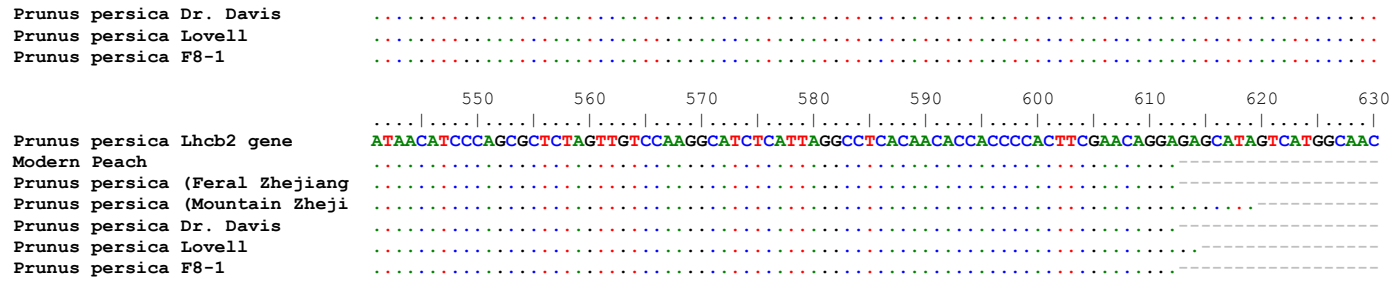
```
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene   TGGT-TGGAGCTGGCATGAACTTCTACAGTACAACTAATC--TTGTGGCCTGCGT--TCACCAACCACAAAAATTGCACC------AAT
Modern Peach                ....-................................................--.......................--......-------...
Prunus persica (Feral Zhejiang  ....-................................................--.......................--......-------..
Prunus persica (Mountain Zheji  ....-................................................--.......................--......-------..
Prunus persica Dr. Davis    ....-................................................--.......................--......-------...
Prunus persica Lovell       ....-................................................--.......................--......-------...
Prunus persica F8-1         ....-................................................--.......................--......-------...
Prunus armeniaca            CA.G-.T.T...CATGG..TT.GTCGA..G.T.CAGG..TGG.G.A..A.CA.T.GG.G.AGG..TGG.TCC.C.TT...CTGGAGGGGC
Prunus avium                CA.G-.T.T...AATGG..TT.GTCGA..G.T.CAGG..TGG.G.A..A.CA.T.GG.G.AGG..TGG.TCC.C.TT...CTGGAGGGGC
Prunus mume                 ..T.C.A.T.AGA.AGG.GTTA...CTTAGC..CG.GGCA--G.CAAAT..GTG---..AT.TTTGTGGTTGG..A..GC------.GG

                               1450      1460      1470      1480      1490      1500      1510      1520      1530
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene   TTGCCAGCCACGTGG-CTAAGAGATAACCTCT--CTCTCACTAGAACAATCAGTTTGGAAATCTAAATC----CAATTGAGAACCAATA
Modern Peach                ...............-...........-......--.......................................----...............
Prunus persica (Feral Zhejiang  ...............-...........-......--.......................................----...............
Prunus persica (Mountain Zheji  ...............-...........-......--.......................................----...............
Prunus persica Dr. Davis    ...............-...........-......--.......................................----...............
Prunus persica Lovell       ...............-...........-......--.......................................----...............
Prunus persica F8-1         ...............-...........-......--.......................................----...............
Prunus armeniaca            C.TTG.C....T...A..GGCT...G.T.CTGAGG.GT.TG..-....TGAAG..GAAAG.GCT...GAATGGG.GGC..GC..TG.C.T
Prunus avium                C.TTG.C....T...A..GGCT...G.T.CTGAGG.GT.TG..-....TGAAG..GAAAG.GCT...GAATGGG.GGC..GC..TG.C.T
Prunus mume                 CCA.A..ATTA..T.TACTGT...AGGT.ATG-----.CAG..CC...C.AA..-ACAA.T..T...TAGGCCC.CG....TG.AACCAT

                               1540      1550      1560      1570      1580      1590      1600      1610      1620
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene   GAATGTT--GCTGATGTGT---AGATATCCTAATCCAC--AACATCTTCAC--ATTCCCGTTATAAC-ATCCCAGCGCTCTAGTTGTC--
Modern Peach                .......--.........---.............---.--.............--................-..................--
Prunus persica (Feral Zhejiang  .......--.........---.............---.--.............--................-..................--
Prunus persica (Mountain Zheji  .......--.........---.............---.--.............--................-..................--
Prunus persica Dr. Davis    .......--.........---.............---.--.............--................-..................--
Prunus persica Lovell       .......--.........---.............---.--.............--................-..................--
Prunus persica F8-1         .......--.........---.............---.--.............--................-..................--
Prunus armeniaca            C......TG.A.TC.T...TCAG.C...TG...CTGGA--..GGG.CCAGTTG.GAA..TC...G..C..GTTGCT.A..C.....C.AA
Prunus avium                C......TG.A.TC.T...TCAG.C...TG...CTGGA--..GGG.CCAGTTG.GAA..TC...G..C..GTTGCT.A..C.....C.AA
Prunus mume                 ...A.CATAAACTT.CCT.---G.G.G.TT.TG.TGG.TT..GG.AAAA.ACGCCAA..TC..A..T--..GA.CT.T..C--...G.TG

                               1630      1640      1650      1660      1670      1680      1690      1700      1710
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Prunus persica Lhcb2 gene   CAAGGCATCTCATTAGGCCTCACAACACCACCCCACTTCGAACAGGAGAGCATAGTCATGGCAACCTCTGCAATCCAACAATCAGCATTT
Modern Peach                .......................................................-----------------------------------------
Prunus persica (Feral Zhejiang  .......................................................-----------------------------------------
Prunus persica (Mountain Zheji  .....................................................-----------------------------------------
Prunus persica Dr. Davis    .......................................................-----------------------------------------
Prunus persica Lovell       .......................................................-----------------------------------------
Prunus persica F8-1         .......................................................-----------------------------------------
Prunus armeniaca            ...T....GGGC...T...A.-....TTTGT...TGGAAA.TG.A..A.C.TA.A.GCAT.GCTGT..AT.TG.AAC.TT.C.C.A.C..
Prunus avium                ...T....GGGC...T...A.-....TTTGT...TGGAAA.TG.A..A.C.TA.A.GCAT.GCTGT..AT.TG.AAC.TT.C.C.A.C..
Prunus mume                 .TGT.A.AAGGT.A.TTTGC.G..GT.-------------------------------------------------------------------

                               1720      1730      1740      1750      1760      1770      1780      1790      1800
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
```

104

```
Prunus persica Lhcb2 gene     ----------------------------------------------------------------------------------------------------
Modern Peach                  ----------------------------------------------------------------------------------------------------
Prunus persica (Feral Zhejiang ----------------------------------------------------------------------------------------------------
Prunus persica (Mountain Zheji ----------------------------------------------------------------------------------------------------
Prunus persica Dr. Davis      ----------------------------------------------------------------------------------------------------
Prunus persica Lovell         ----------------------------------------------------------------------------------------------------
Prunus persica F8-1           ----------------------------------------------------------------------------------------------------
Prunus armeniaca              CTTGCTTTTGTTGAGGTTCTCTTGTTGTGGACTCCATTAGTTCTGATGTAATCTTACTAGAAACTGCTTGTATTGAAATCACTTA-----
Prunus avium                  CTTGCCTTTGTTGAGGTTCTCTTGTTGAGGACTCCATTAGTTCTGATGTAATCTTACTAGAAACTGCTTGTATTGAAATCACTTATCCAG
Prunus mume                   ----------------------------------------------------------------------------------------------------
```

Figure 41. Alignment of *Lhcb2* gene of various *P. persica* cultivars, *Prunus sp.* and of DNA extracts from the seed of peach samples collected from the Zhejiang province in China.