

Genetic origins of the introduced pea weevil (*Bruchus pisorum*) population in Ethiopia

Loraine Cornelia Scheepers

Dissertation submitted in fulfilment of the requirements for the degree of

Magister Scientiae

in the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of the Free State

Supervisor

Prof. J.P. Grobler

Department of Genetics, University of the Free State

2012

Declaration

I declare that the dissertation hereby handed in for the qualification *Magister Scientiae* at the University of the Free State is my own work and that I have not previously submitted the same work for a qualification at/in any other University/Faculty.

Loraine Cornelia Scheepers

2012

Acknowledgements

This thesis would not have been possible if it was not for all the people who helped me throughout this study.

Firstly I would like to thank my supervisor, Prof. Paul Grobler, for his guidance and advice during the course of my studies and for his help in the preparation of this dissertation.

I would like to thank Birhane Asayehagne, for initiating this project and for providing pea weevil specimens from Ethiopia for this study.

I would like to thank Dr. Darryl Hardie, Dr. Helmut Saucke and Elizabeth Roberts for generously providing me with pea weevil specimens for this study. Without your contributions this thesis would not have been possible.

I would like to thank Christiaan Labuschagne and Antoinette van Schalkwyk of Inqaba Biotec for all their help with the sequencing.

I would like to thank the Department of Genetics at the University of the Free State for funding and accommodating this study.

Finally I would like to thank all my friends and family for their support throughout this study.

Table of Contents

Declaration	ii
Acknowledgements	iii
List of Figures	vi
List of Tables.....	viii
1 Introduction	1
1.1 Biology of <i>Bruchus pisorum</i> (Linnaeus, 1758)	1
1.2 Distribution of <i>Bruchus pisorum</i>	4
1.2.1 <i>Bruchus pisorum</i> in Ethiopia.....	4
1.2.2 <i>Bruchus pisorum</i> in Australia.....	6
1.2.3 <i>Bruchus pisorum</i> in the United States of America	7
1.2.4 <i>Bruchus pisorum</i> in Europe	8
1.3 Evolutionary origin of <i>Bruchus pisorum</i>	8
1.4 Molecular phylogeny and population genetics	10
1.4.1 Molecular markers.....	10
1.4.2 Phylogenetic and population genetic studies	12
1.4.3 Use of COX1, Cytb and EF-1 α gene sequences in phylogenetic and population genetic studies.....	13
1.4.4 Use of GenBank records in phylogenetic and population genetic studies	14
1.5 Objectives and scope of the current study	15
2 Research methodology	16
2.1 Populations analysed.....	16
2.1.1 Ethiopian specimens	16
2.1.2 Australian specimens	17
2.1.3 German specimens	17
2.1.4 American specimens.....	18
2.1.5 Supplementary populations used	18
2.2 DNA extraction.....	19
2.2.1 Australian, Ethiopian and German specimens	19
2.2.2 American museum specimens	20
2.3 DNA quantification	21
2.4 Genes chosen for analysis	21

2.5	Primer design.....	21
2.6	PCR (Polymerase Chain Reaction) amplification of gene regions	22
2.7	Sequencing of gene regions	24
2.8	Supplementary sequences	25
2.9	Data analysis	27
2.9.1	Sequence analysis	27
2.9.2	Tree building	29
2.9.3	Haplotype networks.....	33
2.9.4	Population diversity	34
2.9.5	Population differentiation.....	35
3	Results	36
3.1	DNA extraction and quantification.....	36
3.2	DNA amplification and sequencing	38
3.3	Sequence analysis.....	39
3.4	Phylogenetic trees	46
3.4.1	Phylogenetic trees constructed with COX1 sequences.....	46
3.4.2	Phylogenetic trees constructed with Cytb sequences	51
3.4.3	Phylogenetic trees constructed from EF-1 α sequences	58
3.5	Haplotype networks	60
3.6	Population diversity.....	64
3.6.1	Ethiopian subpopulations	64
3.6.2	Diversity within country-specific populations	71
4	Discussion.....	74
4.1	Determining the origin of <i>B. pisorum</i> in Ethiopia	74
4.1.1	Evolutionary origin of <i>B. pisorum</i>	75
4.1.2	Anthropogenic spread of <i>B. pisorum</i>	76
4.1.3	Identifying possible source populations of <i>B. pisorum</i> in Ethiopia ..	79
4.1.4	A possible endemic population of <i>B. pisorum</i> in Ethiopia.....	82
4.2	Diversity and differentiation across Ethiopian subpopulations	84
4.3	Notes on DNA extraction and gene amplification.....	86
4.4	Future avenues for research.....	87
5	Summary.....	88
6	Opsomming.....	89
7	References.....	90

List of Figures

FIGURE 1-1 <i>BRUCHUS PISORUM</i> LIFE CYCLE.....	2
FIGURE 1-2 WORLD MAP OF THE DISTRIBUTION OF <i>B. PISORUM</i> . AREAS IN WHICH <i>B. PISORUM</i> ARE KNOWN TO OCCUR ARE INDICATED IN RED (BERIM, 2008).....	4
FIGURE 1-3 DISTRIBUTION OF <i>B. PISORUM</i> IN ETHIOPIA. AREAS WHERE <i>B. PISORUM</i> IS KNOWN TO BE PRESENT ARE INDICATED IN RED (ASSAYEHGNE, 2002).....	5
FIGURE 1-4 DISTRIBUTION OF <i>B. PISORUM</i> IN AUSTRALIA. AREAS THAT ARE SHADED RED ARE AREAS WHERE PEA WEEVILS ARE KNOWN TO BE PRESENT (WATERHOUSE AND SANDS, 2001).	6
FIGURE 1-5 DISTRIBUTION OF <i>B. PISORUM</i> IN NORTH AMERICA. AREAS WHERE PEA WEEVILS ARE KNOWN TO BE PRESENT ARE SHADED RED (CAPINERA, 2001).....	7
FIGURE 1-6 DISTRIBUTION OF <i>B. PISORUM</i> IN EUROPE. AREAS WHERE PEA WEEVILS ARE KNOWN TO BE PRESENT ARE SHADED IN RED (DAISIE, 2008).	8
FIGURE 1-7 WORLD MAP ILLUSTRATING THE MOST LIKELY ORIGIN (IN RED) AND SPREAD (IN GREEN) OF PISUM SATIVUM CULTIVATION; AND THE AREAS WHERE PEA WEEVILS ARE KNOWN PESTS (PINK TRIANGLES) (ADAPTED BY BYRNE (2005) FROM BOCK, D (2005) AND LEFF <i>ET AL.</i> (2004)).	9
FIGURE 2-1 MAP OF THE AMHARA REGION IN ETHIOPIA. THE EBINAT DISTRICT IS INDICATED IN GREEN AND THE YILMANA DENSA REGION IS INDICATED IN PINK.	17
FIGURE 3-1 EVOLUTIONARY RELATIONSHIPS OF COX1 SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE NEIGHBOUR-JOINING METHOD.....	47
FIGURE 3-2 EVOLUTIONARY RELATIONSHIPS OF COX1 SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE MINIMUM EVOLUTION METHOD.....	47
FIGURE 3-3 EVOLUTIONARY RELATIONSHIPS OF COX1 SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE UPGMA METHOD.	47
FIGURE 3-4 MAXIMUM PARSIMONY ANALYSIS OF COX1 SEQUENCES OF <i>B. PISORUM</i> POPULATIONS.....	47
FIGURE 3-5 MOLECULAR PHYLOGENETIC ANALYSIS BY MAXIMUM LIKELIHOOD USING COX1 SEQUENCES. 48	
THE EVOLUTIONARY HISTORY WAS INFERRED BY USING THE MAXIMUM LIKELIHOOD METHOD BASED ON THE JUKES-CANTOR MODEL (JUKES AND CANTOR 1969). THE BOOTSTRAP CONSENSUS TREE INFERRED FROM 1000 REPLICATES IS TAKEN TO REPRESENT THE EVOLUTIONARY HISTORY OF THE TAXA ANALYSED (FELSENSTEIN 1985). BRANCHES CORRESPONDING TO PARTITIONS REPRODUCED IN LESS THAN 50% BOOTSTRAP REPLICATES ARE COLLAPSED. THE PERCENTAGE OF REPLICATE TREES IN WHICH THE ASSOCIATED TAXA CLUSTERED TOGETHER IN THE BOOTSTRAP TEST (1000 REPLICATES) ARE SHOWN NEXT TO THE BRANCHES. INITIAL TREE(S) FOR THE HEURISTIC SEARCH WERE OBTAINED AUTOMATICALLY AS FOLLOWS: WHEN THE NUMBER OF COMMON SITES WAS < 100 OR LESS THAN ONE FOURTH OF THE TOTAL NUMBER OF SITES, THE MAXIMUM PARSIMONY METHOD WAS USED, OTHERWISE BIONJ METHOD WITH MCL DISTANCE MATRIX WAS USED. THE ANALYSIS INVOLVED 20 NUCLEOTIDE SEQUENCES. ALL POSITIONS CONTAINING GAPS AND MISSING DATA WERE ELIMINATED. THERE WERE A TOTAL OF 642 POSITIONS IN THE FINAL DATASET. EVOLUTIONARY ANALYSES WERE CONDUCTED IN MEGA5 (TAMURA, ET AL. 2011).....	48

FIGURE 3-6 PHYLOGENETIC TREE INFERRED BY THE BAYESIAN ANALYSIS OF PHYLOGENY CONSTRUCTED WITH COX1 SEQUENCES OF <i>B. PISORUM</i>	50
FIGURE 3-7 EVOLUTIONARY RELATIONSHIPS OF CYTB SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE NEIGHBOUR-JOINING METHOD.	52
FIGURE 3-8 EVOLUTIONARY RELATIONSHIPS OF CYTB SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE MINIMUM EVOLUTION METHOD.	52
FIGURE 3-9 EVOLUTIONARY RELATIONSHIPS OF CYTB SEQUENCES OF <i>B. PISORUM</i> POPULATIONS INFERRED USING THE UPGMA METHOD.....	53
FIGURE 3-10 MAXIMUM PARSIMONY ANALYSIS OF CYTB SEQUENCES OF <i>B. PISORUM</i> POPULATIONS.	53
FIGURE 3-11 MOLECULAR PHYLOGENETIC ANALYSES BY MAXIMUM LIKELIHOOD METHOD USING CYTB SEQUENCES.....	55
THE EVOLUTIONARY HISTORY WAS INFERRED USING THE MAXIMUM LIKELIHOOD METHOD BASED ON THE JUKES-CANTOR MODEL (JUKES AND CANTOR 1969). THE BOOTSTRAP CONSENSUS TREE INFERRED FROM 1000 REPLICATES IS TAKEN TO REPRESENT THE EVOLUTIONARY HISTORY OF THE TAXA ANALYSED (FELSENSTEIN 1985). BRANCHES CORRESPONDING TO PARTITIONS REPRODUCED IN LESS THAN 50% BOOTSTRAP REPLICATES ARE COLLAPSED. THE PERCENTAGE OF REPLICATES TREES IN WHICH THE ASSOCIATED TAXA CLUSTERED TOGETHER IN THE BOOTSTRAP TEST (1000 REPLICATES) ARE SHOWN NEXT TO THE BRANCHES. INITIAL TREE(S) FOR THE HEURISTIC SEARCH WERE OBTAINED AUTOMATICALLY AS FOLLOWS. WHEN THE NUMBER OF COMMON SITES WAS < 100 OR LESS THAN ONE FOURTH OF THE TOTAL NUMBER OF SITES, THE MAXIMUM PARSIMONY METHOD WAS USED, OTHERWISE BIONJ METHOD WITH MCL DISTANCE MATRIX WAS USED. THE ANALYSIS INVOLVED 19 NUCLEOTIDE SEQUENCES. CODON POSITIONS INCLUDED WERE 1ST+2ND+3RD+NONCODING. ALL POSITIONS CONTAINING GAPS AND MISSING DATA WERE ELIMINATED. THERE WAS A TOTAL OF 251 POSITIONS IN THE FINAL DATASET. EVOLUTIONARY ANALYSES WERE CONDUCTED IN MEGA5 (TAMURA, ET AL. 2011).....	55
FIGURE 3-12 PHYLOGENETIC TREE INFERRED BY THE BAYESIAN ANALYSIS OF PHYLOGENY CONSTRUCTED WITH CYTB SEQUENCES OF <i>B. PISORUM</i>	56
FIGURE 3-13 PHYLOGENETIC TREE INFERRED BY THE BAYESIAN ANALYSIS OF PHYLOGENY CONSTRUCTED WITH EF-1A SEQUENCES OF <i>B. PISORUM</i>	59
FIGURE 3-14 A MINIMUM-SPANNING NETWORK BETWEEN CYTB HAPLOTYPES OF <i>B. PISORUM</i> . NODES REPRESENT HAPLOTYPES AND THE SIZES OF THE NODES ARE PROPORTIONAL TO THE NUMBER OF INDIVIDUALS THAT SHARE THAT HAPLOTYPE. THE NUMBER OF MUTATIONAL EVENTS BETWEEN TWO HAPLOTYPES ARE INDICATED WITH CROSSBARS. WHEN MORE THAN FOUR MUTATIONAL EVENTS OCCURRED THE NUMBER OF EVENTS IS INDICATED NEXT TO THE INTERRUPTED LINES. COLOURS ARE USED TO INDICATE IN WHICH POPULATIONS THE HAPLOTYPES OCCUR. AUSTRALIA ●, CHINA ●, ETHIOPIA ●, GERMANY ●, JAPAN ●, USA ●, OUTGROUP HAPLOTYPE OF <i>B. RUFIMANUS</i> ●.	62
FIGURE 3-15 A MINIMUM-SPANNING NETWORK BETWEEN COX1 HAPLOTYPES OF <i>B. PISORUM</i> . NODES REPRESENT HAPLOTYPES AND THE SIZES OF THE NODES ARE PROPORTIONAL TO THE NUMBER OF INDIVIDUALS THAT SHARE THAT HAPLOTYPE. THE NUMBER OF MUTATIONAL EVENTS BETWEEN TWO HAPLOTYPES ARE INDICATED WITH CROSSBARS. WHEN MORE THAN FOUR MUTATIONAL EVENTS OCCURRED THE NUMBER OF EVENTS IS INDICATED NEXT TO THE INTERRUPTED LINES. COLOURS ARE USED TO INDICATE IN WHICH POPULATIONS THE HAPLOTYPES OCCUR. AUSTRALIA ●, CHINA ●, ETHIOPIA ●, GERMANY ●, USA ●, OUTGROUP HAPLOTYPE OF <i>B. RUFIMANUS</i> ●.	63

List of Tables

TABLE 2-1 PRIMER SETS DESIGNED FOR THE EF-1A, COX1 AND CYTB GENE REGIONS.	22
TABLE 2-2 LIST OF <i>B. PISORUM</i> SEQUENCES DOWNLOADED FROM GENBANK.	25
TABLE 2-3 LIST OF <i>B. RUFIMANUS</i> SEQUENCES DOWNLOADED FROM GENBANK.	26
TABLE 3-1 NUCLEIC ACID CONSECRATION RANGE OF DNA EXTRACTED FROM SPECIMENS OF DIFFERENT POPULATIONS	36
TABLE 3-2 NUCLEIC ACID CONCENTRATIONS OF DNA EXTRACTED FROM USA MUSEUM SPECIMENS	37
TABLE 3-3 THE NUMBER OF SEQUENCES OBTAINED FOR THE COX1, CYTB AND EF-1A GENES, AS WELL AS THE MINIMUM AND MAXIMUM SEQUENCE LENGTHS IN THE INDIVIDUAL POPULATIONS.	39
TABLE 3-4 THE VARIATION OBSERVED IN THE COX1, CYTB AND EF-1A SEQUENCE ALIGNMENTS OF <i>B. PISORUM</i> FROM ETHIOPIA, AUSTRALIA, GERMANY, THE USA, CHINA AND JAPAN	41
TABLE 3-5 COX1 HAPLOTYPE FREQUENCIES IN THE DIFFERENT POPULATIONS	42
TABLE 3-6 CYTB HAPLOTYPE FREQUENCIES IN THE DIFFERENT POPULATIONS	43
TABLE 3-7 EF-1A GENOTYPE OCCURRENCES IN THE DIFFERENT POPULATIONS	44
TABLE 3-8 GENETIC DIVERSITY IN THE <i>B. PISORUM</i> ETHIOPIAN SUBPOPULATIONS ACCORDING TO COX1 SEQUENCE DATA	64
TABLE 3-9 GENETIC DIVERSITY IN <i>B. PISORUM</i> ETHIOPIAN SUBPOPULATIONS ACCORDING TO CYTB SEQUENCE DATA	65
TABLE 3-10 GENETIC DIVERSITY IN <i>B. PISORUM</i> ETHIOPIAN SUBPOPULATIONS ACCORDING TO EF-1A SEQUENCE DATA	66
TABLE 3-11 F_{ST} VALUES AND F_{ST} P VALUES BETWEEN PAIRS OF ETHIOPIAN SUBPOPULATIONS BASED ON COX1 SEQUENCE DATA. THE POPULATIONS ARE AS FOLLOWS: (1) ETH1, (2) ETH2, (3) ETH3, (4) ET4, (5) ET5 AND (6) AN OUTGROUP GERMAN POPULATION. VALUES HIGHLIGHTED IN GREEN INDICATE NO SIGNIFICANT DIFFERENTIATION BETWEEN POPULATION PAIRS, WHILE VALUES HIGHLIGHT IN RED INDICATE SIGNIFICANT DIFFERENTIATION BETWEEN THE POPULATION PAIRS.	68
TABLE 3-12 F_{ST} VALUES AND F_{ST} P VALUES BETWEEN PAIRS OF ETHIOPIAN SUBPOPULATIONS BASED ON CYTB SEQUENCE DATA. THE POPULATIONS ARE AS FOLLOWS: (1) ETH1, (2) ETH2, (3) ETH3, (4) ET4, (5) ET5 AND (6) AN OUTGROUP GERMAN POPULATION. VALUES HIGHLIGHTED IN GREEN INDICATE NO SIGNIFICANT DIFFERENTIATION BETWEEN POPULATION PAIRS, WHILE VALUES HIGHLIGHT IN RED INDICATE SIGNIFICANT DIFFERENTIATION BETWEEN THE POPULATION PAIRS.	69
TABLE 3-13 F_{ST} VALUES AND F_{ST} P VALUES BETWEEN PAIRS OF ETHIOPIAN SUBPOPULATIONS BASED ON EF-1A SEQUENCE DATA. THE POPULATIONS ARE AS FOLLOWS: (1) ETH1, (2) ETH2, (3) ETH3, (4) ET4, (5) ET5 AND (6) AN OUTGROUP GERMAN POPULATION. VALUES HIGHLIGHTED IN GREEN INDICATE NO SIGNIFICANT DIFFERENTIATION BETWEEN POPULATION PAIRS, WHILE VALUES HIGHLIGHT IN RED INDICATE SIGNIFICANT DIFFERENTIATION BETWEEN THE POPULATION PAIRS.	70

TABLE 3-14 GENETIC DIVERSITY IN *B. PISORUM* POPULATIONS ACCORDING TO CYTB SEQUENCE DATA..... 71

TABLE 3-15 GENETIC DIVERSITY IN *B. PISORUM* POPULATIONS ACCORDING TO COX1 SEQUENCE DATA ... 72

TABLE 3-16 GENETIC DIVERSITY IN *B. PISORUM* POPULATIONS ACCORDING TO EF-1A SEQUENCE DATA ... 73

1 Introduction

1.1 Biology of *Bruchus pisorum* (Linnaeus, 1758)

Bruchus pisorum, the pea weevil, is a cosmopolitan insect pest of *Pisum sativum*, the field pea. The pea weevil is not a true weevil as its name suggests, but belongs to another family of plant-feeding beetles. *Bruchus pisorum* is classified under the order Coleoptera, family Chrysomelidae and subfamily Bruchinae, whereas true weevils belong to the family Curculionidae. The pea weevil has a chromosome number of $2n=34$ ($n\♂=16+Xy$) (Lachowska *et al.*, 1998).

The adult pea weevil is a small beetle, varying between 4 and 5 mm in length. The pea weevil has a black body with white markings on its abdomen (Berim, 2008). Pea weevil eggs are yellow in colour and are cigar shaped with a length of 1.5 mm and width of 0.6 mm. The larvae are legless, cream coloured and grow to a length of approximately 5 mm (Berim, 2008; Baker, 1998).

Bruchus pisorum is a strictly monophagous bruchid and completes its univoltine life cycle on *P. sativum* (Hardie and Clement, 2001). Pea weevils become active in spring, when temperatures reach about 20°C. The female pea weevils are not sexually mature when they leave hibernation, and must feed on pea pollen to enable ovarian development (Hardie and Clement, 2001). On average pea weevils start laying eggs about 2 weeks after invading flowering pea fields. The eggs are deposited singly or in small clusters on the exterior of green pea pods (Baker, 1998). Depending on the temperature, the eggs will usually hatch within three to five days. When the larvae hatch, they bore directly from the egg into the pea pod. They then continue to bore into the pea seed. The larval stage has a duration of about 7-11 weeks (Hardie *et al.*, 1995).

When the larva reaches maturity it has consumed most or all of the seed contents leaving a thin layer of the seed coat in one area of the seed. After completing its larval development the pea weevil pupates in the seed; this stage lasts in the region of 2-3 weeks (Hardie *et al.*, 1995).

The adult normally pushes through the thin layer of seed coat and escapes after completing its development. The adult pea weevils then seek sheltered areas to hibernate through the winter and remain there until spring (Hardie *et al.*, 1995). However, under cool and dry conditions, the weevil may remain within the seed during the winter and emerge in the spring. It may remain in the pea seed for 18-24 months and is consequently easily transported with dried peas (Capinera, 2001). Figure 1-1 illustrates the life cycle of the pea weevil.

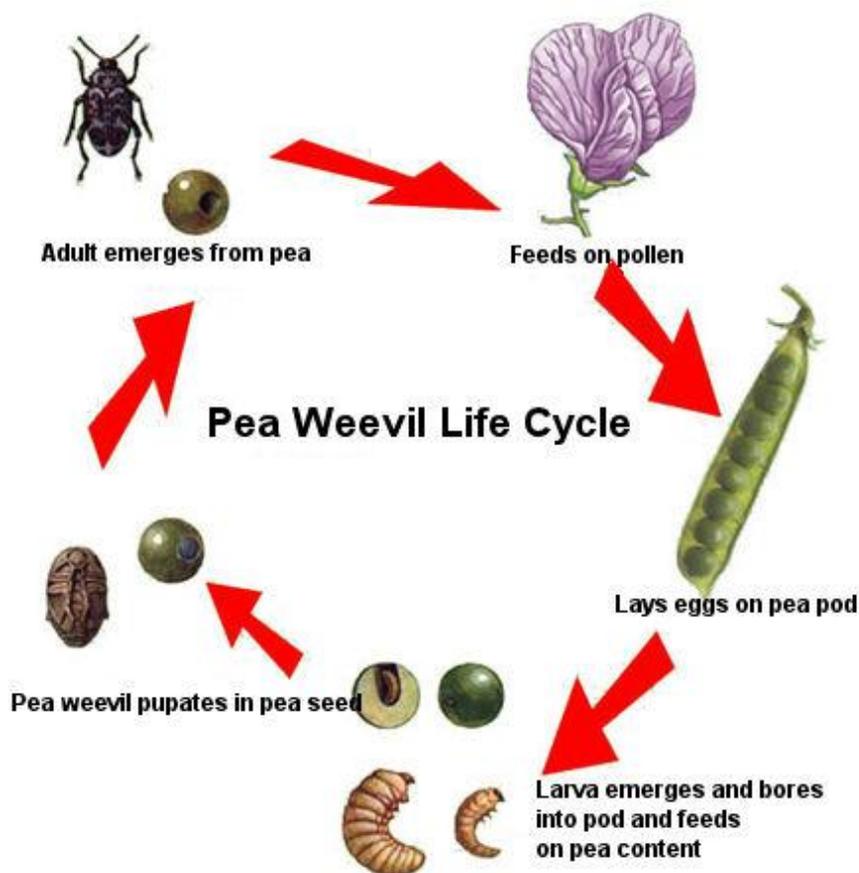


Figure 1-1 *Bruchus pisorum* life cycle

Pea seeds are damaged by the larvae feeding within the pea seeds. Each larva destroys a single pea seed. Peas contaminated with even a few larvae are unsuitable for human consumption (Capinera, 2001). When peas are grown for livestock feed, the infestation lowers the weight and value of the peas. Furthermore, infested peas that are used for livestock feed may be a source of pea weevil infestation if peas are grown nearby. Peas that are damaged by pea weevils have a reduced germination potential, may be a source of infestation and is undesirable for use as seed (Capinera, 2001).

Management of pea weevils is predominantly done by the use of insecticides, which is applied when the crop is in bloom. The pea weevils are killed when they land in the field to consume pollen. When infected peas are intended for seed, fumigation is recommended to kill any pea weevil that might be present in the peas (Capinera, 2001). Cultural techniques are also used to help reduce the incidence of pea weevil infestation. These techniques incorporate the destruction of crop residues that may contain pea weevils and an early planting and harvesting strategy. Peas that are planted early in the season and harvested early show a lower incidence of infestation (Capinera, 2001). When peas are grown for hay, pea hay should be cut no later than the beginning of flowering to prevent the presence of pods containing seeds that could be infested with pea weevils (Baker, 1998).

1.2 Distribution of *Bruchus pisorum*

Pea weevils are widely distributed and can be found in Europe, Asia, North Africa, North, Central, and South America and South-Western Australia (Berim, 2008). Figure 1-2 shows the current global distribution of *B. pisorum*.

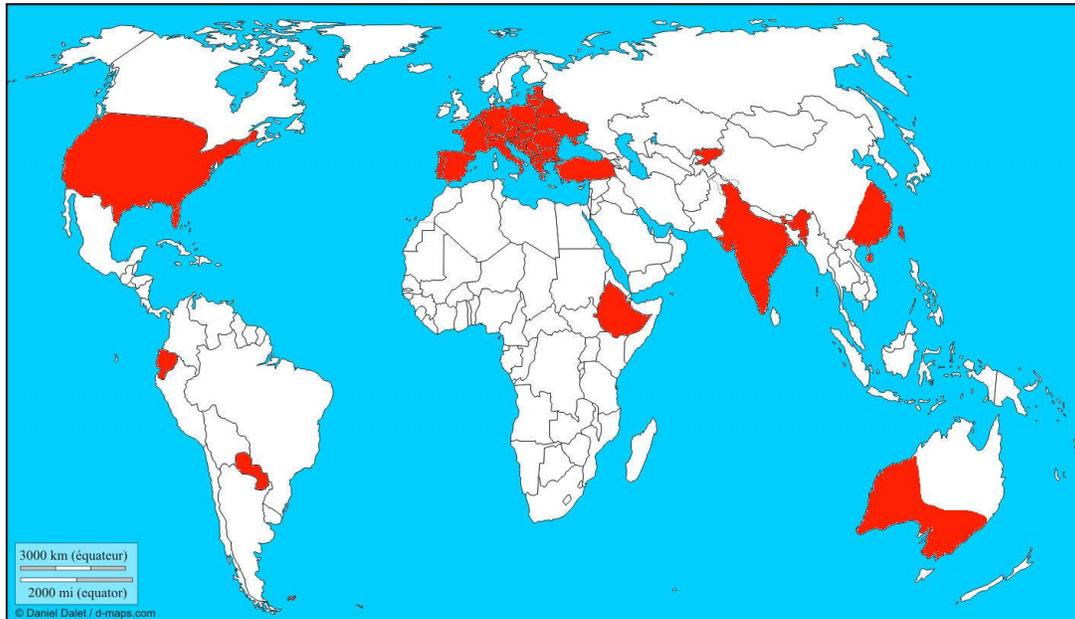


Figure 1-2 World map of the distribution of *B. pisorum*. Areas in which *B. pisorum* are known to occur are indicated in red (Berim, 2008).

1.2.1 *Bruchus pisorum* in Ethiopia

The pea weevil was first documented in Ethiopia in 1985. From 1992 the pea weevil has been reported to be a substantial insect pest of field peas in North West Ethiopia (Esmelealem and Adane, 2007), with the insects found in the warmer areas around Bahir Dar and the highlands of Motta. In Ethiopia, adult pea weevils enter pea fields in early August where they feed on pollen (Assayehegne, 2002). Pea weevils cause crop losses of up to 80% in the Ebinat and South Gonder areas and losses of about 45% in the Wag Himra area (Bekele *et al.*, 2006).

The pea weevil was most likely accidentally introduced to Ethiopia in the mid 1970s. It was possibly introduced with food aid received by Ethiopia during the severe famine the country experienced during that period. The pea weevil is now extensively distributed in the Amhara National Regional State (ANRS) (Birhane Asayehegne, personal communication, November 2010). The pea weevil is also rapidly spreading to neighbouring regions. Farmers in Ethiopia have reported that the pea weevil is the most destructive pest yet seen on pea plants and crop losses as high as 85% have been recorded (Esmelealem and Adane, 2007). Figure 1-3 shows a map of the distribution of *B. pisorum* in Ethiopia.

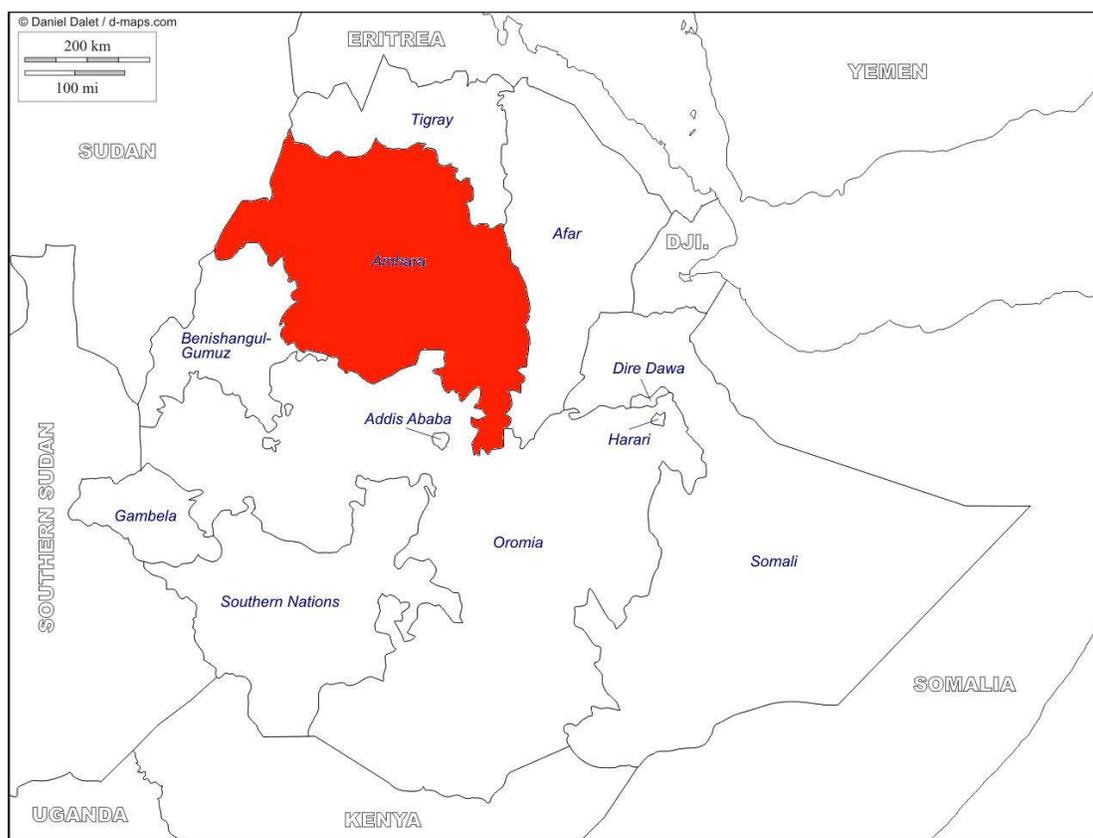


Figure 1-3 Distribution of *B. pisorum* in Ethiopia. Areas where *B. pisorum* is known to be present are indicated in red (Assayehegne, 2002).

1.2.2 *Bruchus pisorum* in Australia

In Australia the pea weevil was first recorded in 1931 in Western Australia (Waterhouse and Sands 2001). Within 30 years *B. pisorum* spread to other southern mainland states. Presently *B. pisorum* is distributed throughout South-western Australia. In high rainfall areas more than 70% of field pea seeds may be infected, leading to significant economic losses. (Waterhouse and Sands, 2001). Figure 1-4 shows the current distribution of pea weevils in Australia.

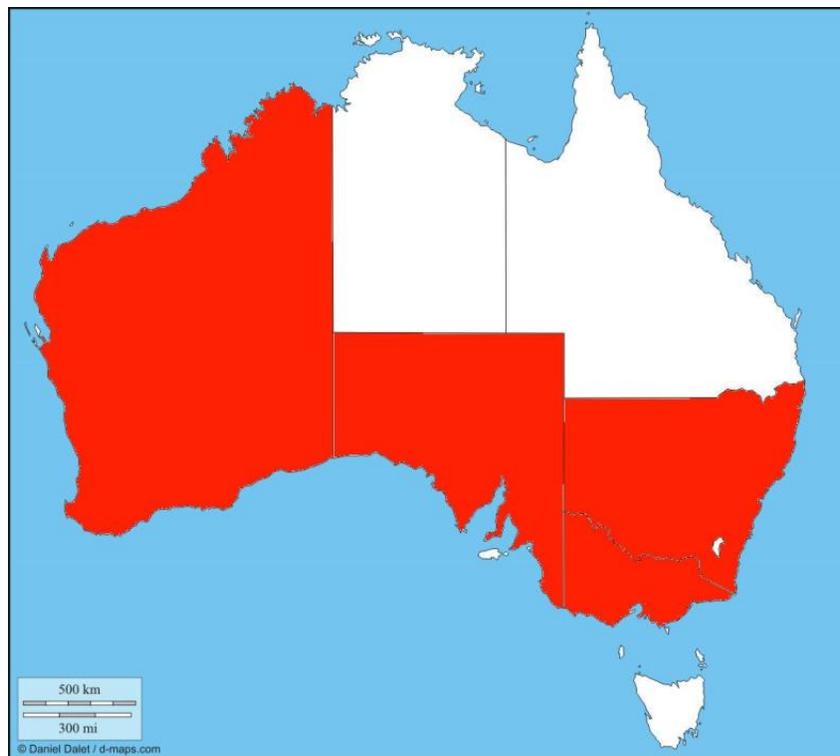


Figure 1-4 Distribution of *B. pisorum* in Australia. Areas that are shaded red are areas where pea weevils are known to be present (Waterhouse and Sands, 2001).

1.2.3 *Bruchus pisorum* in the United States of America

In the USA documentary records indicate that the pea weevil was doing damage to crops in the colonies as early as 1675 (Bain, 1998). Records also show that shipments of dry pea seeds arrived in the Massachusetts Bay Colony as early as 1628 (Bain, 1998). These shipments could have harboured pea weevils and thus provided the source population of *B. pisorum* in the USA.

The pea weevil was first documented near Philadelphia, Pennsylvania, in the 1740s and then noted in nearby states in the 1750s (Capinera, 2001). By the 1890s the pea weevil had spread across the USA to Washington and Oregon states (Capinera, 2001). Presently the pea weevil is distributed throughout the USA and southern Canada. Figure 1-5 shows the current distribution of the pea weevil in North America.



Figure 1-5 Distribution of *B. pisorum* in North America. Areas where pea weevils are known to be present are shaded red (Capinera, 2001).

1.2.4 *Bruchus pisorum* in Europe

The pea weevil is distributed widely across Europe. Their distribution only excludes the coldest of the European regions. The first record of the pea weevil in Europe was in the Czech Republic in 1850 (Beenen and Roques, 2010). Figure 1-6 indicates the current distribution of *B. pisorum* in Europe.

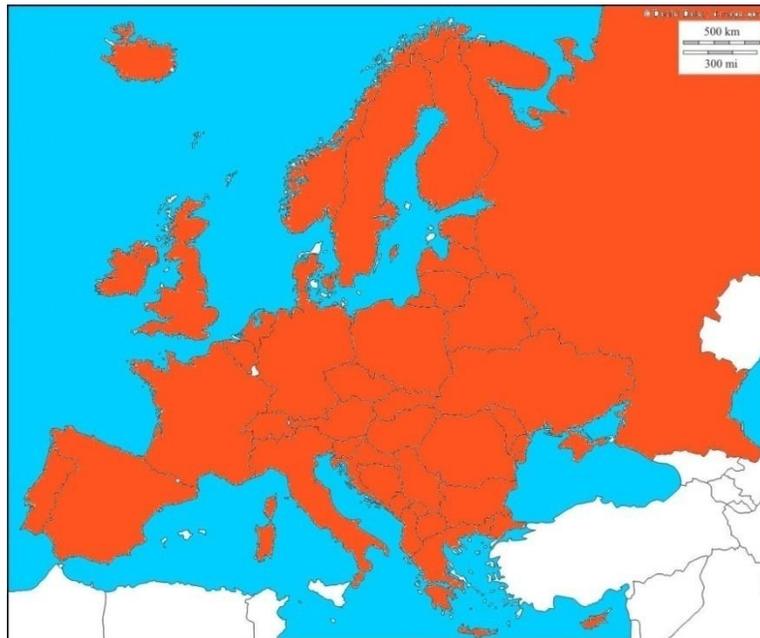


Figure 1-6 Distribution of *B. pisorum* in Europe. Areas where pea weevils are known to be present are shaded in red (DAISIE, 2008).

1.3 Evolutionary origin of *Bruchus pisorum*

The evolutionary origin of the pea weevil is uncertain but it is believed to have co-evolved with its host *P. sativum* (Byrne, 2005). If *B. pisorum* co-evolved with *P. sativum*, then the evolutionary origin of *B. pisorum* will most likely mirror the origin and spread of *P. sativum* and it could thus have originated in the same geographical region. The origin of *P. sativum* is not well known but the Mediterranean region, western and central Asia and Ethiopia have been indicated as possible centres of origin (Byrne, 2005).

Archaeological evidence of the use of peas has been found in the Fertile Crescent, dating back to 8000 BC (Messiaen *et al.*, 2006). Peas were most likely first cultivated in western Asia, from where it spread to Europe, China and India. In the mountainous regions of Central and East Africa peas were already well known before the arrival of Europeans. Presently *P. sativum* is cultivated in all temperate countries and in most tropical highlands (Messiaen *et al.*, 2006). Figure 1-7 illustrates the most likely origin and spread of the cultivation of *P. sativum*.

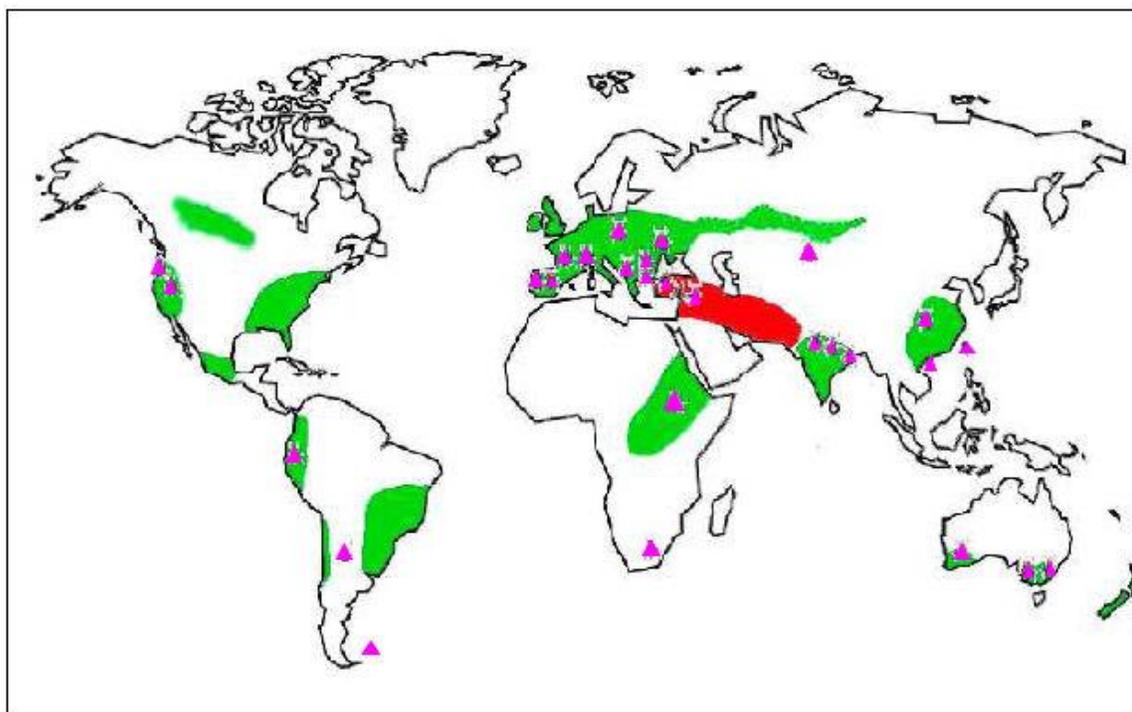


Figure 1-7 World map illustrating the most likely origin (in red) and spread (in green) of *Pisum sativum* cultivation; and the areas where pea weevils are known pests (pink triangles) (adapted by Byrne (2005) from Bock, D (2005) and Leff *et al.* (2004)).

1.4 Molecular phylogeny and population genetics

Molecular phylogenetics is the study of the evolutionary development and the line of descent of species by using molecular data such as nucleic acid (DNA and RNA) and protein sequences. Nucleic acid and proteins are information molecules since they retain a record of the organism's evolutionary history. When molecular phylogenetics is used, nucleic acid or protein sequences from different organisms are compared by using various statistical approaches (Wiser, 2008). These approaches are used to estimate the evolutionary relationships based on the degree of homology between the sequences. The evolutionary distance between two organisms is reflected in the differences of the nucleotides and amino acids. The closer related organisms are, the fewer sequence differences there will be (Wiser, 2008).

Population genetics studies the genetic composition of biological populations, patterns of connectivity and the changes in genetic composition that result from the occurrence of various factors, including natural selection. Population genetics uses mathematical models of gene frequency dynamics to determine the likely pattern of genetic variation in populations (Okasha, 2008).

1.4.1 Molecular markers

Molecular markers are used in the study of phylogeny, evolution and population genetics. A molecular marker is a DNA sequence used to mark a particular location (locus) on a particular chromosome (Okumuş and Çiftci, 2003). There are two categories of markers, namely protein and DNA markers. In population genetic and phylogenetic studies, three classes of genetic markers are generally used: allozymes, mitochondrial DNA and nuclear DNA (Okumuş and Çiftci, 2003).

Allozymes were the first true molecular markers to be used to study variation (Sclötterer, 2004). Genetic variations give rise to protein variations called allozymes. Allozymes differ in electrical charge, allowing them to be separated with the use of electrophoreses. Allozyme variation provides data on single locus genetic variation (Okumuş and Çiftci, 2003). One of the criticisms against allozyme markers is that they are an indirect and insensitive method of detecting variation in DNA, due to the redundancy of the genetic code (Sclötterer, 2004).

In recent times, a multitude of markers based on direct study of DNA have been developed. These markers include RFLPs (Restriction Fragment Length Polymorphism), RAPDs (Randomly Amplified Polymorphic DNA), VNTRs (Variable Number of Tandem Repeats), SNPs (Single Nucleotide Polymorphisms) and nucleotide sequences (Sclötterer, 2004). A RFLP is a polymorphism defined by restriction fragment lengths that is produced by a specific restriction endonuclease (Okumuş and Çiftci, 2003). RAPDs are based on random amplification of anonymous loci by PCR. RAPD markers do not require any prior knowledge of primer sequences in the targeted species (Sclötterer, 2004). SNPs are polymorphisms due to a single nucleotide substitution or a single nucleotide insertion or deletion (Okumuş and Çiftci, 2003). VNTRs are repeated segments of nuclear DNA. The number of tandem repeats varies in number at different loci and is dispersed throughout the genome (Okumuş and Çiftci, 2003). VNTRs can be classified as either mini- or microsatellites. A minisatellite is composed of tandem repeats with a length of between 9 and 65bp. Microsatellites are composed of tandem repeats with a length of between 2 and 8bp (Okumuş and Çiftci, 2003).

Mitochondrial and nuclear gene sequences are regularly used as molecular markers when studying phylogeny, molecular evolution and population genetics. In general mitochondrial genes have several advantages over nuclear genes in phylogenetic and population genetic studies due to the fact that these have higher evolutionary (mutation) rates, lack recombination, are inherited in haploid mode, lack introns and appear to be selectively neutral. Mitochondrial genes have highly conserved gene content (Saeb, 2006; Gasser *et al.*, 2002; Hu *et al.*, 2002b; Liu *et al.*, 1999; Avise, 1994). The high evolution rates of mitochondrial DNA (mtDNA) genes allow for their use to compare both inter and intraspecific variation. Sequences of nuclear genes are nevertheless invaluable for specific applications.

1.4.2 Phylogenetic and population genetic studies of insects

DNA sequencing has become the preferred method for most molecular systematic studies. Mitochondrial DNA and nuclear rDNA (ribosomal DNA) are the most commonly sequenced regions in insect systematics (Caterino *et al.*, 2000). For mtDNA the most frequently sequenced genes are the COI (Cytochrome c oxidase subunit 1), COII, 16S rDNA and 12S rDNA gene regions. The COI (COX1) gene is one of the genes used in the Barcode of Life Project. For almost all animal groups a 648 base-pair region in the COX1 gene is being used as the standard barcode (Ratnasingham and Hebert, 2007). The Cytochrome b (Cytb), NADH dehydrogenase 5 (ND5) and COIII genes have also been sequenced and used, but to a lesser extent (Caterino *et al.*, 2000). For nuclear rDNA the 18S rRNA and 28S rRNA genes have been the most extensively used (Caterino *et al.*, 2000).

Nuclear protein-coding genes are used far less frequently in phylogenetic studies of insects than mtDNA or nuclear rDNA. Recently a few nuclear protein-coding genes have come into wider use. Of these genes, Elongation factor 1-alpha (EF-1 α) has been the most widely used (Caterino *et al.*, 2000).

1.4.3 Use of COX1, Cytb and EF-1 α gene sequences in phylogenetic and population genetic studies

The protein coding mitochondrial genes COX1 and Cytb have functions in cell respiration. Cytb and COX1 are the only protein coding genes among mitochondrial genes that occur in all eukaryotes (Stoeckle and Asubel, 2003). The protein coding mitochondrial gene Cytochrome c oxidase subunit 1 (COX1) provides an ideal marker for studying population genetic structure and molecular evolution (Blouin, 2002; Saeb, 2006). COX1 has therefore been shown to be of use in population-based studies of a range of invertebrates (Hu *et al.*, 2002a).

Cytochrome b is an additional example of a mitochondrial gene that is frequently used in molecular phylogeny studies (Kergoat *et al.*, 2007a). Cytb has been proven to have the same level of sequence variation as COX1 in phylogenetic analysis of numerous insect orders (Simmons and Weller, 2001). Sequences of the Cytb region together with sequences of the COX1 region have thus been used for the analysis of genetic variation and phylogeny of insect populations (Ito *et al.*, 2011).

The EF-1 α gene has been one of the most widely used nuclear protein-coding genes in the study of invertebrates. EF-1 α gene sequences have been proven of use in studies among species groups and genera within sub-families (Caterino *et al.*, 2000). EF-1 α gene sequences have been used in numerous molecular phylogeny studies of insects (Johnson *et al.*, 2001; Schoville *et al.*, 2011).

1.4.4 Use of GenBank records in phylogenetic and population genetic studies

GenBank is the NCBI (National Centre for Biotechnology Information) sequence database. It is an open access, annotated collection of all publicly available nucleotide sequences. The most important source of new data for GenBank is direct submissions from scientists. In February 2012 there were 149,819,246 sequence records in the traditional GenBank divisions (NCBI, 2012).

When gene sequences of a specific species is required, GenBank can be searched for sequences of this species either by using search terms like the scientific or common name of a species or by doing a comparison search with an available gene sequence by using the BLAST function. BLAST (Basic Local Alignment Search Tool) is an algorithm that compares a query sequence with a database of sequences and identifies sequences that resembles the query sequence. GenBank's database can also be searched for protein and translated nucleotide sequences.

GenBank is a valuable resource when undertaking phylogenetic and population genetic studies. Sequences obtained from GenBank can be used to increase the number of sequences in the data sets when conducting phylogenetic and population genetic studies and so increase the significance of the results. It simplifies the process of finding sequence data of a species of interest in different geographical regions as well as the process of finding sequences for use as an outgroup.

1.5 Objectives and scope of the current study

To determine the origin of *Bruchus pisorum* found in Ethiopia, *B. pisorum* specimens from different populations in several countries were obtained. These specimens served as reference populations to which the Ethiopian population could be compared. Three gene regions, namely COX1, Cytb and EF-1 α , were then sequenced in all of these specimens and compared with gene sequences from Ethiopian specimens as well as sequences downloaded from GenBank.

To determine the current population structure of *B. pisorum* across Ethiopia, *B. pisorum* specimens from different Ethiopian populations in different geographical regions in Ethiopia were collected. Gene sequences from these specimens were then produced. Molecular and statistical techniques were used to determine the diversity and differentiation within and among the Ethiopian subpopulations.

2 Research methodology

2.1 Populations analysed

Pea weevil specimens were acquired from four countries, specifically Ethiopia, Australia, Germany and the United States of America. Supplementary sequence data of pea weevils from two additional countries, China and Japan, were obtained from GenBank.

2.1.1 Ethiopian specimens

In Ethiopia, pea weevil specimens were obtained by collecting contaminated peas. In total 74 pea weevil specimens collected in Ethiopia were used in this study. Peas were collected from five localities in October 2007. In the Ebinat district, peas were collected from three locations with diverse average temperatures. The locations were: Qualias, which experiences high temperatures; Selamaya, which experiences moderate temperatures; and Jiman Derega, which experiences comparatively colder temperatures. In the Yilmana Densa district peas were collected from the Geboya and Adet Hanna areas. Both the Ebinat and Yilmana Densa districts are located in the Amhara Region. Figure 2-1 shows a map of the Amhara Region. The Ebinat and Yilmana Densa districts, where the samples were collected, are highlighted in green and pink respectively.

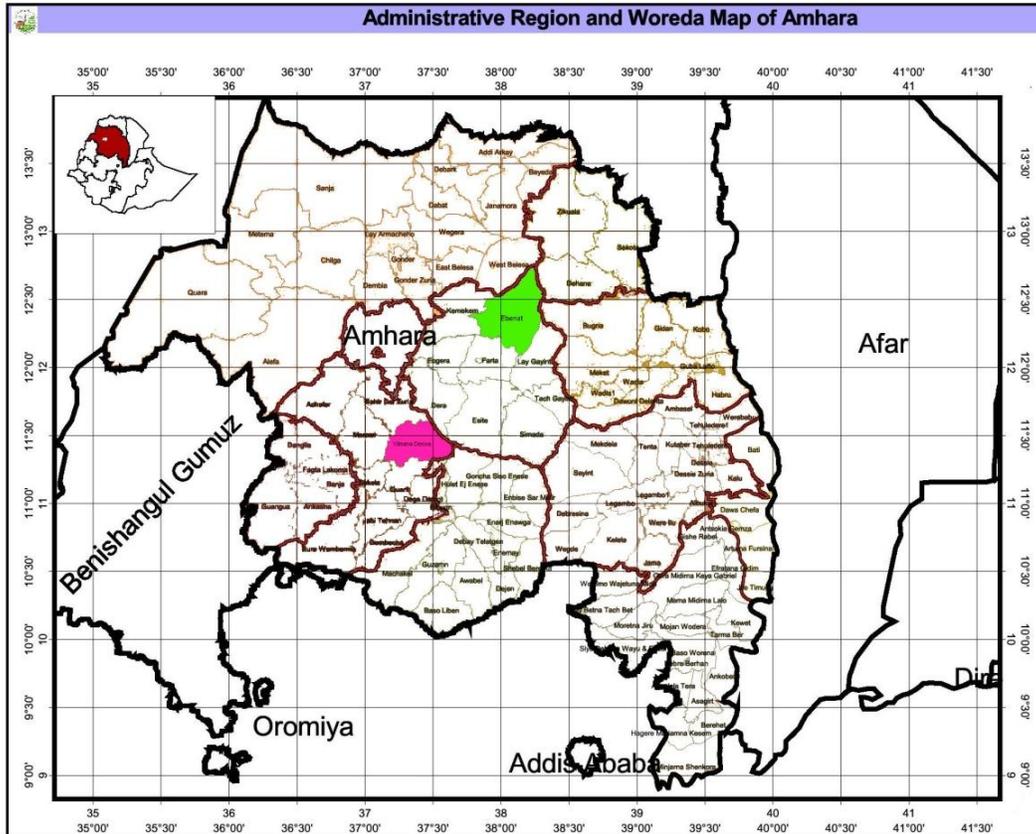


Figure 2-1 Map of the Amhara region in Ethiopia. The Ebinat district is indicated in green and the Yilmana Densa region is indicated in pink.

2.1.2 Australian specimens

As representative of Australian pea weevils, laboratory raised specimens of *B. pisorum* were received from Dr Darryl Hardie, Senior Entomologist at the Department of Agriculture & Food in Western Australia.

2.1.3 German specimens

For Germany, six dried pea weevil specimens were received from Dr. Helmut Saucke of the Kassel University in Germany. The six pea weevils were collected from dried organic peas in eastern Germany in 2010.

2.1.4 American specimens

From the United States of America, 21 pea weevil specimens were obtained from the Smithsonian Museum collection. The specimens were collected in eight different locations at different times 70 - 100 years ago. Seven specimens were collected in Walla-Walla County, Washington, between 30 July and 30 August 1938. During September 1929 six specimens were collected in Moscow, Idaho. One specimen was collected on 5 February 1931 in Willamette Valley, Oregon. In Portland, Oregon, one specimen was collected during November 1923. One specimen was collected in Indiana. In 1913 one specimen was collected in New York. On July 13 1923 one specimen was collected in Pennsylvania, and three specimens were collected in Washington, District of Columbia.

2.1.5 Supplementary populations used

In addition to the specimens collected in the four different countries, gene sequence data of pea weevils from two additional countries that were available on GenBank were downloaded. These sequences originated from China and Japan and were used as additional reference populations. This was done to increase the number of populations to which the Ethiopian population could be compared, in keeping with the objective of obtaining data from as many different pea weevil populations as possible in order to determine the source of the pea weevils in Ethiopia.

2.2 DNA extraction

2.2.1 Australian, Ethiopian and German specimens

DNA of the pea weevils from each location sampled was isolated using the Roche High Pure PCR Template Preparation Kit (Roche 2007). The protocol for isolation of nucleic acids from a mouse tail was followed.

- The whole pea weevil was placed in a 1.5ml micro-centrifuge tube. To the micro-centrifuge tube 200 µl Tissue Lysis Buffer and 40 µl Proteinase K was added.
- The pea weevil was then homogenised in the tube and incubated at 55°C overnight to digest the sample.
- After digestion 200µl Binding Buffer and 100µl isopropanol was added to the tube and the tube was vortexed.
- The tube was then centrifuged at 13 000 x g for five minutes to remove the insoluble debris.
- The supernatant was transferred to a High Pure Filter Tube in a Collection Tube and centrifuged at 8000 x g for one minute.
- 500µl of Inhibitor Removal Buffer was then added to the High Pure Filter Tube and centrifuged at 8000 x g for 1 minute.
- 500µl Wash Buffer was added to the spin column and centrifuged at 8000 x g for one minute. This step was repeated for a second time.
- The High Pure Filter Tube was then transferred to a clean 1.5ml micro-centrifuge tube and 50µl pre-heated (65°C) DNA Elution Buffer was added to the High Pure Filter Tube.
- It was then centrifuged at 8000 x g for one minute to elute the DNA.

2.2.2 American museum specimens

The museum specimens from the USA were made available to this study as a loan. Part of the loan agreement stated that the specimens could not be destroyed. For this reason a non-destructive DNA extraction method was needed for the museum specimens. This method was modified from the method described by Gilbert *et al.* (2007). The Roche High Pure PCR Template Preparation Kit was used for the DNA extractions (Roche, 2007).

- The whole pea weevil specimens were incubated in a mixture of 200µl Tissue Lysis Buffer and 40µl Proteinase K at 55 °C for 24 hours.
- The liquid was pipetted out and used for the DNA extraction.
- The specimens were then submerged in 100% EtOH for 4 hours to stop further digestion and air-dried for 72 hours.
- To the liquid product from the digestion, 200µl Binding Buffer was added and incubated for 10 min at 70 °C.
- Thereafter 100µl isopropanol was added to the mixture and the mixture was vortexed.
- The mixture was then transferred to a High Pure Filter Tube in a Collection Tube and centrifuged at 8000 x g for one minute.
- 500µl of Inhibitor Removal Buffer was then added to the High Pure Filter Tube and centrifuged at 8000 x g for 1 minute.
- Thereafter 500µl Wash Buffer was added to the spin column and centrifuged at 8000 x g for one minute. This step was repeated for a second time.
- The High Pure Filter Tube was then transferred to a clean 1.5ml micro-centrifuge tube and 25µl pre-heated (65°C) DNA Elution Buffer was added to the High Pure Filter Tube. It was then centrifuged at 8000 x g for one minute to elute the DNA.

2.3 DNA quantification

In order to determine the concentration of the extracted DNA, the isolated DNA was quantified on a NanoDrop ND1000 spectrophotometer. The NanoDrop spectrophotometer calculates the concentration of nucleic acids by measuring the absorbance value of light at a wavelength of 260nm. The absorbance value is then used to calculate the concentration of DNA in the sample by using the Beer-Lambert equation. The concentration of the DNA in the sample is presented in the measurement unit of ng/μl.

2.4 Genes chosen for analysis

In this study nucleotide sequences of three gene regions were used. Two mitochondrial genes were studied, specifically Cytochrome b (Cytb) and Cytochrome c oxidase subunit 1 (COX1). The COX1 and Cytb genes were chosen for the reason that reference sequences were available on GenBank and in view of the fact that Cytb and COX1 have been successfully used in previous phylogenetic studies of insects (Ito *et al.*, 2011). In addition to the mitochondrial genes, the nuclear protein-coding gene Elongation Factor 1-alpha (EF-1α) was sequenced to assist in the population study of the pea weevils currently in Ethiopia.

2.5 Primer design

Primers for the Cytb and COX1 gene regions were designed by using sequences of *B. pisorum* downloaded from GenBank. Since no EF-1α sequences of *B. pisorum* were available, EF-1α sequences of beetles closely related to *B. pisorum* were downloaded as a point of departure.

The program CLC Main Workbench (CLC bio 2005) was then used to analyse the sequences downloaded from GenBank. The CLC software was used to construct assembly sequences from which consensus sequences were constructed. Areas of consensus were identified in the sequences and these areas were used for primer design. After the primer design was completed, the primers were BLASTed against GenBank to test the specificity of the primers. The primer sets of the selected genes are listed in Table 2-1.

Table 2-1 Primer sets designed for the EF-1 α , COX1 and Cytb gene regions.

Elongation Factor 1-α	
Forward	5' CGTGGTATCACYATTGACATYGC 3'
Reverse	5' CAGCTTTTCCTTCYTTACGTTTC 3'
Cytochrome c oxidase subunit 1	
Forward	5' C TTCAGGATTTGGTATAATTTTC 3'
Reverse	5' GGAAGTTCAGAATAACTATG 3'
Cytochrome b	
Forward	5' GAGGTGCAACTGTTATTAC 3'
Reverse	5' CCTAATTTATTAGGAATAGATCG 3'

2.6 PCR (Polymerase Chain Reaction) amplification of gene regions

The polymerase chain reaction (PCR) is a method used to amplify a section of DNA. PCR uses multiple cycles of a 3-step process. The first step involves denaturation of a double-stranded DNA template. In the second step sequence-specific primers anneal to complementary sites. In the third step DNA polymerase extends the annealed primers in a 5'-3' direction, thereby copying the original DNA sequence. The PCR process uses the ability of DNA polymerase to synthesise a new strand of DNA that is complementary to the offered template strand.

The DNA polymerase enzyme can only add a nucleotide onto a pre-existing 3'-OH group. For this reason it needs a primer to which it can add the first nucleotide. This requirement makes it possible to specify the exact region of template sequence to amplify. When the PCR process has been completed the specific sequence will mount up in billions of copies.

The PCR process was applied to the extracted DNA from *B. pisorum*, using appropriate specific primers to amplify the three genes of interest. The same reaction conditions were used to amplify all three genes. Negative controls consisting of the PCR reaction mix excluding the DNA was also set up to detect DNA contamination. The reactions mix was set up in 0.2ml thin walled PCR tubes.

The PCR reactions mix consisted of the following components: 1µl each of the forward and reverse primers with an initial concentration of 10µM, 1 to 2µl of DNA (depending on the DNA concentration), 9.5 µl water and 12.5 µl EconoTaq® PLUS GREEN 2X Master Mix. The EconoTaq® PLUS GREEN 2X Master Mix is a ready-to-use PCR master mix also containing agarose gel loading buffer and tracking dyes. The Master Mix contains the following components: 0.1 units/µl of EconoTaq DNA Polymerase, reaction buffer (pH 9.0), 400µM dATP, 400µM dGTP, 400µM dCTP, 400µM dTTP, 3mM MgCl₂, and a proprietary mix of PCR Enhancer/Stabilizer and blue and yellow Tracking Dyes (Lucigen® Corporation, 2006).

The cycling was performed on the ABI9700 Thermo cycler that was programmed as follows: the first step was the initial denaturation at 95°C for 5 min. This was followed by 45 cycles each consisting of denaturing at 95°C for 30 seconds, annealing at 50°C for 30 seconds and elongation at 72°C for 60 seconds. This was followed by the final elongation step at 72°C for 10 min.

To determine if amplicons were present, 2µl of the PCR product was loaded on a 1% agarose gel, pre-stained with EtBr (ethidium bromide) and ran for 15 minutes at 100V and 350 mA. The gel was then visualised under UV light.

2.7 Sequencing of gene regions

DNA sequencing is the determination of the precise sequence of nucleotides in a sample of DNA. Various methods and technologies can be used to do this. The classical chain-termination method (Sanger *et al.*, 1977) requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleotidetriphosphates (dNTPs), and modified nucleotides (dideoxynucleotides) that terminate DNA strand elongation. Dye-terminator sequencing makes use of labelling of the chain terminator ddNTPs, which allows sequencing in a single reaction. In dye-terminator sequencing, each of the four dideoxynucleotide chain terminators is labelled with fluorescent dyes which allows for the sequence to be analysed by use of a capillary electrophoresis-based genetic analyser (Montesino and Prieto, 2012). For the current study the dye-terminator sequencing method was used.

In the current study, the amplified DNA was sequenced in one direction as follows: firstly the PCR products were purified by using the Exo/SAP Amplicon Purification protocol (Werle *et al.*, 1994). The purification reaction was set up with 10µl PCR product, 0.5µl Exonuclease1 and 2µl FastAP™ Thermosensitive Alkaline Phosphatase which was mixed in a tube and incubated at 37°C for 15 minutes. The reaction was then stopped by heating the mixture to 85°C for 15 minutes. This procedure removes the contaminating primers and dNTPs. Thereafter the purified amplicons were used for sequencing. The sequencing reactions were performed using ABI Big dye V3.1 and analysed on an ABI 3500XL genetic analyser.

2.8 Supplementary sequences

In addition to the sequences obtained from the collected *B. pisorum* specimens, supplementary sequences were downloaded from GenBank. All Cytb and COX1 sequences available for *B. pisorum* on GenBank were downloaded. In total nine Cytb and six COX1 sequences of *B. pisorum* were available. The downloaded sequences of *B. pisorum* with their GenBank access numbers and origins are listed in Table 2-2.

For the current study sequence data of the Cytb, COX1 and EF-1 α genes of an outgroup was also needed. For that reason Cytb, COX1 and EF-1 α sequences of *Bruchus rufimanus* available on GenBank were downloaded. The downloaded sequences are listed in Table 2-3 (with GenBank access numbers).

Table 2-2 List of *B. pisorum* sequences downloaded from GenBank

Abbreviation assigned	GenBank access no	Sequence origin	Gene name	Sequence length
Ch1	EF570096.1	China	COX1	803 bp
Ch2	EF570095.1	China	COX1	803 bp
Ch3	EF570094.1	China	COX1	803 bp
Ch4	EF484373.1	China	COX1	803 bp
Ch5	EF484372.1	China	COX1	803 bp
Ch6	EF570097.1	China	COX1	803 bp
Ch7	EF570109.1	China	Cytb	463 bp
Ch8	EF570108.1	China	Cytb	463 bp
Ch9	EF570107.1	China	Cytb	463 bp
Ch10	EF484395.1	China	Cytb	463 bp
Ch11	EF484396.1	China	Cytb	463 bp
Ch12	EF484394.1	China	Cytb	463 bp
Ch13	EF484393.1	China	Cytb	463 bp
Ch14	EF484392.1	China	Cytb	463 bp
Jap1	DQ307672.1	Japan	Cytb	471 bp

Table 2-3 List of *B. rufimanus* sequences downloaded from GenBank

Abbreviation assigned	GenBank access no	Sequence origin	Gene name	Sequence length
BR1	DQ155987.1	Britain	COX1	820 bp
BR2	EF570084.1	China	COX1	822 bp
BR3	EF570083.1	China	COX1	822 bp
BR4	EF570082.1	China	COX1	822 bp
BR5	EF570081.1	China	COX1	822 bp
BR6	EF570079.1	China	COX1	822 bp
BR7	EF570078.1	China	COX1	822 bp
BR8	EF484354.1	China	COX1	822 bp
BR9	EF484353.1	China	COX1	822 bp
BR10	EF484352.1	China	COX1	822 bp
BR11	EF484378.1	China	Cytb	463 bp
BR12	EF484377.1	China	Cytb	463 bp
BR13	EF570101.1	China	Cytb	463 bp
BR14	EF570100.1	China	Cytb	463 bp
BR15	EF570099.1	China	Cytb	463 bp
BR16	EF570098.1	China	Cytb	463 bp
BR17	EF484378.1	China	Cytb	463 bp
BR18	EF484377.1	China	Cytb	463 bp
BR19	EF484376.1	China	Cytb	463 bp
BR20	EF484375.1	China	Cytb	463 bp
BR21	EF484374.1	China	Cytb	463 bp
BR22	AY997367.1	USA	EF-1 α	556 bp

2.9 Data analysis

2.9.1 Sequence analysis

The DNA sequence of each gene was first screened to evaluate the quality and determine the length of the sequences. All the sequences were edited to remove low quality sequences using the Mega5 (Tamura *et al.*, 2011) software. The sequences of each gene were then aligned by using ClustalW (Higgins *et al.*, 1994) in the program Mega5.

ClustalW is an extensively used system for aligning any number of homologous nucleotide or protein sequences. The ClustalW approach uses progressive alignment methods for multi-sequence alignments. In these methods, the most similar sequences, that is to say the sequences with the best alignment score, are aligned first. Thereafter increasingly more distant groups of sequences are aligned until a global alignment is obtained. This heuristic approach is necessary because finding the global optimal solution is costly in both memory and time requirements. The algorithm used by ClustalW starts by computing a rough distance matrix between each pair of sequences based on pairwise sequence alignment scores. Following the assignment of pairwise alignment scores, the algorithm uses the neighbour-joining method with midpoint rooting to create a guide tree; which is used to generate a global alignment (Higgins *et al.*, 1994).

After alignment sequences with a comparatively short length were removed from the alignments. The sequences downloaded from GenBank were then added to the alignments and the sequences were realigned using ClustalW in Mega5.

The mitochondrial gene sequences were then evaluated to identify distinct sequences, henceforth referred to as haplotypes, for each gene in each population. This was done by using the software Arlequin ver. 3.5 (Excoffier and Lischer, 2010). Arlequin was configured to estimate the haplotype frequencies and to search for shared haplotypes across the populations.

The nuclear gene sequences were treated as genotypes. The sequences were analysed to determine the number of unique genotypes in each population. Sequences were evaluated and the nuclear DNA nucleotide positions that showed perfect double peaks were scored as heterozygotes.

All the sequences were analysed with Mega5 to identify the number of conserved, variable, parsimony-informative, singleton, 0-fold, 2-fold and 4-fold degenerate sites in the sequences. Conserved or constant sites are sites that contain the same nucleotide or amino acid in all sequences. A site is labelled as a constant site if at least two sequences contain unambiguous nucleotides or amino acids (Tamura *et al.*, 2011). Variable sites are sites that contain at least two types of nucleotides or amino acids. Variable sites can be singleton or parsimony-informative (Tamura *et al.*, 2011). Parsimony-informative sites are sites that contain at least two types of nucleotides or amino acids, and at least two of them occur with a minimum frequency of two (Tamura *et al.*, 2011). Sites that contain at least two types of nucleotides or amino acids, with one occurring multiple times, are singleton sites. A site is identified as a singleton site if at least three sequences contain unambiguous nucleotides or amino acids (Tamura *et al.*, 2011).

A site where all the changes are nonsynonymous is a 0-fold degenerate site (Tamura *et al.*, 2011). Sites where one out of three changes is synonymous is 2-fold degenerate sites. All sites at which two out of three changes are synonymous are also included in this category (Tamura *et al.*, 2011). Sites in which all changes are synonymous are called 4-fold degenerate sites (Tamura *et al.*, 2011).

Changes are nonsynonymous when a nucleotide change results in changes to amino acid encoded by the original codon. A nucleotide site in which one or more changes are nonsynonymous is referred to as a nonsynonymous site. If only one of three possible nucleotide changes at that site is nonsynonymous, then the site is 1/3 nonsynonymous. If two of three nucleotide changes are nonsynonymous, then the site is 2/3 nonsynonymous. Finally, if all three possible nucleotide changes are nonsynonymous, then the site is completely nonsynonymous (Tamura *et al.*, 2011).

A synonymous change is when a nucleotide change does not change the amino acid encoded by the original codon. A nucleotide site in which one or more changes is synonymous is referred to as a synonymous site. If only one of three possible nucleotide changes at that site is synonymous, then the site is 1/3 synonymous. If two of three nucleotide changes are synonymous, then the site is 2/3 synonymous, and if all three possible nucleotide changes are synonymous, then the site is completely synonymous (Tamura *et al.*, 2011).

2.9.2 Tree building

The DNA sequences of each gene were used to construct phylogenetic trees. To select the best substitution model, the FindModel program (available online at <http://hcv.lanl.gov/content/sequence/findmodel/findmodel.html>) was used. FindModel selected the Jukes-Cantor model as the best substitution model for all the sequenced genes, namely EF-1 α , Cytb and COX1.

The Jukes-Cantor model is a model in which the rate of nucleotide substitution is the same for all pairs of the four nucleotides A, T, C, and G. The multiple hit correction equation used for the Jukes-Cantor model produces a maximum likelihood estimate of the number of nucleotide substitutions between two sequences. The Jukes-Cantor model assumes firstly that the substitution rates among sites are equal and secondly that the

nucleotide frequencies are equal. The Jukes-Cantor model does not correct for a higher rate of transitional substitutions as compared to transversional substitutions (Jukes and Cantor, 1969).

The distinct haplotypes identified in the mtDNA gene sequences in each population were used to construct phylogenetic trees of the Cytb and COX1 sequence data. Distinct genotypes of the nuclear gene, EF-1 α in each population were used to construct the EF-1 α phylogenetic tree. Gene sequences of *B. rufimanus* was used as the outgroup in all these trees.

To test all possibilities and alternative hypotheses of evolution, the Mega5 software was used to construct phylogenetic trees by using the neighbour joining (NJ) (Saitou and Nei, 1987), unweighted pair-group arithmetic mean (UPGMA) (Nei and Kumar, 2000), minimum evolution (ME) (Rzhetsky and Nei, 1993), maximum likelihood (ML) (Felsenstein, 1981) and maximum parsimony (MP) (Nei and Kumar, 2000) methods. For all the trees constructed the Jukes-Cantor (Jukes and Cantor, 1969) model was used and tested with 1000 Bootstrap replicas (Felsenstein, 1985).

The minimum evolution method of phylogenetic inference is based on the assumption that the tree with the smallest sum of branch length estimates is most likely to be the true one. The ME method uses distance measures that correct for multiple hits at the same sites. A topology showing the smallest value of the sum of all branches (S) is chosen as an estimate of the correct tree. The construction of a minimum evolution tree is time-consuming since the S values for all topologies must be evaluated. The number of possible topologies (un-rooted trees) quickly increases with the number of taxa. The phylogenetic tree is an un-rooted tree but, for ease of inspection, it is frequently displayed in a style similar to that of rooted trees (Rzhetsky and Nei, 1993; Tamura *et al.*, 2011).

The neighbour joining method is a distance-based method. NJ uses a distance matrix to calculate the tree. The NJ method is a simplified version of the ME method. It chooses a topology showing the smallest value of the sum of all branches (S) as an estimate of the correct tree. However, unlike ME, the S value is not computed for all topologies, but the examination of different topologies is embedded in the algorithm, so that only one final tree is produced (Saitou and Nei, 1987; Tamura *et al.*, 2011).

The Unweighted Pair Group Method with Arithmetic method is a hierarchical clustering method. This method assumes that the rate of nucleotide or amino acid substitution is the same for all evolutionary lineages. Since the assumption of a constant rate of evolution is made, this method produces a rooted tree (Sneath and Sokal, 1973; Tamura *et al.*, 2011).

The maximum parsimony method was originally developed for morphological characteristics. To construct an MP tree, only parsimony-informative sites are used. The approach in *MEGA* software estimates MP tree branch lengths by using the average pathway method for un-rooted trees. For a given topology, the sum of the minimum possible substitutions over all sites is known as the Tree Length. The topology with the minimum tree length is known as the Maximum Parsimony tree (Eck and Dayhoff, 1966; Fitch, 1971; Tamura *et al.*, 2011).

The maximum likelihood method estimates a tree for which the observed data is most probable. Firstly an initial tree is built using a fast but suboptimal method such as Neighbour Joining. Its branch lengths are then adjusted to maximise the likelihood of the data set for that tree topology under the desired model of evolution. Thereafter variants of the topology are created using the NNI (nearest neighbour interchange) method to search for topologies that match the data better. Maximum likelihood branch lengths are then computed for these variant tree topologies and the greatest likelihood observed is reserved as the best selection up to that point. This search

continues until no greater likelihoods are found (Felsenstein, 1981; Tamura *et al.*, 2011).

The MrBayes3.2 software (Huelsenbeck *et al.*, 2001) (Ronquist and Huelsenbeck, 2003) was used to construct trees by use of the Bayesian analysis of phylogeny (BAP) and the trees were visualised on the Figtree v1.3.1 software (Rambaut, 2009). MrBayes was configured as follows: the nucleotide model was set to the 4by4 model and the variation rates across all sites were set to equal. To determine the posterior probability values, the Markov Chain Monte Carlo (MCMC) method in MrBayes was used and set to run for 1 million generations and to sample trees every 100 generations.

The Bayesian analysis of phylogeny is build upon a likelihood foundation. Specifically, it is based on a quantity called the posterior probability of a tree. Bayes' theorem is used to combine the prior probability of a phylogeny ($\text{Pr}[\text{Tree}]$) with the likelihood ($\text{Pr}[\text{Data}|\text{Tree}]$) to produce a posterior probability distribution on trees ($\text{Pr}[\text{Tree}|\text{Data}]$) (Williams and Bernard 2003).

$$\text{Pr}[\text{Tree}|\text{Data}] = \frac{\text{Pr}[\text{Data}|\text{Tree}] \times \text{Pr}[\text{Tree}]}{\text{Pr}[\text{Data}]}$$

The posterior probability represents the probability that the tree is correct. The tree with the highest posterior probability can be selected as the best approximation of phylogeny. All trees are considered as equally probable, and likelihood is calculated using a substitution model of evolution. To compute the posterior probability a summary must be calculated over all trees, and all possible combinations of branch length and substitution model parameter values must be summarised for each tree. The MCMC method is the most frequently used method to estimate the posterior probability (Williams and Bernard, 2003).

2.9.3 Haplotype networks

The software Network version 4.6 (Free Phylogenetic Network Software 1999-2010) was used to construct minimum spanning networks between haplotypes of *B. pisorum*. One haplotype of *B. rufimanus* was also included in each network. Networks were constructed for each of the mitochondrial genes, namely Cytb and COX1. Median Joining (MJ) (Bandelt, *et al.* 1999) was used to calculate the network.

Haplotype networks allow for the visualisation of the relationships between sequences. Haplotype networks are phylogenetic networks and can be constructed using several different approaches such as the union of maximum parsimony trees (UMP) (Cassens *et al.*, 2005), statistical parsimony (SP) (Templeton *et al.*, 1992), split decomposition (SD) (Huson, 1998), Neighbor-Net (NN) (Huson and Bryant, 2006), unreduced median networks (MED) (Huson and Bryant, 2006), reduced median-joining (RMD) (Bandelt *et al.*, 1995), minimum-spanning network (MSN) (Excoffier and Smouse, 1994) and Median Joining (MJ) (Bandelt, *et al.* 1999).

A minimum spanning tree is a tree which connects a set of sequence types without creating any cycles of inferring additional nodes, so that the total length is minimal (Bandelt *et al.*, 1999). A modification to the algorithm used to construct minimum spanning trees allows for the construction of a union of all minimum spanning trees. This union is called a minimum spanning network (Bandelt *et al.*, 1999). A minimum spanning network is typically not useful for the direct representation of genetic data, because a minimum spanning tree is not the most parsimonious (Bandelt *et al.*, 1999). Minimum spanning networks, nevertheless, serve as good points of departure in each recursive step of the MJ network construction for generating additional inferred sequence types which reduce tree lengths.

The Median-Joining method begins by combining minimum spanning trees into a single network (Bandelt *et al.*, 1999; Posada and Crandall, 2001). Thereafter median vectors, representing missing intermediates, are added to the network using parsimony criteria (Posada and Crandall, 2001). The Median-Joining method is a fast method and can handle large data sets and multistate characters (Posada and Crandall, 2001).

2.9.4 Population diversity

Population diversity is the amount of variation between individuals in a given population. This variation may be measured in terms of genetic or morphological variation. Different genetic variation measures are used to determine the amount of diversity in a population. These measures include the number of polymorphic sites, nucleotide diversity (π), haplotype diversity (h) and the mean number of pairwise differences.

Nucleotide diversity (π) is used to measure the degree of polymorphism within a population (Nei and Li, 1979). The nucleotide diversity measure is defined as the average number of nucleotide differences per site between any two DNA sequences chosen randomly from the sample population. Haplotype diversity reflects the number of unique haplotypes in a population and is derived from the haplotype frequencies in a given population. The mean number of pairwise differences is the number of nucleotide differences between pairs sampled in a population. The number of polymorphic sites is the number of sites in the DNA sequence that shows variation between sequences.

The population diversity was computed using Arlequin version3.5 software (Excoffier and Lischer, 2010). Arlequin was used to compute the genetic diversity within all of the studied populations. Four diversity measures were calculated for each gene in each population, namely the number of

polymorphic sites, nucleotide diversity (π), haplotype or genotype diversity (h) and the mean number of pairwise differences. The genetic diversity within the subpopulations in Ethiopia was also calculated using the same strategy. A list of the subpopulations into which the Ethiopian population is divided can be found in Table 3-2.

2.9.5 Population differentiation

Population differentiation is measured using the F_{ST} index. The F_{ST} index was developed by Sewall Wright (Wright, 1951). The lower the F_{ST} value, the smaller the differentiation between the pair of populations. Associated p values are calculated to test the significance of the F_{ST} values.

Arlequin software was also used to compute the population differentiation between population pairs by calculating pairwise F_{ST} and associated p values. A significance level of 0.05 was used.

3 Results

3.1 DNA extraction and quantification

In the Ethiopian population, DNA was extracted from a total of 74 specimens. The nucleic acid concentrations of the DNA extracted from these insects were in the range of 10-132.8 ng/μl. The number of specimens from which DNA was extracted in each population is listed in Table 3-1. Also listed here are the sample ID ranges for the specimens in each population and the minimum and maximum DNA yield obtained in each population.

Table 3-1 Nucleic acid concentration range of DNA extracted from specimens of different populations

Population	No. of specimens	Sample ID range	Minimum concentration	Maximum concentration
Ethiopia Geboya	15	Et1A to Et1O	15.9 ng/μl	129.5 ng/μl
Ethiopia Jiman Derega	15	Et2A to Et2O	10 ng/μl	104.5 ng/μl
Ethiopia Adet Hanna	14	Et3A to Et3N	13.2 ng/μl	132.8 ng/μl
Ethiopia Selamaya	15	Et4A to Et4O	11.4 ng/μl	104.9 ng/μl
Ethiopia Qualias	15	Et5A to Et5O	20.9 ng/μl	129.9 ng/μl
Australia	10	AsA to AsJ	56.2 ng/μl	172.6 ng/μl
Germany	6	Ger1 to Ger6	169.6 ng/μl	206 ng/μl
USA	21	USA1 to USA21	13.1 ng/μl	29.7 ng/μl

DNA was successfully extracted from 21 museum specimens from the USA by using a modified non-destructive DNA extraction method. After the DNA extraction was completed, it was confirmed that specimens sustained minimal morphological damage. The nucleic acid concentration of the DNA extracted from the museum specimens were in the range of 13.1-29.7ng/μl. The nucleic acid yield from each museum specimen as well as the date the specimen was collected is listed in Table 3-2.

Table 3-2 Nucleic acid concentrations of DNA extracted from USA museum specimens

Date collected	Sample ID	Nucleic acid concentration
30/07/1938	USA1	18 ng/μl
30/07/1938	USA2	21.1 ng/μl
30/08/1938	USA3	21.9 ng/μl
30/08/1938	USA4	19.4 ng/μl
30/08/1938	USA5	17.8 ng/μl
30/08/1938	USA6	20.3 ng/μl
30/08/1938	USA7	13.1 ng/μl
10/1929	USA8	22.4 ng/μl
10/1929	USA9	29.7 ng/μl
10/1929	USA10	20.3 ng/μl
10/1929	USA11	17.0 ng/μl
10/1929	USA12	21.7 ng/μl
10/1929	USA13	19.7 ng/μl
5/02/1931	USA14	20.5 ng/μl
11/1923	USA15	21.2 ng/μl
unknown	USA16	17.3 ng/μl
1913	USA17	17.5 ng/μl
unknown	USA18	13.5 ng/μl
13/07/1923	USA19	21.5 ng/μl
13/07/1923	USA20	15.4 ng/μl
13/07/1923	USA21	16.7 ng/μl

3.2 DNA amplification and sequencing

The extracted DNA was amplified and sequenced for the COX1, Cytb and EF-1 α genes. The expected maximum sequence length of these genes according to sequences on GenBank, were as follows: COX = 803bp, Cytb = 469bp and EF-1 α = 556bp.

In the Ethiopian population, including all subpopulations in Ethiopia, a total of 62 COX1, 73 Cytb and 68 EF-1 α gene sequences were obtained. All three gene regions were successfully amplified in 60 individual specimens. The Cytb and COX1 gene regions were both successfully amplified in 61 individual specimens. In 67 individual specimens the Cytb and EF-1 α gene regions were successfully amplified and the COX1 and EF-1 α gene regions were both successfully amplified in 61 individual specimens.

Of the 21 DNA samples extracted from the USA museum specimens, DNA was successfully amplified and sequenced to produce 11 gene sequences. These 11 sequences were obtained from eight individual specimens. The COX1 gene was successfully amplified and sequenced for two individual specimens namely USA8 and USA11. The Cytb and EF-1 α genes were both successfully amplified and sequenced in three individual specimens namely USA13, USA14 and USA17. The Cytb gene was successfully amplified and sequenced for the individual USA20, and for individual specimens USA15 and USA2, the EF-1 α gene was successfully amplified and sequenced.

Across all the Australian, Ethiopian, German and USA populations sampled a total of 80 COX1, 93 Cytb and 89 EF-1 α gene sequences were obtained. The number of sequences that originated in each population for each gene is listed in Table 3-3. The maximum and minimum sequence lengths observed for each gene in each population are also listed.

Table 3-3 The number of sequences obtained for the COX1, Cytb and EF-1 α genes, as well as the minimum and maximum sequence lengths in the individual populations.

Population	COX1			Cytb			EF-1 α		
	no	Minimum length	Maximum length	no	Minimum length	Maximum length	no	Minimum length	Maximum length
Ethiopia Geboya	15	700 bp	711 bp	13	369 bp	387 bp	15	388 bp	400 bp
Ethiopia Jiman Derega	13	645 bp	715 bp	15	369 bp	387 bp	14	388 bp	400 bp
Ethiopia Adet Hanna	13	699 bp	709 bp	14	369 bp	387 bp	14	388 bp	400bp
Ethiopia Selamaya	7	700 bp	709 bp	15	369 bp	387 bp	10	388 bp	399 bp
Ethiopia Qualias	14	686 bp	709 bp	15	369 bp	381 bp	15	391 bp	400 bp
Australia	10	709 bp	726 bp	10	307 bp	377 bp	10	385 bp	388 bp
Germany	6	703 bp	713 bp	6	368 bp	381 bp	6	407 bp	409 bp
USA	2	718 bp	719 bp	4	307 bp	316 bp	5	373 bp	397 bp

3.3 Sequence analysis

The COX1, Cytb and EF-1 α sequences obtained from the sampled populations of *B. pisorum* were aligned with the *B. pisorum* sequences downloaded from GenBank. In total six COX1 and nine Cytb sequences of *B. pisorum* were downloaded from GenBank. Refer to Table 2-2 for a list of the downloaded *B. pisorum* sequences that include the GenBank accession

numbers, sequence lengths and abbreviations used for each sequence in this study.

After aligning the sequences and removing low quality and comparatively short sequences, the COX1 sequence alignment consisted of 84 sequences, the Cytb sequence alignment of 102 sequences and the EF-1 α alignment of 89 sequences. The COX1 sequence alignment displayed the longest final length of 642 bp followed by the EF-1 α sequence alignment with a length of 303 bp and Cytb sequence alignment with a length of 251bp.

The sequences were analysed to determine the degree of sequence variation. The COX1 gene displayed the highest number of Parsinomic informative sites, with 24 sites. The Cytb and EF-1 α genes possessed the same number of Parsinomic informative sites with 12 sites each. A summary of the overall variation observed in the COX1, Cytb and EF-1 α sequences is listed in Table 3-4, along with the final sequence lengths and the number of haplotypes or genotypes identified in the alignments.

In the COX1 sequence alignment, nine distinct haplotypes were identified. Of these haplotypes, no single haplotype was present in all five populations. Three haplotypes were present in at least two populations. The remaining six haplotypes were only present in single populations. The haplotypes identified in the COX1 sequence alignment are listed in Table 3-5, together with the haplotype frequency in each population as well as the sample name of a representative sequence of the haplotype in the given population.

Ten Cytb haplotypes were identified in the final Cytb sequence alignment. Of these, six haplotypes were present in single populations only. Two haplotypes were present in two populations and two were present in three populations. The 10 haplotypes identified are listed in Table 3-6, together with the haplotype frequency in each population as well as the sample name of the representative sequence of the haplotype in the given population.

A total of 70 EF-1 α genotypes were identified in the EF-1 α sequence alignment. Of these, only one genotype was present in two populations, with the remaining 69 genotypes present in single populations only. The genotypes identified in the EF-1 α sequence alignment are listed in Table 3-8, together with the genotype frequency in each population as well as the sample name of the representative sequence of the genotype in the given population.

Table 3-4 The variation observed in the COX1, Cytb and EF-1 α sequence alignments of *B. pisorum* from Ethiopia, Australia, Germany, the USA, China and Japan

	COX1	Cytb	EF-1 α
Alignment length	642 bp	251 bp	303 bp
No. haplotypes	9	10	70 (genotypes)
No. conserved sites	613	236	288
No. variable sites	29	15	15
No. parsinomic informative sites	24	12	12
No. singleton sites	5	3	3
0-fold degenerate sites	410	152	205
2-fold degenerate sites	160	70	57
4-fold degenerate sites	42	18	33

Table 3-5 COX1 haplotype frequencies in the different populations

Haplotype name, GenBank access number, haplotype frequencies (in brackets) and the representative sequence in each population for the COX1 sequences data						
Haplotype name	GenBank access number	Ethiopia	Australia	China	Germany	USA
COX-1	JQ434491	-	(0.9) AsJ	-	-	-
COX-2	JQ434490	-	(0.1) As B	(0.333) Ch3	-	-
COX-3	EF570097.1	-	-	(0.667) Ch6	-	-
COX-4	JQ434485	(0.017) Et1A	-	-	-	-
COX-5	JQ434489	(0.483) Et5O	-	-	-	(0.5) USA8
COX-6	JQ434487	(0.45) Et5k	-	-	-	(0.5) USA11
COX-7	JQ434488	(0.033) Et5L	-	-	-	-
COX-8	JQ434486	(0.017) Et4B	-	-	-	-
COX-9	JQ434492	-	-	-	(1) Ger6	-

Table 3-6 Cytb haplotype frequencies in the different populations

Haplotype name, GenBank access number, haplotype frequencies (in brackets) and the representative sequence in each population for the Cytb sequences data							
Haplotype name	GenBank access number	Ethiopia	Australia	China	Germany	Japan	USA
CB1	JQ478759	(0.822) Et5M	(0.600)AsC	-	-	(1) Jap1	-
CB2	JQ478758	(0.110) Et5I	-	-	-	-	-
CB3	JQ478760	(0.069) Et5O	(0.300)AsA	-	-	-	-
CB4	JQ478757	-	(0.100)AsB	(0.125) Ch9	-	-	-
CB5	JQ478765	-	-	-	-	-	(0.25) USA20
CB6	JQ478761	-	-	(0.625) Ch10	(0.833)Ger5	-	(0.25) USA17
CB7	EF570108.1	-	-	(0.25) Ch8	-	-	-
CB8	JQ478762	-	-	-	(0.167)Ger6	-	-
CB9	JQ478763	-	-	-	-	-	(0.25) USA13
CB10	JQ478764	-	-	-	-	-	(0.25) USA14

Table 3-7 EF-1 α genotype occurrences in the different populations

Genotype name, GenBank access number, Genotype frequencies (in brackets) and the representative sequence in each population for the EF-1 α sequences data					
Genotype name	GenBank access number	Ethiopia	Australia	Germany	USA
EF1	JQ478822	Et 1A(0.015)	-	-	-
EF2	JQ478823	Et1B (0.015)	-	-	-
EF3	JQ478766	Et1C (0.015)	-	-	-
EF4	JQ478824	Et1D (0.015)	-	-	-
EF5	JQ478781	Et2L (0.029)	-	-	-
EF6	JQ478767	Et1F (0.015)	-	-	-
EF7	JQ478768	Et1G (0.015)	-	-	-
EF8	JQ478769	Et1H (0.015)	-	-	-
EF9	JQ478770	Et1I (0.015)	-	-	-
EF10	JQ478771	Et1J (0.015)	-	-	-
EF11	JQ478772	Et1K (0.015)	-	-	-
EF12	JQ478773	Et1L (0.015)	-	-	-
EF13	JQ478812	Et5I (0.029)	-	-	-
EF14	JQ478800	Et4J (0.029)	-	-	-
EF15	JQ478774	Et1O (0.015)	-	-	-
EF16	JQ478792	Et3I (0.029)	-	-	-
EF17	JQ478813	Et5J (0.059)	-	-	-
EF18	JQ478791	Et3G (0.029)	-	-	-
EF19	JQ478775	Et2E (0.015)	-	-	-
EF20	JQ478776	Et2F (0.015)	-	-	-
EF21	JQ478777	Et2G (0.015)	AsC (0.1)	-	-
EF22	JQ478789	Et3E (0.029)	-	-	-
EF23	JQ478778	Et2I (0.015)	-	-	-
EF24	JQ478779	Et2J (0.015)	-	-	-
EF25	JQ478780	Et2K (0.015)	-	-	-
EF26	JQ478782	Et2M (0.015)	-	-	-
EF27	JQ478783	Et2N (0.015)	-	-	-
EF28	JQ478784	Et2O (0.015)	-	-	-
EF29	JQ478785	Et3A (0.015)	-	-	-
EF30	JQ478786	Et3B (0.015)	-	-	-
EF31	JQ478787	Et3C (0.015)	-	-	-
EF32	JQ478788	Et3D (0.015)	-	-	-
EF33	JQ478790	Et3F (0.015)	-	-	-
EF34	JQ478797	Et4A (0.029)	-	-	-
EF35	JQ478793	Et3J (0.015)	-	-	-
EF36	JQ478794	Et3K (0.015)	-	-	-
EF37	JQ478796	Et3N (0.015)	-	-	-

Genotype name	GenBank access number	Ethiopia	Australia	Germany	USA
EF38	JQ478798	Et4B (0.015)	-	-	-
EF39	JQ478799	Et4D (0.015)	-	-	-
EF40	JQ478801	Et4K (0.015)	-	-	-
EF41	JQ478802	Et4L (0.015)	-	-	-
EF42	JQ478803	Et4M (0.015)	-	-	-
EF43	JQ478818	Et5O (0.029)	-	-	-
EF44	JQ478804	Et5A (0.015)	-	-	-
EF45	JQ478805	Et5B (0.015)	-	-	-
EF46	JQ478806	Et5C (0.015)	-	-	-
EF47	JQ478807	Et5D (0.015)	-	-	-
EF48	JQ478808	Et5E (0.015)	-	-	-
EF49	JQ478809	Et5F (0.015)	-	-	-
EF50	JQ478810	Et5G (0.015)	-	-	-
EF51	JQ478811	Et5H (0.015)	-	-	-
EF52	JQ478814	Et5K (0.015)	-	-	-
EF53	JQ478815	Et5L (0.015)	-	-	-
EF54	JQ478816	Et5M (0.015)	-	-	-
EF55	JQ478817	Et5N (0.015)	-	-	-
EF56	JQ478819	-	AsG (0.6)	-	-
EF57	JQ478820	-	AsH(0.1)	-	-
EF58	JQ478821	-	AsJ (0.2)	-	-
EF59	JQ478825	-	-	Ger1(0.167)	-
EF60	JQ478826	-	-	Ger2(0.167)	-
EF61	JQ478827	-	-	Ger3(0.167)	-
EF62	JQ478828	-	-	Ger4(0.167)	-
EF63	JQ478829	-	-	Ger5(0.167)	-
EF64	JQ478830	-	-	Ger6(0.167)	-
EF65	JQ478831	-	-	-	USA13 (0.2)
EF66	JQ478795	Et3L (0.015)	-	-	-
EF67	JQ478832	-	-	-	USA14 (0.2)
EF68	JQ478833	-	-	-	USA15 (0.2)
EF69	JQ478834	-	-	-	USA17 (0.2)
EF70	JQ478835	-	-	-	USA21 (0.2)

3.4 Phylogenetic trees

Phylogenetic trees were constructed for each of the studied genes. For each tree, sequences of *B. rufimanus* that were downloaded from GenBank were used as an outgroup. Sequences of *B. rufimanus* downloaded from GenBank are listed in Table 2-3, together with sequence length and GenBank access numbers.

3.4.1 Phylogenetic trees constructed with COX1 sequences

To construct trees only sequences representative of each haplotype in each population were used. The sample IDs of the haplotype representative in each population are listed in Table 3-6.

The phylogenetic trees constructed by NJ, ME, UPGMA and MP methods are presented in Figures 3-1 to 3-4. The evolutionary analyses for all trees were conducted on Mega 5 (Tamura *et al.*, 2011). A bootstrap test with 1000 replicas was performed on each tree. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. All four trees displayed highly similar structures. Differences in taxon resolution on branches can be explained by the variation of bootstrap values among the trees, causing the collapse of some branches.

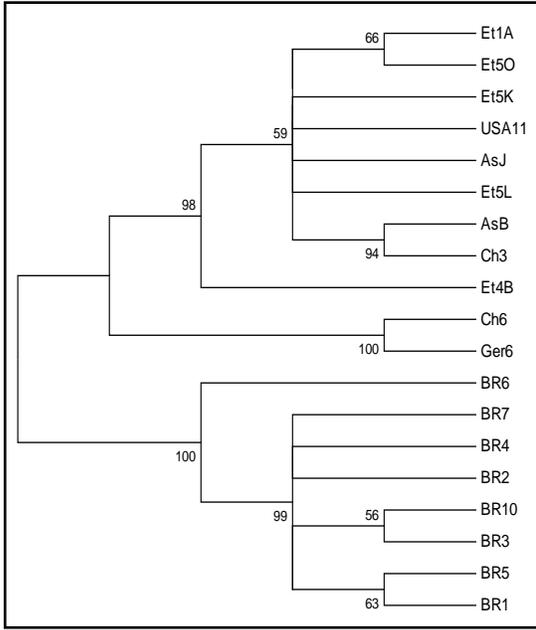


Figure 3-1 Evolutionary relationships of COX1 sequences of *B. pisorum* populations inferred using the Neighbour-Joining method.

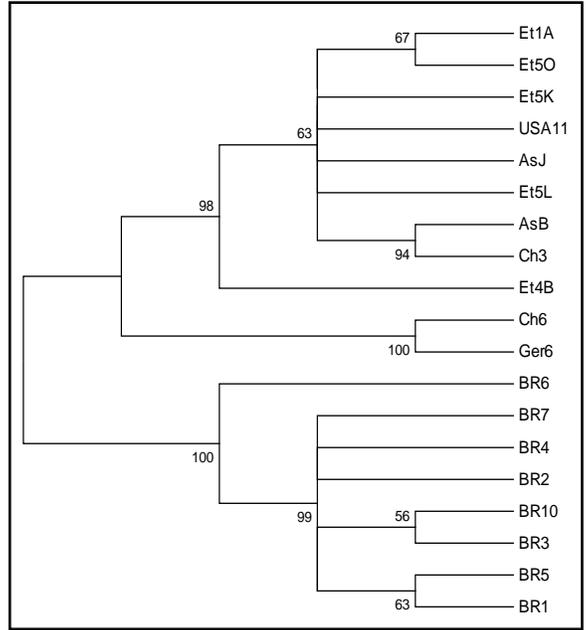


Figure 3-2 Evolutionary relationships of COX1 sequences of *B. pisorum* populations inferred using the Minimum Evolution method.

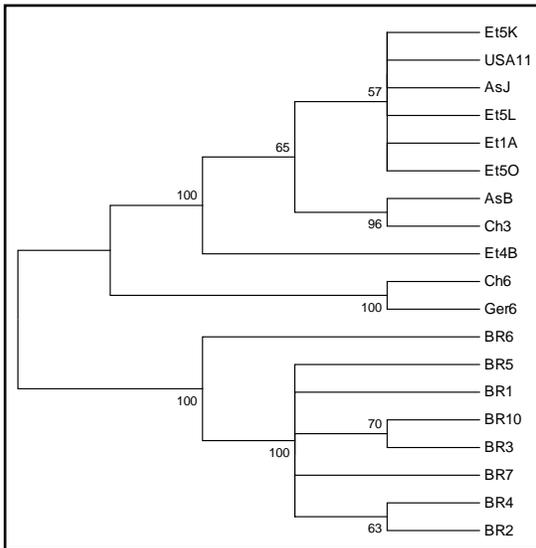


Figure 3-3 Evolutionary relationships of COX1 sequences of *B. pisorum* populations inferred using the UPGMA method.

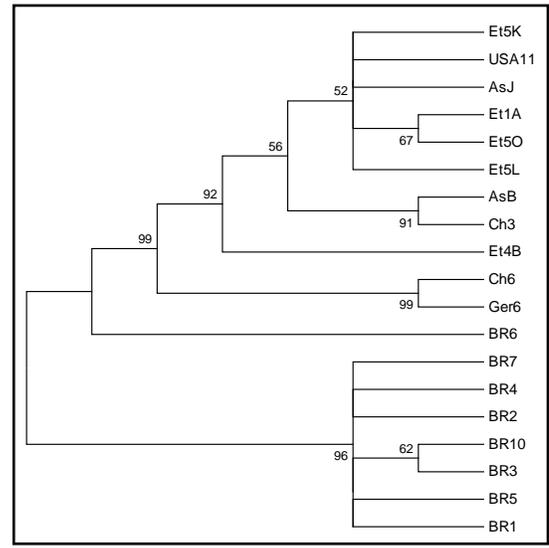


Figure 3-4 Maximum Parsimony analysis of COX1 sequences of *B. pisorum* populations.

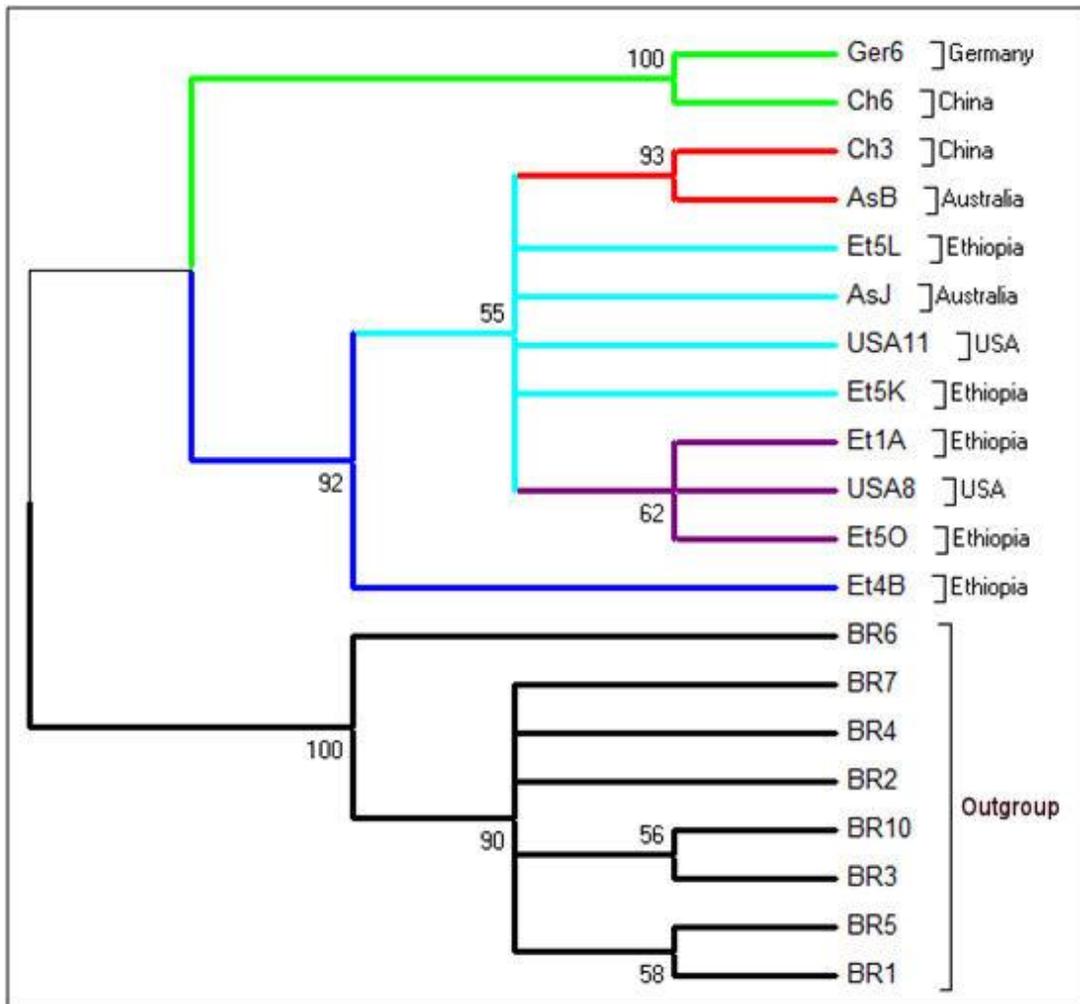


Figure 3-5 Molecular phylogenetic analysis by Maximum Likelihood using COX1 sequences

The evolutionary history was inferred by using the Maximum Likelihood method based on the Jukes-Cantor model (Jukes and Cantor, 1969). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically as follows: When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used, otherwise BIONJ method with MCL distance matrix was used. The analysis involved 20 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 642 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura, *et al.* 2011).

Figure 3-5 shows a phylogenetic tree constructed by Maximum Likelihood analysis from COX1 sequences of *B. pisorum* (with *B. rufimanus* used as the outgroup). The different branches in the tree are highlighted using different colours. The bootstrap values for each node is given. It can be seen that four distinct groups are formed in the tree.

The first group, indicated in green, consists of two population representatives, one from the German population (Ger6) and one from the Chinese population (Ch6). This grouping is supported by a 100% bootstrap value. The second grouping (in red) consists of a Chinese (Ch3) and Australian (AsB) population representative. This grouping is supported by a 93% bootstrap value. The third grouping (in blue) consists of two Ethiopian (Et5k, Et5L), one Australian (AsJ) and one American (USA11) population representative. This grouping is supported by a 55% bootstrap value. The fourth grouping (in purple) is a split from the third grouping. This split is supported by a 62% bootstrap value. This grouping consists of two Ethiopian (Et1A, Et5O) and one American (USA8) representative.

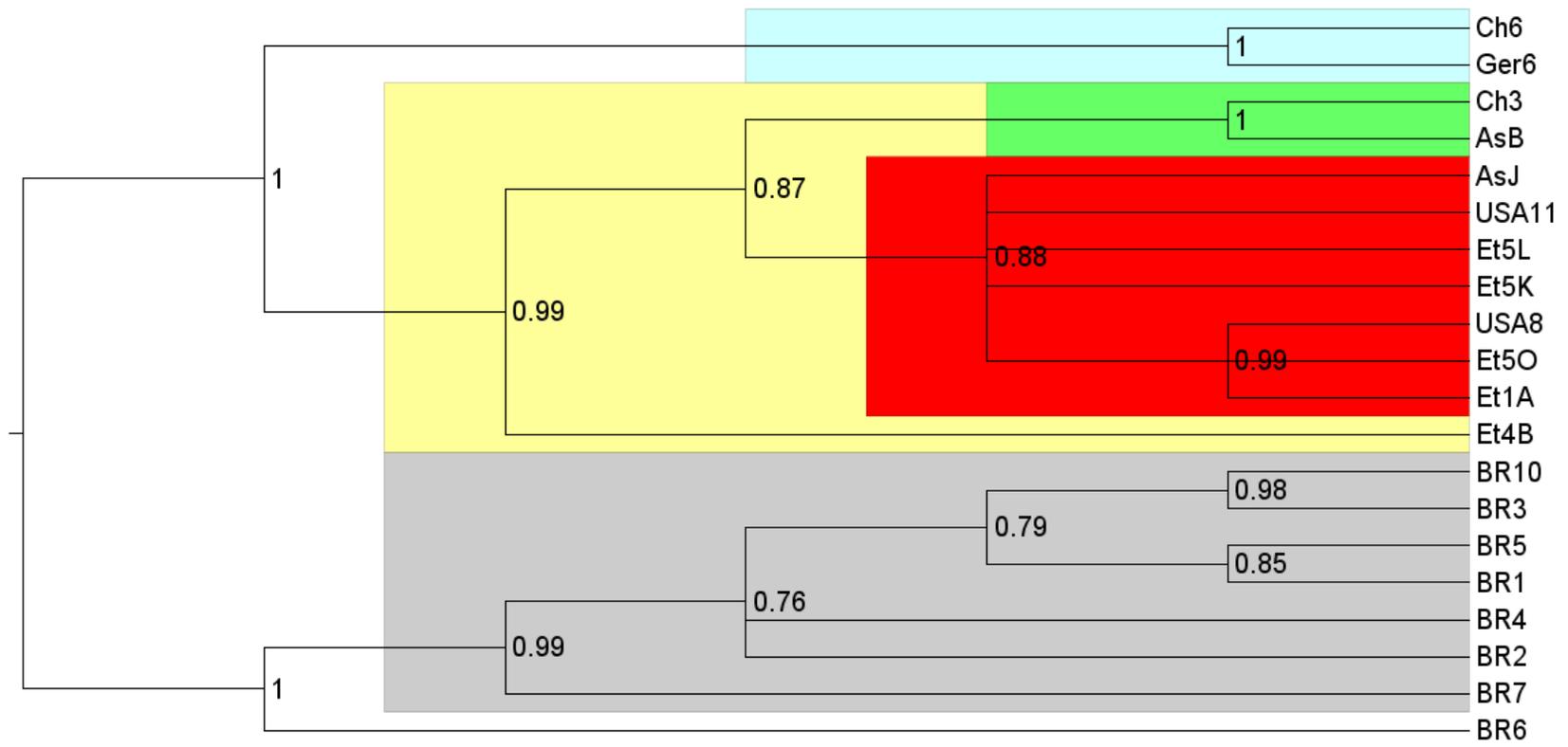


Figure 3-6 Phylogenetic tree inferred by the Bayesian analysis of phylogeny constructed with COX1 sequences of *B. pisorum*

Figure 3-6 shows a Phylogenetic tree of the COX1 sequences of *B. pisorum* constructed using the MrBayes software, implementing the Bayesian analysis of phylogeny method. The posterior probability of each node is given. COX1 sequences of *B. rufimanus* were used as the outgroup. Four final groupings are formed in the tree.

The first grouping (highlighted blue) is supported by a posterior probability value of 1. This grouping consists of two population representatives, one from the German population (Ger6) and one from the Chinese population (Ch6). The second grouping (highlighted green) is also supported by a posterior probability value of 1. This grouping consists of a Chinese (Ch3) and an Australian (AsB) population representative. The third grouping (highlighted red) is supported by a posterior probability value of 0.88 and consists of two Ethiopian (Et5k, Et5L), one Australian (AsJ) and one American (USA11) population representative. The fourth grouping is a split from the third grouping. This grouping is supported by a posterior probability value of 0.99 and consists of two Ethiopian (Et1A, Et5O) representatives and one American (USA8) representative.

The Phylogenetic trees constructed by the Maximum Likelihood and Bayesian analysis of phylogeny methods displayed highly similar structures, and the taxon groupings were identical.

3.4.2 Phylogenetic trees constructed with Cytb sequences

Phylogenetic trees were constructed using single Cytb sequences that are representative of each Cytb haplotype in each population. The sample IDs of sequences that are representative of the haplotypes in each population are listed in Table 3-7.

As for COX1, the phylogenetic trees presented in Figures 3-7 to 3-10 were constructed by NJ, ME, UPGMA and MP methods, using similar software and analysis approaches. A bootstrap test with 1000 replicates was performed on each tree, and branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The four phylogenetic trees displayed patterns indicative of similar evolutionary structures. Variation of bootstrap values among the trees most likely caused the collapse of some branches. This explains the differences in taxon grouping on the branches of different trees.

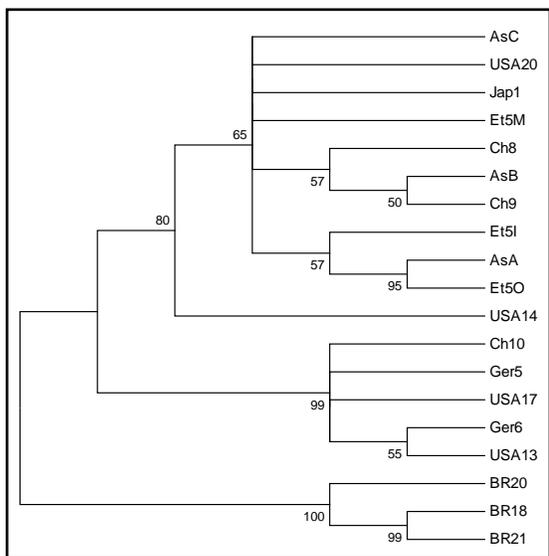


Figure 3-7 Evolutionary relationships of *Cytb* sequences of *B. pisorum* populations inferred using the Neighbour-Joining method.

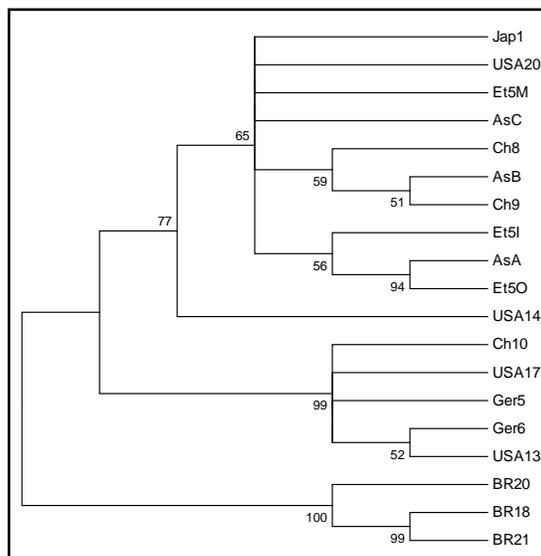


Figure 3-8 Evolutionary relationships of *Cytb* sequences of *B. pisorum* populations inferred using the Minimum Evolution method.

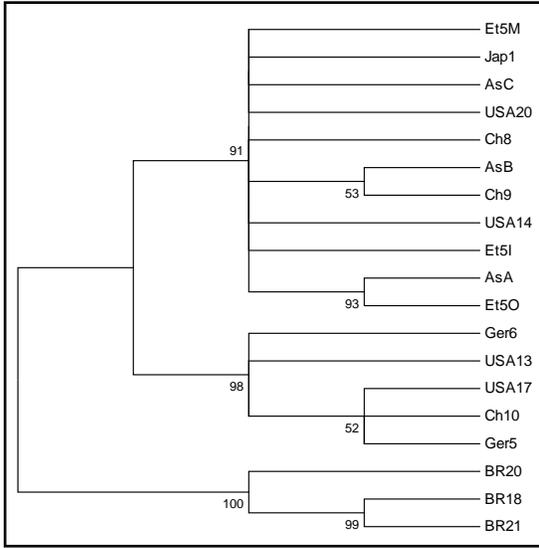


Figure 3-9 Evolutionary relationships of *B. pisorum* populations inferred using the UPGMA method.

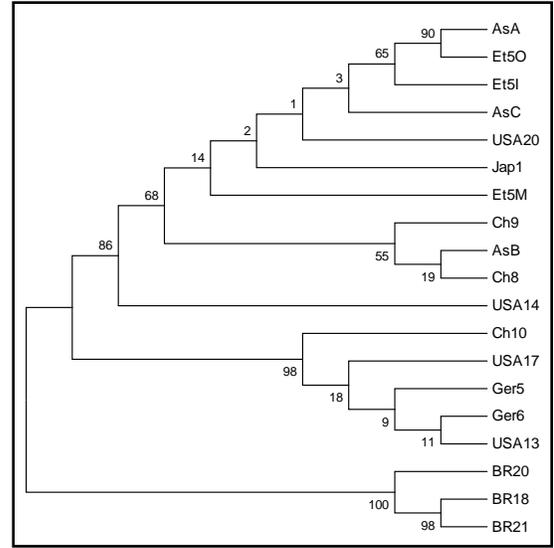


Figure 3-10 Maximum Parsimony analysis of *B. pisorum* populations.

A phylogenetic tree constructed by Maximum Likelihood analysis from *Cytb* sequences of *B. pisorum* is shown in Figure 3-11 (with *Cytb* sequences of *B. rufimanus* used as the outgroup). The different branches in the tree are highlighted in different colours, and the bootstrap value for each node is given. The following groupings can be observed on the tree.

The first grouping (in green) is supported by a 100% bootstrap value. This grouping consists of one American (USA17), one German (Ger5) and one Chinese (Ch10) population representative. A split occurs in the green branch to produce a secondary grouping supported by a 52% bootstrap value. This grouping consists of one American (USA13) and one German (Ger6) population representative.

The second grouping (in blue) is supported by a 68% bootstrap value and consists of one American (USA20), one Australian (AsC), one Japanese (Jap1) and one Ethiopian (Et5m) population representative. The third grouping (in red) is a split from the second grouping. There are actually two splits, the first supported by a 57% bootstrap value and the second split and

grouping supported by a 53% bootstrap value. The grouping consists of one Australian (AsB) and one Chinese (Ch9) population representative. The fourth grouping (in purple) is a split from the second grouping. Two splits occur: the first split is supported by a 61% bootstrap value and the second split and grouping is supported by a 91% bootstrap value and consists of an Australian (AsA) and an Ethiopian (Et5O) population representative.

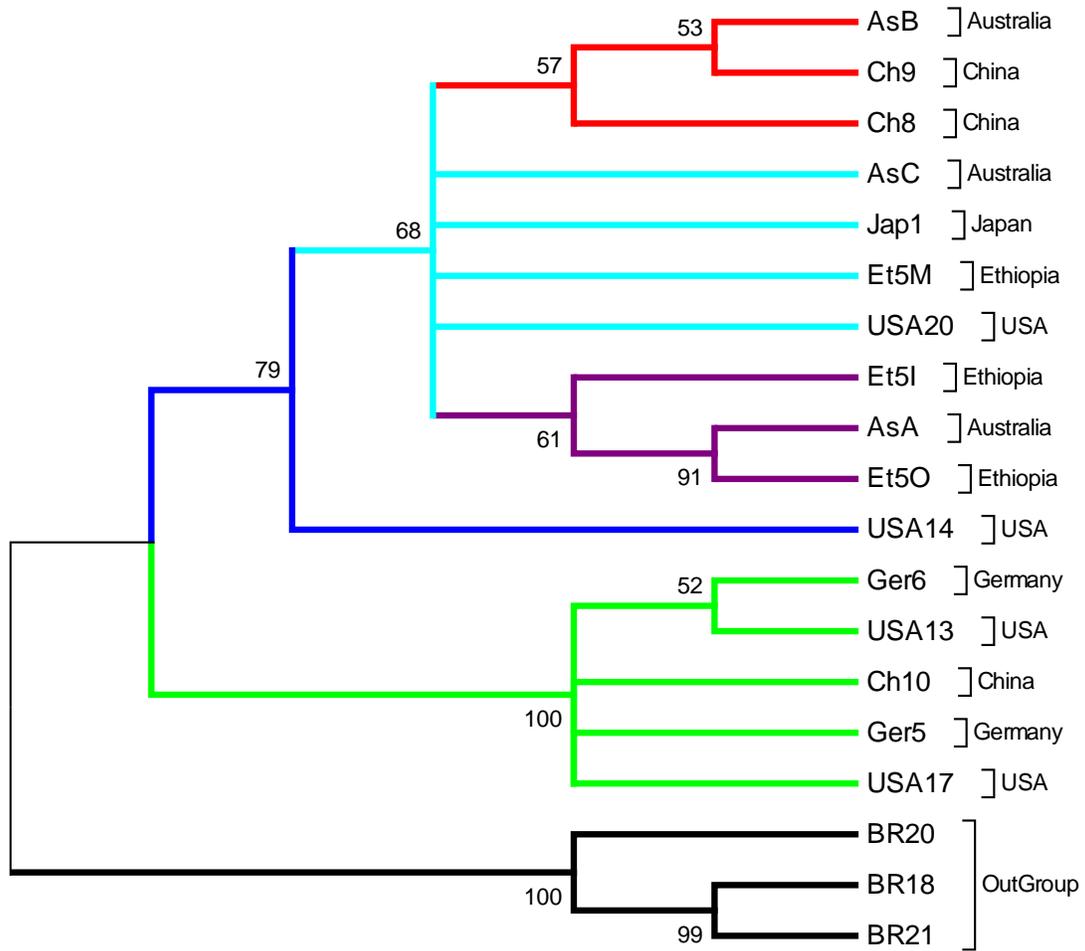


Figure 3-11 Molecular Phylogenetic analyses by Maximum Likelihood method using Cytb sequences

The evolutionary history was inferred using the Maximum Likelihood method based on the Jukes-Cantor model (Jukes and Cantor, 1969). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used, otherwise BIONJ method with MCL distance matrix was used. The analysis involved 19 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 251 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura, *et al.* 2011).

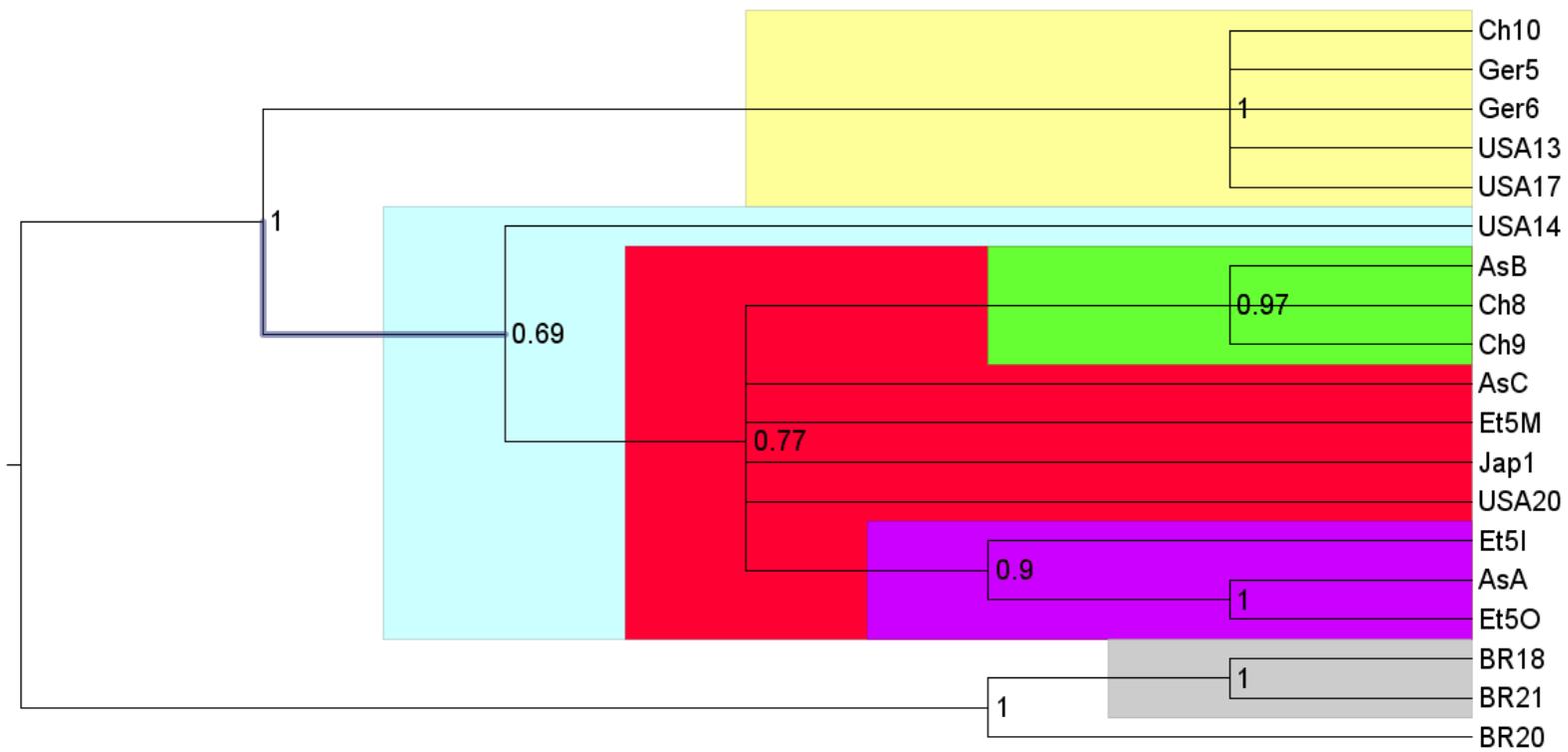


Figure 3-12 Phylogenetic Tree inferred by the Bayesian analysis of phylogeny constructed with Cytb sequences of *B. pisorum*

Figure 3-12 shows a phylogenetic tree of the Cytb sequences of *B. pisorum* constructed with MrBayes using the Bayesian analysis of phylogeny method. The posterior probabilities of each node are given, with Cytb sequences of *B. rufimanus* used as an outgroup. Four final groupings are formed in the tree.

The first grouping (highlighted in yellow) is supported by a posterior probability value of 1. This grouping consists of five population representatives, two from the German population (Ger5, Ger6), one from the Chinese population (Ch10) and two from the American population (USA13, USA17). The second grouping (highlighted in green) is supported by a posterior probability value of 0.97. This grouping consists of two Chinese (Ch8, Ch9) and one Australian (AsB) population representative. The third grouping (highlighted in red) is supported by a posterior probability value of 0.77 and consists of one Ethiopian (Et5M), one Australian (AsC), one Japanese (Jap1) and one American (USA20) population representative. The fourth grouping (highlighted in purple) is a split from the third grouping. Two splits occur: the first split is supported by a 0.9 posterior probability value while the second split and grouping is supported by a posterior probability value of 1. This grouping consist of one Ethiopian (Et5O) and one Australian (AsA) representative.

The Phylogenetic trees constructed by the maximum likelihood and Bayesian analysis of phylogeny method displayed similar structures, but the Bayesian analysis of phylogeny method provided a tree with greater resolution and with fewer splits in the branches.

3.4.3 Phylogenetic trees constructed from EF-1 α sequences

Sequences of EF-1 α which were representative of EF-1 α genotypes in each population were used to construct phylogenetic trees. The sample names of sequences that are representative of the genotypes in each population are listed Table 3-8. The phylogenetic tree was constructed with sequences which contained degenerate bases. These degenerate bases were the result of heterozygous nucleotide positions within the diploid gene.

Figure 3-13 shows the phylogenetic tree that was constructed using the Bayesian analysis of phylogeny method. The posterior probability of each node is given. EF-1 α gene sequences of *B. rufimanus* were used as an outgroup.

Little variation was observed in the tree, with all of the *B. pisorum* sequences grouping together (highlighted in yellow). Three splits occur in this grouping forming three additional groupings. The first grouping (highlighted in green) is supported by a 0.54 posterior probability value and consists of two Ethiopian population representatives (Et1H, Et3L). The second grouping (highlighted in purple) is supported by a 0.52 posterior probability value and consists of a single American population representative (USA13). The third grouping (highlighted in red) is supported by a 0.77 posterior probability value and consists of two American population representatives (USA14, USA21).

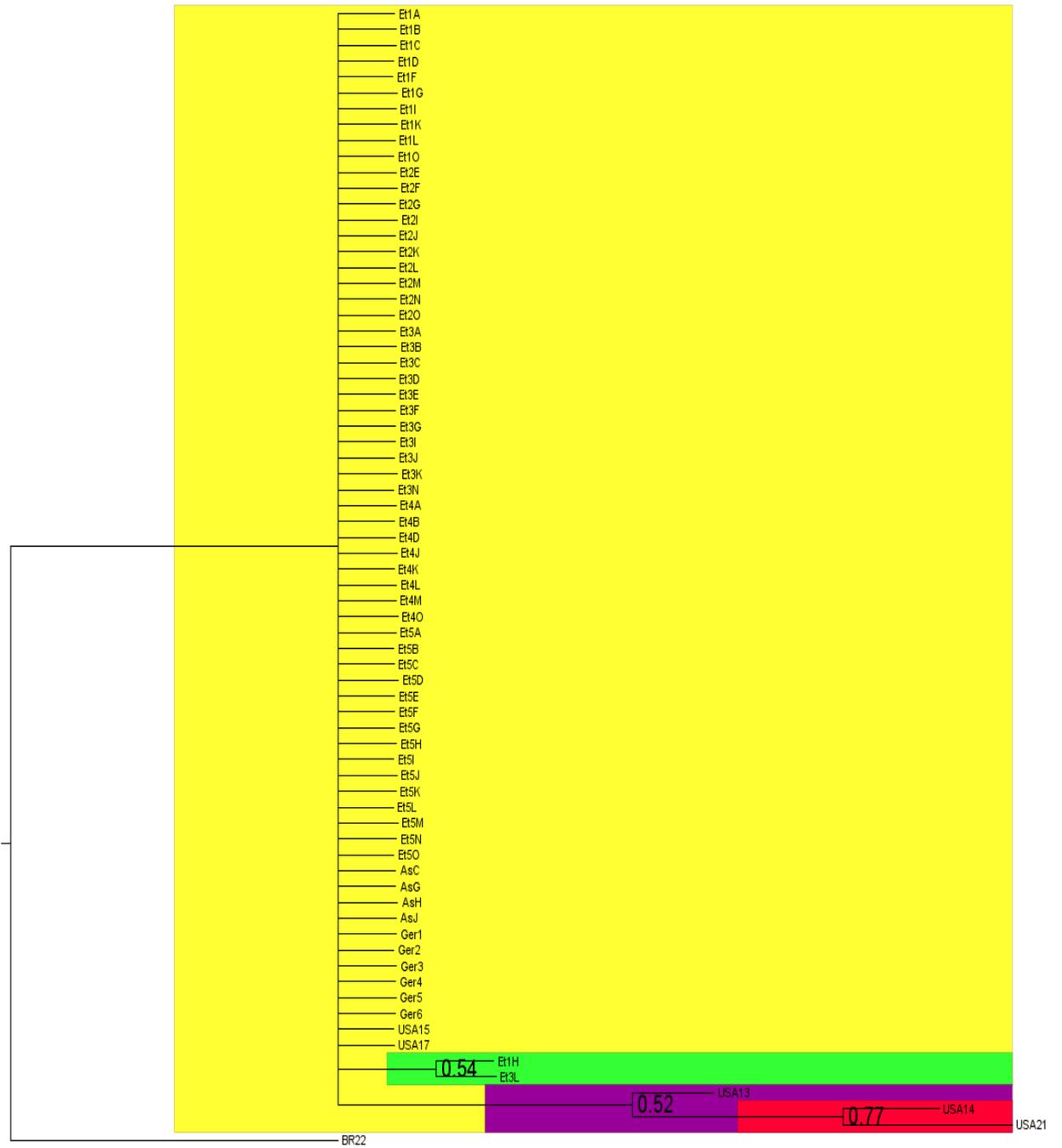


Figure 3-13 Phylogenetic tree inferred by the Bayesian analysis of phylogeny constructed with EF-1α sequences of *B. pisorum*.

3.5 Haplotype networks

The minimum spanning network based on Cytb haplotypes is shown in Figure 3-19. The haplotype CB1 was present in most individual specimens sampled across the populations and was present in the Ethiopian, Australian and Japanese populations. The haplotypes which are closest related to the CB1 haplotype are CB4 (present in the Australian and Chinese populations) and CB5 (present in the USA population) - both are separated from the CB1 haplotype by single mutational events. The CB7 haplotype (present in the Chinese population) is separated from CB4 by a single mutational event. The CB2 (present in the Ethiopian population) and CB10 (present in the USA population) haplotypes are both separated from the CB1 haplotype by two mutational events. Haplotype CB3 (present in the Ethiopian and Australian populations) is separated from the CB1 haplotype by three mutational events. The haplotypes that are the least related to the CB1 haplotype are CB6 (present in the Chinese and German populations) which is separated from the CB1 haplotype by nine mutational events, and CB9 (present in the German population) and CB8 (present in the German population), which are separated from the CB1 haplotype by 10 mutational events. The CB11 haplotype is a haplotype of *B. rufimanus*. The haplotype frequencies in each population are listed in Table 3-6.

The minimum spanning network built from COX1 haplotypes is shown in Figure 3-20. The haplotype that was present in most of the individuals sampled across the populations is the COX-5, which is present in the Ethiopian and American populations. The haplotypes most closely related to the COX-5 haplotype are COX-6 (present in the Ethiopian and USA populations), COX-7 (present in the Ethiopian population) and the COX-1 (present the Australian population). These are separated from the COX-5 haplotype by single mutational events. COX-4 (present in the Ethiopian population) is separated from the COX-6 haplotype by a single mutational event. The haplotypes which are the most distantly related to COX-5 are

COX-2 (present in the Australian and Chinese population) separated by three mutational events, COX-8 (present in the Ethiopian population) separated by five mutational events and the COX-3 haplotype (present in the Chinese population), separated by 21 mutational events. The COX-9 haplotype (present in the German population) is separated by one mutational event from the COX-3 haplotype and by 22 mutational events from COX-5. The haplotype COX-10 is a haplotype from *B. rufimanus* and was used as outgroup. The haplotype frequencies in each population are listed in Table 3-5.

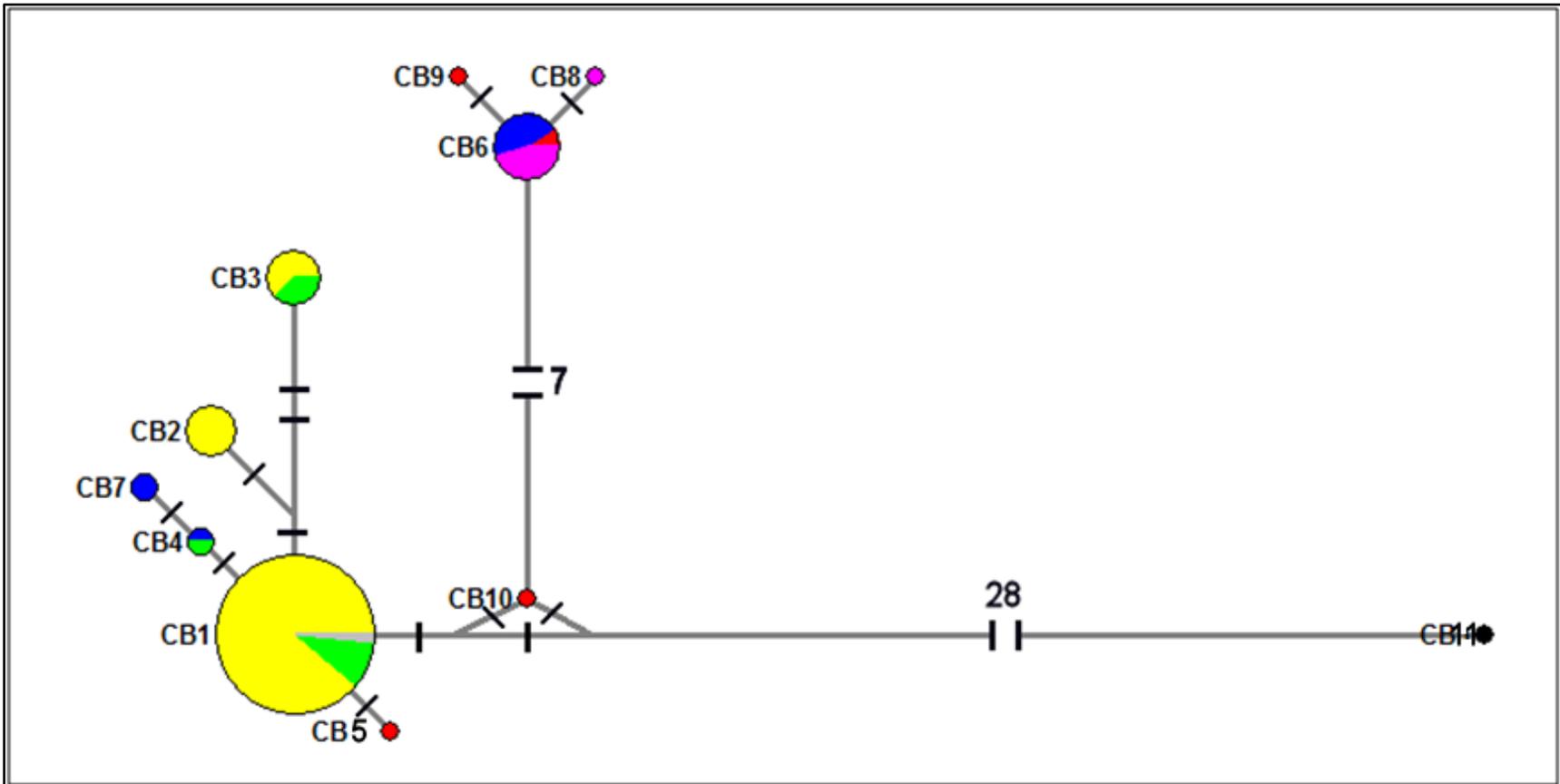


Figure 3-14 A minimum-spanning network between Cytb haplotypes of *B. pisorum*. Nodes represent haplotypes and the sizes of the nodes are proportional to the number of individuals that share that haplotype. The number of mutational events between two haplotypes are indicated with crossbars. When more than four mutational events occurred the number of events is indicated next to the interrupted lines. Colours are used to indicate in which populations the haplotypes occur. Australia ●, China ●, Ethiopia ●, Germany ●, Japan ○, USA ●, outgroup haplotype of *B. rufimanus* ●

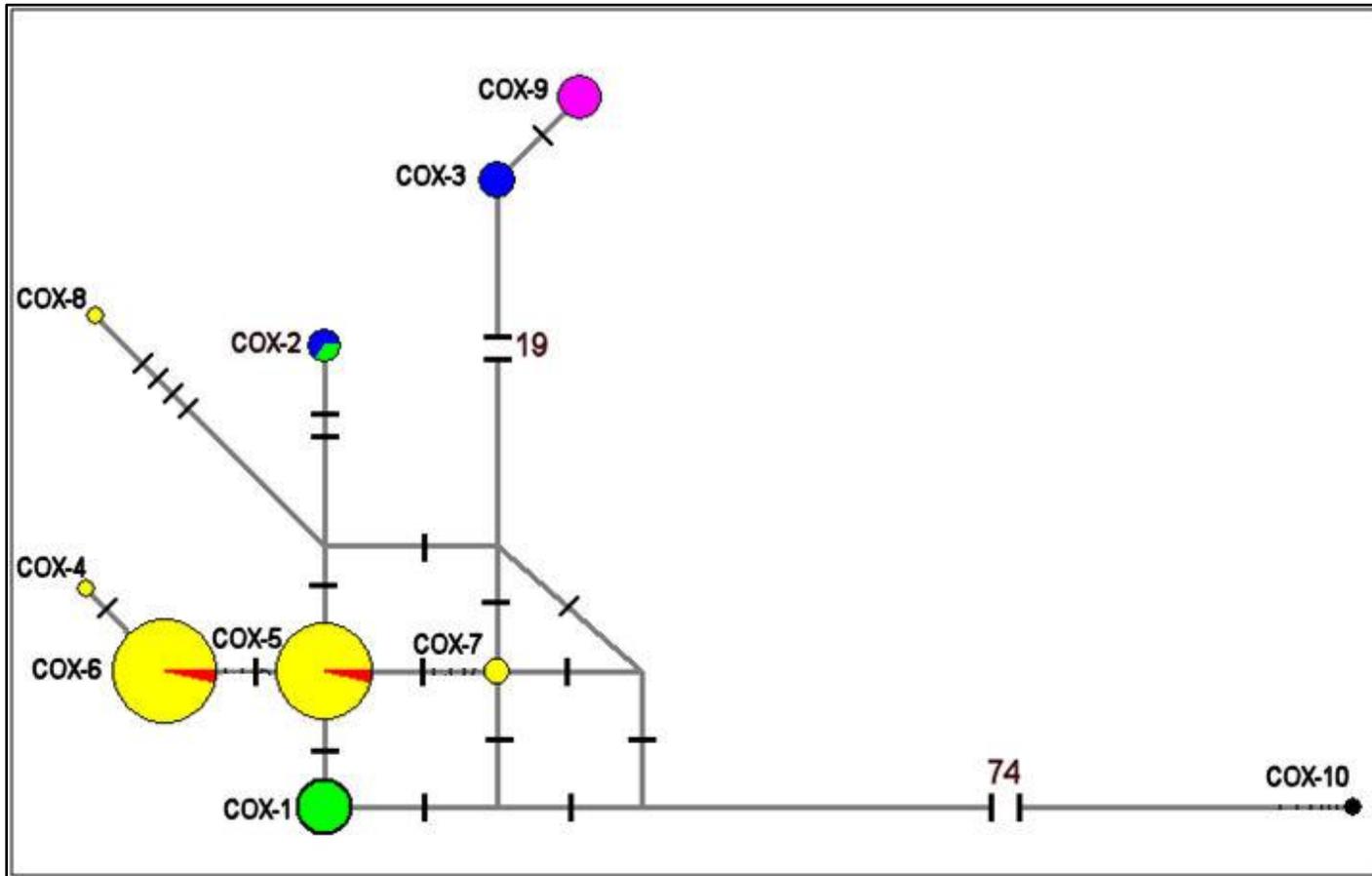


Figure 3-15 A minimum-spanning network between COX1 haplotypes of *B. pisorum*. Nodes represent haplotypes and the sizes of the nodes are proportional to the number of individuals that share that haplotype. The number of mutational events between two haplotypes are indicated with crossbars. When more than four mutational events occurred the number of events is indicated next to the interrupted lines. Colours are used to indicate in which populations the haplotypes occur. Australia ●, China ●, Ethiopia ●, Germany ●, USA ●, outgroup haplotype of *B. rufimanus* ●

3.6 Population diversity

3.6.1 Ethiopian subpopulations

The complete list of the diversity values of COX1 for each measure and subpopulation is presented in Table 3-8. The subpopulation with the most polymorphic sites was the Et4 population with 6. The nucleotide diversity in the populations ranged between 0.001 and 0.003, with the Et4 population displaying the highest nucleotide diversity. The haplotype diversity ranged between 0.507 and 0.614. The Et3 population displayed the highest haplotype diversity. The mean pairwise differences ranged between 0.545 and 1.905, with the Et4 population displaying the highest number of pairwise differences.

Table 3-8 Genetic diversity in the *B. pisorum* Ethiopian subpopulations according to COX1 sequence data

		<i>Bruchus pisorum</i> Ethiopian subpopulations COX1				
Haplotype	GenBank no	Et1	Et2	Et3	Et4	Et5
COX-4	JQ434485	0.067	-	-	-	-
COX-5	JQ434489	0.267	0.500	0.462	0.714	0.615
COX-6	JQ434487	0.667	0.500	0.462	0.143	0.308
COX-7	JQ434488	-	-	0.077	-	0.077
COX-8	JQ434486	-	-	-	0.143	-
No. polymorphic sites (<i>n</i>)		2	1	2	6	2
Nucleotide diversity (π)		0.001 ± 0.001	0.001 ± 0.001	0.001 ± 0.001	0.003 ± 0.002	0.001 ± 0.001
No. haplotypes		3	2	3	3	3
Haplotype diversity (<i>h</i>)		0.507	0.545	0.614	0.524	0.564
Mean number of pairwise differences		0.610 ± 0.511	0.545 ± 0.482	0.692 ± 0.560	1.905 ± 1.228	0.667 ± 0.546

The diversity values of the Cytb gene for each measure and population are listed in Table 3-9. The subpopulations with the most polymorphic sites were the Et4 and Et5 populations with 4 sites each. The nucleotide diversity in the populations ranged between 0 and 0.006. The haplotype diversity ranged between 0 and 0.591. The Et4 population displayed the highest haplotype diversity. The mean pairwise differences ranged between 0 and 1.6, with the Et5 population displaying the highest number of pairwise differences. The Et1 and Et3 populations displayed the lowest level of diversity.

Table 3-9 Genetic diversity in *B. pisorum* Ethiopian subpopulations according to Cytb sequence data

<i>Bruchus pisorum</i> Ethiopian subpopulations Cytb						
Haplotype	GenBank no	Et1	Et2	Et3	Et4	Et5
CB1	JQ478759	1	0.933	1	0.600	0.600
CB2	JQ478758	-	0.067	-	0.333	0.133
CB3	JQ478760	-	-	-	0.067	0.267
No. polymorphic sites (<i>n</i>)		0	2	0	4	4
Nucleotide diversity (π)		0	0.001 \pm 0.001	0	0.005 \pm 0.004	0.006 \pm 0.004
No. haplotypes		1	2	1	3	3
Haplotype diversity (<i>h</i>)		0	0.134	0	0.562	0.591
Mean number of pairwise differences		0	0.267 \pm 0.309	0	1.257 \pm 0.839	1.600 \pm 1.004

The complete list of the diversity values of the EF-1 α gene for each measure and subpopulation is presented in Table 3-10. The nucleotide diversity in the populations ranged between 0 and 0.001. The genotype diversity for all populations was identical at 1.0. The mean pairwise differences ranged between 0.152 and 0.22, with the Et3 population displaying the highest number of pairwise differences.

Table 3-10 Genetic diversity in *B. pisorum* Ethiopian subpopulations according to EF-1 α sequence data

<i>Bruchus pisorum</i> Ethiopian subpopulations EF-1α					
	Et1	Et2	Et3	Et4	Et5
Nucleotide diversity (π)	0.001 \pm 0.001	0	0.001 \pm 0.001	0	0.001 \pm 0.001
No. genotypes	15	14	14	10	15
Genotype diversity (h)	1	1	1	1	1
Mean number of pairwise differences	0.190 \pm 0.255	0	0.22 \pm 0.278	0	0.152 \pm 0.226
Number of polymorphic sites	2	0	2	0	1

Population differentiation between population pairs was established by calculating the pairwise F_{ST} and associated p values; a significance level of 0.05 was used. The German population of *B. pisorum* was used as an outgroup, because the phylogenetic trees constructed with the COX1 and Cytb gene sequences indicated that the German population was not closely related to the Ethiopian population.

Table 3-11 shows the F_{ST} and F_{ST} p values between population pairs based on the COX1 sequence data. The values indicate that there is significant differentiation between the Et1 group (Geboya population) and the Et4 group (Selamaya). There is no significant differentiation between pairwise combinations of the remaining populations, except for the outgroup population, which shows significant differentiation from all the Ethiopian subpopulations.

Table 3-12 shows the F_{ST} and F_{ST} p values between the population pairs based on the Cytb sequence data. The Et4 (Selamaya) and Et5 (Qualias)

populations both show significant differentiation from the Et1 (Geboya) and Et3 (Adet Hanna) populations, while the Et4 (Selamaya) population also shows significant differentiation from the Et2 (Jiman Derega) population. The outgroup population shows significant differentiation from all the Ethiopian subpopulations.

Table 3-13 shows the F_{ST} and F_{ST} p values between the populations pairs based on the EF-1 α sequence data. There is no significant differentiation between the populations. The outgroup population was also not significantly different from the remaining populations.

Table 3-11 F_{ST} values and F_{ST} p values between pairs of Ethiopian subpopulations based on COX1 sequence data. The populations are as follows: (1) Eth1, (2) Eth2, (3) Eth3, (4) Et4, (5) Et5 and (6) an outgroup German population. Values highlighted in green indicate no significant differentiation between population pairs, while values highlight in red indicate significant differentiation between the population pairs.

COX1 <i>B. pisorum</i> Ethiopian population F_{ST} and F_{ST} p values						
	1	2	3	4	5	6
1	*					
	*					
2	0.02 0.369±0.046	*				
		*				
3	0.031 0.667±0.024	0.073 0.991±0.003	*			
			*			
4	0.128 0.018±0.012	0.03 0.288±0.03	0.045 0.144±0.028	*		
				*		
5	0.065 0.180±0.038	0.051 0.712±0.04	0.044 0.838±0.043	0.001 0.604±0.047	*	
				0.001 0.604±0.047	*	
6	0.98 0	0.983 0	0.978 0	0.955 0	0.979 0	*
						*

Table 3-12 F_{ST} values and F_{ST} p values between pairs of Ethiopian subpopulations based on Cytb sequence data. The populations are as follows: (1) Eth1, (2) Eth2, (3) Eth3, (4) Et4, (5) Et5 and (6) an outgroup German population. Values highlighted in green indicate no significant differentiation between population pairs, while values highlight in red indicate significant differentiation between the population pairs.

Cytb <i>B. pisorum</i> Ethiopian population F_{ST} and F_{ST} p values						
	1	2	3	4	5	6
1	*					
	*					
2	-0.005 0.991±0.003	*				
		*				
3	0.000 0.991±0.003	-0.005 0.991±0.003	*			
			*			
4	0.265 0.009 ± 0.009	0.156 0.045±0.0203	0.265 0.045±0.020	*		
				*		
5	0.241 0.027±0.014	0.173 0.090±0.019	0.241 0.036 ±0.015	0.017 0.351±0.070	*	
					*	
6	0.990 0.000	0.969 0.000	0.990 0.000	0.893 0.000	0.872 0.000	*
						*

Table 3-13 F_{ST} values and F_{ST} p values between pairs of Ethiopian subpopulations based on EF-1 α sequence data. The populations are as follows: (1) Eth1, (2) Eth2, (3) Eth3, (4) Et4, (5) Et5 and (6) an outgroup German population. Values highlighted in green indicate no significant differentiation between population pairs, while values highlight in red indicate significant differentiation between the population pairs.

EF-1 α <i>B. pisorum</i> Ethiopian population F_{ST} and F_{ST} p values						
	1	2	3	4	5	6
1	*					
	*					
2	-0.113 0.667±0.053	*				
		*				
3	-0.078 0.712±0.029	-0.197 0.883±0.03	*			
			*			
4	-0.114 0.775±0.043	0 0.991±0.003	-0.185 0.748±0.039	*		
				*		
5	-0.071 0.649±0.035	-0.143 0.829±0.027	-0.08504 0.757±0.024	-0.379 0.955±0.02	*	
				*		
6	-0.5484 0.946±0.021	0 0.991±0.003	-0.631 0.973±0.013	0 0.991±0.003	-3. 0.991±0.003	*
						*

3.6.2 Diversity within country-specific populations

The within-region diversity values of the Cytb gene for each measure and population are listed in Table 3-14. The population with the most polymorphic sites was the USA population with 11, followed by the Chinese population with 10. The nucleotide diversity in populations ranged between 0.001 and 0.027. The USA population displayed the highest nucleotide diversity. The haplotype diversity ranged between 0.312 and 1 with the USA population, displaying the highest haplotype diversity of 1. The mean pairwise differences ranged between 0.333 and 5.071, with the Chinese population displaying the highest number of pairwise differences. The population displaying the least genetic diversity was the German population, with the lowest values for all the diversity measures.

Table 3-14 Genetic diversity in *B. pisorum* populations according to Cytb sequence data

Haplotype	GenBank no	<i>Bruchus pisorum</i> Cytb				
		Ethiopia	Australia	China	Germany	USA
CB1	JQ478759	0.822	0.600	-	-	-
CB2	JQ478758	0.110	-	-	-	-
CB3	JQ478760	0.069	0.300	-	-	-
CB4	JQ478757	-	0.100	0.125	-	-
CB5	JQ478765	-	-	-	-	0.250
CB6	JQ478761	-	-	0.625	0.833	0.250
CB7	EF570108.1	-	-	0.250	-	-
CB8	JQ478762	-	-	-	0.167	-
CB9	JQ478763	-	-	-	-	0.250
CB10	JQ478764	-	-	-	-	0.250
No. polymorphic sites (<i>n</i>)		4	4	10	1	11
Nucleotide diversity (π)		0.003 \pm 0.002	0.006 \pm 0.002	0.020 \pm 0.012	0.001 \pm 0.002	0.027 \pm 0.019
No. haplotypes		3	3	3	2	4
Haplotype diversity (<i>h</i>)		0.312	0.600	0.607	0.334	1
Mean number of pairwise differences		0.753 \pm 0.562	1.600 \pm 1.033	5.071 \pm 2.752	0.333 \pm 0.380	0.667 \pm 3.984

The diversity values of the COX1 gene for each measure and population are listed in Table 3-15. The population with the most polymorphic sites was the Chinese population with 22. The nucleotide diversity in the populations ranged between 0.001 and 0.018, with the Chinese population displaying the highest nucleotide diversity. The haplotype diversity ranged between 0.2 and 1 with the USA population displaying the highest haplotype diversity of 1. The mean pairwise differences ranged between 0.774 and 11.733. The Chinese population displayed the highest number of pairwise differences. The population displaying the least genetic diversity was the German population which displayed the lowest values for all the diversity measures.

Table 3-15 Genetic diversity in *B. pisorum* populations according to COX1 sequence data

		<i>Bruchus pisorum</i> COX1				
Haplotype	GenBank no	Ethiopia	Australia	China	Germany	USA
COX-1	JQ434491	-	0.900	-	-	-
COX-2	JQ434490	-	0.100	0.333	-	-
COX-3	EF570097.1	-	-	0.667	-	-
COX-4	JQ434485	0.017	-	-	-	-
COX-5	JQ434489	0.483	-	-	-	0.500
COX-6	JQ434487	0.450	-	-	-	0.500
COX-7	JQ434488	0.033	-	-	-	-
COX-8	JQ434486	0.017	-	-	-	-
COX-9	JQ434492	-	-	-	1	-
No. polymorphic sites (<i>n</i>)		8	4	22	0	1
Nucleotide diversity (π)		0.001 \pm 0.001	0.001 \pm 0.001	0.018 \pm 0.011	0	0.002 \pm 0.002
No. haplotypes		5	2	2	1	2
Haplotype diversity (<i>h</i>)		0.811	0.200	0.666	0	1
Mean number of pairwise differences		0.774 \pm 0.573	0.800 \pm 0.628	11.733 \pm 6.207	0	1 \pm 1

The complete list of the diversity values based on the EF-1 α gene, for each measure and population, is presented in Table 3-16. The nucleotide diversity within populations ranged between 0 and 0.0145, with the USA population displaying the highest nucleotide diversity. The genotype diversity ranged between 0.644 and 1. The USA and German populations displayed the highest genotype diversity of 1. The mean pairwise differences ranged between 0 and 4.4, with the USA population displaying the highest number of pairwise differences. The population displaying the least genetic diversity was the Australian population which displayed the lowest values in all the diversity measures.

Table 3-16 Genetic diversity in *B. pisorum* populations according to EF-1 α sequence data

	<i>Bruchus pisorum</i> EF-1α			
	Ethiopia	Australia	Germany	USA
Nucleotide diversity (π)	0.0004 \pm 0.0007	0	0	0.015
No. genotypes	57	4	6	5
Genotype diversity (h)	0.994	0.644	1	1
Mean number of pairwise differences	0.111 \pm 0.182	0	0	4.400 \pm 2.608
No. of polymorphic sites	2	0	0	10

4 Discussion

4.1 Determining the origin of *B. pisorum* in Ethiopia

The main objective of the research described in this dissertation was to determine the origin of the introduced pea weevils currently found in Ethiopia. This species causes great economical losses throughout the world. The species was first documented in Ethiopia in 1985 (Esmelealem and Adane, 2007). The pea weevil is a cosmopolitan insect pest of the field pea and was most likely introduced to Ethiopia sometime during the 1970s. After the pea weevil has completed its larval development, the adult pea weevil can remain within pea seeds for 18 to 24 months and it is consequently easily transported with dried peas (Capinera, 2001). Ethiopia experienced severe famine during the 1970s and was dependent on food aid. The fact that the pea weevil can remain in pea seeds for such an extended period of time makes it highly probable that it was accidentally introduced with the food aid received by Ethiopia. Between 1974 and 1984 the USA was the primary source of food aid to Ethiopia (Kissi, 2000). This fact makes the USA a likely candidate for being the source of the pea weevils currently found in Ethiopia (though the *B. pisorum* population in the USA is itself most likely the result of an introduction). A second possibility is that pea weevils historically occurred in the area, but was not reported.

In order to determine the origin of the pea weevils in Ethiopia, a suitable sampling strategy was devised. This strategy consisted of finding populations of pea weevils across the globe and then comparing these populations, on a genetic level, with the pea weevil population currently found in Ethiopia. Similar sampling strategies have been used in previous studies on introduced insect pests. An example includes determining the origin of the cabbage root fly found in the north-eastern coast of North America, Biron *et al.* (2000)

sampled a total of 16 populations and generated RADP data to evaluate genetic distances and construct phylogenetic trees. In order to determine the geographical origin of the Argentine stem weevil found in New Zealand, Williams *et al.* (1994) collected samples from nine locations in South America, one location in Australia and five locations in New Zealand. RAPD data generated from the specimens were then used to construct phylogenetic trees and to calculate a similarity matrix. Finally, Cognato *et al.* (2005) attempted to determine the origin of the red turpentine beetle found in China. Specimens were collected from 32 locations and COX1 gene sequence data were generated and used to construct phylogenetic trees.

4.1.1 Evolutionary origin of *B. pisorum*

The genus *Bruchus* is composed of seed-beetles and consists of 36 valid species (Kergoat, *et al.* 2007b). Seed-beetle larvae only develop inside seeds and are thus characterised by strong plant tissue specialisation. They also show a high tendency to host-specialisation (Kergoat, *et al.* 2007b). Species of the *Bruchus* genus occur mostly in the Palearctic Region; however a few species also occur in North Africa and Asia (Kergoat, *et al.* 2007b). The Palearctic Region is composed of Europe, Asia north of the Himalaya foothills, northern Africa, and the northern and central parts of the Arabian Peninsula. The genus *Bruchus* does not occur naturally in the other world regions, but a few species have been accidentally introduced to North America, tropical Africa, Australia and Japan (Kergoat, *et al.* 2007b).

The evolutionary origin of *Bruchus pisorum* is uncertain but it is believed to have co-evolved with its host *P. sativum* (Byrne, 2005). If *B. pisorum* co-evolved with its host *P. sativum*, then the evolutionary origin of *B. pisorum* may show similarities with the evolutionary development of *P. sativum* and it would then have originated in the same geographical region. The origin of *P. sativum* has not been determined with certainty but the Mediterranean

region, western and central Asia and Ethiopia have been indicated as possible centres of origin (Byrne, 2005). Peas were most likely first cultivated in western Asia, from where it spread to Europe, China and India. The natural distribution of *B. pisorum* likely expanded with the spread of pea cultivation. This leaves the possibility that *B. pisorum* may not have been introduced to Ethiopia, or the broad region, in historically recent times, but may be endemic to the region.

4.1.2 Anthropogenic spread of *B. pisorum*

As discussed in the preceding section, *B. pisorum* possibly co-evolved with its host *P. sativum* and thus possibly originated in the Mediterranean region, western and central Asia or Ethiopia (Byrne, 2005). From there *B. pisorum* most likely spread with the cultivation of peas. In later years it was introduced to other world regions, possibly with dried peas transported to other regions to serve as food or seed.

When studying the dynamics of introduced populations, the level of diversity and differentiation within introduced populations can be highly informative when determining the mode of introduction. It can also provide insight into the geographical origin of introduced populations (Williams *et al.*, 1994). Genetic diversity in an introduced population, as measured using genetic markers, can provide information on the amount of genetic variation lost during the colonisation bottleneck or provide evidence of multiple sources of introduction (Grapputo *et al.*, 2005). The genetic diversity of invasive insect pests have been determined in several studies (Grapputo *et al.*, 2005; Ahern, *et al.*, 2008; Nicol *et al.*, 1997; Williams *et al.*, 1994; Biron *et al.*, 2000). The genetic diversity in an introduced population can be reduced or enhanced, depending on the mode of introduction. For example, enhanced genetic diversity in single populations may result from the introduction of individuals from genetically distinct populations, while a reduction of genetic diversity may

occur when colonisation occurs with the introduction of a few individuals from a common population (Ahern *et al.*, 2008).

In the Americas, pea weevils are known to be an introduced pest. Records indicate that pea weevils were doing damage to crops in the colonies as early as 1675 (Bain, 1998). The pea weevils were most likely introduced from Europe with shipments of peas in the 1600s, as records state that shipments of dry pea seeds arrived in the Massachusetts Bay Colony as early as 1628 (Bain, 1998). The Cytb sequence data generated during the current study indicated the USA population as the most diverse of the studied populations. The high diversity observed in this population may be explained by multiple founder events that consisted of the introduction of individuals from genetically distinct populations. The pea weevil specimens collected in the USA for this study originated in different regions. It may therefore be possible that in the approximately 350 years since they were introduced to America, different evolutionary pressures and regional isolation may have resulted in the diversification of the overall American population. Similar observations were made in a study of the Colorado potato beetle populations (Grapputo *et al.*, 2005), in which introduced populations found in the central United States (originally from Mexico) showed high levels of variation among populations, possibly explained by a series of distinct founder events that resulted in genetically diverse introduced populations.

Germany is located in the Palearctic Region which is indicated as the centre of origin of the *Bruchus* genus, but is not among the regions that are indicated as possible centers of origin for *B. pisorum*. The first record of the pea weevil in Europe was in the Czech Republic in 1850 (Beenen and Roques, 2010). Taking into account that both the Cytb and COX1 sequence data indicated the German population as the population with the least diversity, and that the EF-1 α sequence data indicated it as the second least diverse population studied, it is highly probable that the German population of pea weevils are a highly inbred population and not endemic to Germany. It is possible that this

population lost diversity during the colonisation bottleneck, in which the colonisation occurred with the introduction of a few individuals from a common population. A similar observation was made in a study of the Colorado potato beetle populations (Grapputo *et al.*, 2005), in which an introduced European population (originally from Mexico) showed little to no variation and was fixed to a single mitochondrial haplotype. It was suggested that this was the result of a unique or a single successful founder event. The possibility also exists that the six specimens from Germany could by coincidence be from a single closely related group. A larger sample of pea weevils from this area is needed to definitively evaluate the genetic diversity of the population.

Bruchus pisorum is believed to be an alien species in China (Xu, *et al.* 2006). China is located in the Palearctic Region, and depending on definition, the western part may be included or excluded from the central Asian region, which is indicated as a possible centre of origin of *B. pisorum* and its host *P. sativum* (Byrne, 2005). The Chinese population was indicated as the most diverse population according to the COX1 sequence data. This high level of diversity could be indicative that *B. pisorum* may have originated in this region. However the high level of diversity observed in the population may also be due to multiple founder events which consisted of the introduction of individuals from genetically distinct populations.

In Australia the pea weevil is a recently introduced species, first observed in 1931 in Western Australia (Waterhouse and Sands, 2001). Presently *B. pisorum* is distributed throughout South-western Australia. The diversity values obtained from the Cytb sequence data indicated that the Australian population possesses the second highest haplotype diversity of all the studied populations. The COX1 sequence data indicated that the Australian population displayed the second highest number of mean pairwise differences. This may indicate that colonisation of *B. pisorum* in Australia consisted of individuals from genetically distinct populations. In the

phylogenetic trees constructed with the Cytb and COX1 sequence data, it was observed that the Australian population grouped with all the populations except the German population, indicating that the Australian population is related to several different populations. This could indicate that the pea weevils in Australia have been introduced from several different source populations originating in different countries.

4.1.3 Identifying possible source populations of *B. pisorum* in Ethiopia

In the current study, pea weevil specimens were obtained from the USA, Germany and Australia. GenBank was also searched for sequence data of pea weevils from other countries. The search produced Cytb and COX1 sequence data of pea weevils from China and Japan. Including the Ethiopian population a total of six populations across the globe were thus used in this study. These populations spanned over five continents.

To resolve the relationships between *B. pisorum* from the populations sampled, phylogenetic trees were constructed by using sequences of the Cytb, COX1 and EF-1 α gene regions. In order to explore all possibilities and alternative hypotheses of evolution, these trees were constructed by using the neighbour joining, unweighted pair-group arithmetic mean, minimum evolution, Bayesian analysis of phylogeny, maximum likelihood and maximum parsimony methods. The trees all yielded fundamentally similar structures for each gene. Nevertheless, the trees constructed with the Bayesian analysis of phylogeny method seemed to have the highest values in support of the tree branches. Trees constructed with this method have also been shown to be the most accurate with the lowest percentage of error rates of all likelihood methods (Williams and Bernard, 2003). Furthermore it has been shown that when the neighbour joining method is compared with likelihood methods, it is the worst performer in terms of accuracy (Williams and Bernard, 2003). For this reason the final conclusions were drawn from the analyses performed on

the phylogenetic trees constructed with the Bayesian analysis of phylogeny method.

Phylogenetic trees constructed from genetic data have previously been used in studies to determine the relationships between populations of insect pests with the purpose of establishing the origin of an introduced pest, for example phylogenetic trees constructed with RAPD data (Biron *et al.*, 2000; Williams *et al.*, 1994) and COX1 sequence data (Cognato *et al.*, 2005).

From the phylogenetic tree constructed with the COX1 sequence data presented in Figure 3-12 (p. 56) it can be deduced that the specimens from the Ethiopian population is closest related to the specimens from the USA and Australian populations. The phylogenetic tree constructed with Cytb sequence data is presented in Figure 3-6 (p. 50). Only one Cytb sequence of a Japanese pea weevil was available on GenBank and could therefore only be compared to the studied populations using the Cytb gene sequences. In this tree it can be seen that the specimens from the Ethiopian population were related to the specimens from the Australian, USA and Japanese populations.

When both the phylogenetic trees, constructed from the COX1 and Cytb sequence data, are considered, it can be deduced that the Ethiopian population is related to the USA, Australian and Japanese populations but not to the German or Chinese populations.

The phylogenetic tree constructed from the EF-1 α sequences shown in Figure 3-13 (p. 59) consisted of specimens from the Ethiopian, Australian, German and American populations. In this tree no definitive groupings between the sequences from the Ethiopian population and any of the sequences from any other specific populations were formed. Superficially, this result implies that all of the populations are related to each other and that there are no drastic differences between the sequences of the different

populations. However, this result may also be indicative that the EF-1 α gene is a highly conserved gene, with little value for phylogenetic investigations.

For the final conclusion drawn in this study, only the results obtained for the mitochondrial genes were used, as the data obtained from the nuclear gene did not provide any clear pattern in terms of the relationships between the different populations. When compared to nuclear genes, mitochondrial genes appear to be of greater use when studying intra-species variation, due to the higher evolutionary rates of mitochondrial genes (Saeb, 2006). The differences observed between the trees constructed with mtDNA genes and the tree constructed with the nuclear gene can be explained by the differences in evolutionary rates and the mode of inheritance of these genes. In general, mitochondrial genes have advantages over nuclear genes in phylogenetic and population genetic studies due to the fact that they have higher evolutionary rates, lack recombination, are inherited in haploid mode, lack introns and appear to be selectively neutral (Saeb 2006; Gasser *et al.*, 2002; Hu *et al.*, 2002b; Liu *et al.*, 1999; Avise, 1994). The high evolutionary rates of mtDNA genes allow for their use in comparing both inter and intra-specific variation. Consequently it is possible that for this particular study, which is based on intra-specific variation, the results obtained from the Cytb and COX1 sequence data is more significant than the results obtained from the EF-1 α sequence data.

In the phylogenetic trees constructed with the Cytb and COX1 sequences, it was observed that: the Ethiopian population never grouped with the Chinese or German populations. It can therefore be interpreted that these two populations can be excluded as possible direct sources of the pea weevils currently in found Ethiopia. These populations could nevertheless still be the original source. The Ethiopian population did, however, frequently group with the USA and Australian populations. In the Cytb tree, the Ethiopian population also grouped with the sequence from the Japanese population, therefore these three populations cannot be excluded with certainty as

possible origins of the pea weevil in Ethiopia. Note that the Japanese population was only represented by a single Cytb sequence. This result may therefore not be significant. Sequences of additional specimens and sequences of the COX1 gene could give a different perspective of the relationship between the Ethiopian and Japanese populations.

As mentioned in the preceding sections, the pea weevil is a recently introduced species in Australia, first observed in 1931 in Western Australia (Waterhouse and Sands, 2001). In the phylogenetic trees constructed with the Cytb and COX1 sequence data, it was observed that the Australian population grouped with all the populations except the German population, indicating that the Australian population is related to several different populations. This could indicate that the pea weevils in Australia have been introduced to Australia from several different source populations originating in several different countries. The relationships between the Ethiopian and Australian populations can therefore be explained by the possibility that they might have shared a source population.

Given the fact that the USA was the primary source of food aid to Ethiopia between 1974 and 1984 (Kissi, 2000), it can be concluded from all data currently available that the USA is the most likely country of origin of the pea weevils currently found in Ethiopia.

4.1.4 A possible endemic population of *B. pisorum* in Ethiopia

Ethiopia has been indicated as a possible centre of origin for *Pisum sativum*, and thus for *B. pisorum*. This fact gives rise to the possibility that a naturally occurring population of *B. pisorum* was historically present in Ethiopia. It is possible that the population numbers were low enough that the pea weevil did not reach pest status and thus went unnoticed by the general populace. Ecological factors could have caused a rise in the number of pea weevils

causing it to reach pest status. Several different ecological factors are known to cause a rapid increase in insect pest populations. These include variation of soil quality, change in predation pressure, change in the density of competitors, fluctuations in temperature, the quality of food and the level of precipitation (Hunter, 2002). Droughts have been associated with outbreaks of species of beetles, aphids, saw-flies, Lepidoptera and grasshoppers (Hunter, 2002). It was suggested by White (1984) that plants stressed by drought mobilise nitrogen in their tissue and that nitrogen availability increases herbivore performance, pointing to a link between droughts and outbreaks of pests.

The possibility also exists that a second non-endemic population of pea weevil was introduced to Ethiopia and that the pea weevils currently in Ethiopia are a mixture of endemic and introduced pea weevils. The differentiation from F_{ST} and associated p values calculated for the COX1 and Cytb genes indicated that the Geboya (Et1) and Selamaya (Et4) populations in Ethiopia were significantly different from each other. The reason for this could be that one population is an endemic population while the other is an introduced population.

Three COX1 and one Cytb haplotype, unique to the Ethiopian population were identified. In the phylogenetic tree constructed with the COX1 sequence data, one Ethiopian haplotype was located on a branch supported by a posterior probability value of 0.99 and did not group with any of the other populations. The phylogenetic tree constructed with the Cytb sequence data also showed an Ethiopian haplotype appearing alone on a branch that is supported by a posterior probability value of 0.9. This haplotype did not form a group with any other population. This indicates that there are Ethiopian individuals that do not show close relationships with any of the studied populations. It is therefore possible that an endemic population of pea weevils have existed historically in Ethiopia.

4.2 Diversity and differentiation across Ethiopian subpopulations

A secondary objective of the research described here was to determine the current population structure of pea weevils across Ethiopia. This was done by determining genetic differentiation among the pea weevil specimens collected from five different regions in Ethiopia. Differentiation between specific introduced populations can be the result of the number of founding events, genetic variation in the founder populations and selection pressure (Williams *et al.*, 1994).

To determine the degree of differentiation between the different geographical populations of the pea weevil in Ethiopia, the pairwise F_{ST} and associated p values were calculated for all three genes studied. The German population of pea weevils was used as the outgroup because the phylogenetic trees which were constructed with the COX1 and Cytb gene sequences indicated that the German population was the least related to the Ethiopian populations.

The F_{ST} and associated p values calculated for population pairs using the COX1 gene sequence data indicated significant differentiation between the Geboya (Et1) population and the Selamaya (Et4) population. By comparison, the F_{ST} values indicated that the outgroup population was significantly different from all the Ethiopian subpopulations. For the Cytb gene sequences, the values indicated significant differentiation of the Selamaya (Et4) and Qualias (Et5) populations from the Geboya (Et1) and Adet Hanna (Et3) populations, while the Selamaya (Et4) population also showed significant differentiation from the Et2 (Jiman Derega) population. The outgroup population again showed significant differentiation from all the Ethiopian subpopulations.

For both the Cytb and COX1 sequence data the Geboya (Et1) and Selamaya (Et4) populations were shown to be significantly different from each other. This could indicate that these two subpopulations were isolated from each

other for a significant period of time (possibly with diverse forms of selection pressure), or that these populations were introduced separately to Ethiopia from two different sources, or that one population is endemic to Ethiopia while the other is introduced. Taking into account the possibility that the pea weevil is a recently introduced species since it was first documented in Ethiopia in 1985 (Esmelealem and Adane, 2007), it is highly doubtful that enough generations have passed to cause differentiation between the populations due to selection pressure. The most likely explanation is therefore that the two populations represent separate source populations.

For the EF-1 α gene sequences, the values indicated that there is no significant differentiation between any of the population pairs. Even the outgroup population was shown not to be significantly different from any of the populations. This suggests that the nuclear gene, EF-1 α , has much lower resolution properties than the mitochondrial genes, Cytb, and COX1, when evaluating intra-species variation, as was the case during the phylogenetic analysis.

The levels of diversity in each gene for each of the Ethiopian subpopulations were determined by calculating five diversity measures. According to the COX1 sequence data the most diverse population was the Selamaya (Et4) population, whereas the Jiman Derega (Et2) population showed the lowest diversity. The values obtained from the Cytb sequence data indicated that the Qualias (Et5) population displayed the highest diversity and the Geboya (Et1) and Jiman Derega (Et2) populations showed the lowest diversity. The EF-1 α gene sequence indicated the Geboya (Et1) population as the most diverse population and the Jiman Derega (Et2) population as the least diverse.

4.3 Notes on DNA extraction and gene amplification

DNA was extracted and sequenced from the specimens collected in Ethiopia, Australia, Germany and the USA. From the 74 Ethiopian specimens, a total of 62 COX1, 73 Cytb and 68 EF-1 α gene region sequences were produced. All three gene regions were successfully sequenced in the six German and 10 Australian specimens. The failure to produce sequences of all three gene regions in all individuals is most likely due to low quality of the extracted DNA rather than mutations or deletions that occurred in the primer binding sites, causing the binding of primers to fail.

The specimens obtained from the USA presented a challenge, since these are museum specimens collected between 1913 and 1938. The age of the specimen alone was problematic, but the specimens were also provided as a loan by the Smithsonian museum and instructions were given that the specimens may not be damaged. DNA was extracted by using a modification of the method described by Gilbert *et al.*(2007). The specimens sustained minimum morphological damage during the DNA extraction. From eight individual museum specimens a total 11 sequences were obtained, namely two COX1, four Cytb and five EF-1 α sequences. Two specimens that yielded EF-1 α sequences did not yield COX1 or Cytb sequences. This was surprising as it would be expected that more mitochondrial gene sequences would be produced than nuclear gene sequences, due to the higher amount of mtDNA copies in cells. This anomaly may possibly be caused by degrading in the primer binding sites of the COX1 and Cytb gene regions.

4.4 Future avenues for research

Further studies of *B. pisorum* should aim to further elucidate the geographic origin of this species. It is currently believed that the pea weevil co-evolved with its host *P. sativum* (Byrne, 2005). The origin of *P. sativum* is not well known but the Mediterranean region, western and central Asia and Ethiopia have been indicated as possible centres of origin (Byrne, 2005). The exact location of the evolutionary geographical origin may be determined if pea weevil populations from as many regions as possible are located (including populations of pea weevils in the possible centres of origin) and an in-depth phylogenetic study is conducted on these populations.

5 Summary

This study aimed to determine the origin of pea weevils (*Bruchus pisorum*) in Ethiopia and to determine the current population structure across that country. The pea weevil is presently a widely distributed pest of peas in Ethiopia, causing huge financial losses. Conflicting hypotheses exist on the origin of *B. pisorum* in Ethiopia. It was possibly introduced to Ethiopia sometime in the 1970s, or it might have occurred historically in the area in very low numbers. The methodology of this study consisted of finding populations of pea weevils across the globe and then comparing these populations with the population currently found in Ethiopia. Specimens were obtained from Ethiopia, the USA, Germany and Australia. Gene sequences of pea weevils from China and Japan were also downloaded from GenBank to serve as reference material. DNA was extracted, amplified and sequenced using standard protocols, with the exception of the USA sample which was composed of museum specimens that demanded a non-destructive DNA extraction method. Three gene regions were used in this study: the Elongation Factor 1alpha (EF-1 α), Cytochrome oxidase subunit one (COX1) and Cytochrome b (Cytb). The COX1 and Cytb sequence data provided insight into a possible source population of pea weevils in Ethiopia, whereas results from EF-1 α were uninformative. Pea weevils from the USA were identified as a possible direct source, but it should be noted that these pea weevils are not endemic to the USA. The possibility of an endemic population of pea weevils in Ethiopia is also discussed. Tests for differentiation indicated that there was some differentiation between the Ethiopian subpopulations. This variation is discussed with reference to possible multiple sources of introduction for the current population in Ethiopia, genetic drift since introduction, and the possibility of a mixture of endemic and introduced genetic material in *B. pisorum* in Ethiopia.

6 Opsomming

Die doel van hierdie studie was om die oorsprong van die ertjiekalanders (*Bruchus pisorum*) in Ethiopië te bepaal en om die huidige populasie struktuur van hierdie insekte te ondersoek. Die ertjiekalander is tans 'n wydverspreide pes van ertjies in Ethiopië wat groot ekonomiese verliese veroorsaak. Botsende hipoteses oor die oorsprong van *B. pisorum* in Ethiopië bestaan. Die ertjiekalander is moontlik per abuis in die 1970s in Ethiopië ingevoer of alternatiewelik het dit histories in baie lae getalle in die gebied voorgekom. Die metodiek van hierdie studie het bestaan uit die soek na populasies van ertjiekalanders regoor die wêreld, en die vergelyking van hierdie populasies met die huidige Ethiopiese populasie. Eksemplare is verkry vanaf Ethiopië, die VSA, Duitsland en Australië. DNS volgordes van ertjiekalanders in China en Japan is ook vanaf GenBank afgelaai om te dien as verwysingmateriaal. DNS is onttrek, ge-amplifiseer en die DNS volgordes van die gene is bepaal deur die gebruik van standaard protokolle. Die VSA eksemplare was museum eksemplare en het en 'n nie-vernietigende DNA onttrekkings metode vereis. Drie gene is in hierdie studie gebruik, naamlik Verlenging Faktor 1alpha (EF-1 α), sitochroom oksidasie subeenheid een (COX1) en sitochroom b (Cytb). Die COX1 en Cytb volgorde data het insig in 'n moontlike bron-populasie van ertjiekalanders in Ethiopië verskaf, terwyl die resultate van die EF-1 α geen nie insiggewend was nie. Ertjiekalanders van die VSA is as 'n moontlike direkte bron geïdentifiseer, maar dit moet in ag geneem word dat ertjiekalanders nie endemies aan die VSA is nie. Die moontlikheid van 'n endemiese populasie ertjiekalanders in Ethiopië word ook bespreek. Toetse vir differensiasie het aangedui dat daar differensiasie tussen die Ethiopiese subpopulasies bestaan. Hierdie variasie word bespreek met verwysing na die moontlikheid van veelvuldige bronpopulasies vir die huidige Ethiopiese populasie, genetiese drywing sedert invoering en die moontlikheid van 'n mengsel van inheemse en uitheemse genetiese materiaal in *B. pisorum* in Ethiopië.

7 References

- Ahern, RG, DJ Hawthorne, and MJ Raupp. (2008) "Founder effects and phenotypic variation in *Adelges cooleyi*, an insect pest introduced to the eastern United States." *Biological Invasions* 11, no. 4: p959-971.
- Assayehegne, B. (2002): "The Biology and Ecology of Pea Weevil (Beetle) (*Bruchus Pisorum*) L, Coleoptera, Bruchidea." *Management of pea weevil, bruchus pisorum, ARARI/EARO.- Addis Abeba (Ethiopia).*p 37-46.
- Avise, J. C. (1994). "Molecular Markers, Natural History and Evolution". New York, NY: Chapman and Hall,
- Bain, A. (1998) "A Seventeenth-Century Beetle Fauna from Colonial Boston." *Historical Arhaeology* 32, no. 2: p38-48.
- Baker, G. (1998) "*Bruchus pisorum* (Linnaeus) (Coleoptera:Chrysomelidae)." South Australian Research and Development Institute. November. http://www.sardi.sa.gov.au/__data/assets/pdf_file/0007/44872/peaweevil.pdf (accessed 2011).
- Bandelt, H.J, P Forster, and A Röhl. (1999) "Median-joining networks for inferring intraspecific phylogenies." *Molecular Biology and Evolution* 16: p37-48.
- Bandelt, HJ, P Forster, BC Sykes, and MB Richards. (1995) "Mitochondrial portraits of human populations using median networks." *Genetics* 141: p743–753.
- Beenen, R, and A Roques. (2010): "Leaf and Seed Beetles (Coleoptera, Chrysomelidae). Chapter 8.3." *BioRisk* 4, no. 1: p267–292.
- Bekele, E, F Azerefeagne, and T Abate. (2006) "Facilitating the Implementation and Adoption of Integrated Pest Management (IPM) in Ethiopia." *Drylands Coordination Group.*
- Berim, M.N. (2008) *Interactive Agricultural Ecological Atlas of Russia and Neighboring Countries. Economic Plants and their Diseases, Pests and Weeds.* S.L. Greene, N.I. Dzyubenko, A.N. Frolov (eds.). *Lonicera edulis*. In A.N.Afoninhttp://www.agroatlas.ru/en/content/pests/Bruchus_pisorum/ (accessed 211).

- Biron, D.G, B.S Landry, J.P Neron, D Coderre, and G Boivin. (2000)
“Geographical origin of an introduced pest species, *Delia radicum* (Diptera: Anthomyiidae), determined by RAPD analysis and egg micromorphology.”
Bulletin of Entomological Research 90: p23-32.
- Blouin, S. M. (2002). “Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer.” International Journal for Parasitology 32: p527–531.
- Bock, D. World distribution of field peas (map)(2005). <http://www.mpiz-koeln.mpg.de>.
- Byrne, O.M.T(2010) “Incorporation of Pea Weevil Resistance from Wild Pea (*Pisum Fulvum*) into Cultivated Pea (*Pisum Sativum*).” Edited by UWA Library - Digital Theses. Web. 07 Nov. 2010 Thesis. The University of Western Australia. 2005. <http://theses.library.uwa.edu.au/adt-WU2005.0132/> (accessed 2011).
- Capinera, J.L. (2001). “Handbook of Vegetable Pests.” San Diego: Academic Press,
- Cassens, I, P Mardulyn, and MC Milinkovitch. (2005) “Evaluating intraspecific “network” construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach?” Systematic Biology 54: p363–372.
- Caterino, M.S, S. Cho, and F.A.H Sperling. (2000) “THE CURRENT STATE OF INSECT MOLECULAR SYSTEMATICS: A Thriving Tower of Babel.” Annual Review of Entomology 45: p1-54.
- CLC bio. “CLC Main Workbench.” (2005). <http://www.clcbio.com/index.php?id=92>.
- Cognato, A.I, J. Sun, M.A Anducho-Reyes, and D.R Owen. (2005): “Genetic variation and origin of red turpentine beetle (*Dendroctonus valens* LeConte) introduced to the People's Republic of China.” Agricultural and Forest Entomology 7: p87-94.
- DAISIE. *Bruchus pisorum* "(Linnaeus, 1758)". (2008). <http://www.europe-alien.org/speciesFactsheet.do?speciesId=51116#> (accessed 2011).
- Eck, RV, and MO Dayhoff. (1966). “Atlas of Protein Sequence and Structure”. Silver Springs, Maryland: National Biomedical Research Foundation,

- Esmelealem, M and T. Adane. (2007). "Effect of harvesting and threshing time and grain fumigation of field." Proceedings of the 2nd Annual Regional Conference on Completed Crops Research Activities 18 - 21 September 2007 Bahir Dar, Ethiopia. p109-120.
- Excoffier, L, and H.E.L. Lischer. (2010) "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows." *Molecular Ecology Resources* 10: p564-567.
- Excoffier, L, and PE Smouse. (1994) "Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: Molecular variance parsimony." *Genetics* 136: p343–359.
- Felsenstein, J. (1985) "Confidence limits on phylogenies: An approach using the bootstrap." *Evolution* 39: p783-791.
- Felsenstein, J. (1981) "Evolutionary trees from dna sequences: a maximum likelihood approach." *The Journal of Molecular Evolution* 17: p368–376.
- Fitch, WM. (1971) "Towards defining the course of evolution: Minimum change for a specific tree topology." *Systematic Zoology* 20: p406-416.
- Free Phylogenetic Network Software. Fluxus Technology Ltd. 1999-2010. <http://www.fluxus-engineering.com/sharenet.htm> (accessed 2011).
- Gasser, R. B., H Min, N.B Chilton, and X Zhu. (2002): "Single –strand conformation polymorphism-based analysis of mitochondrial cytochrome c oxidase subunit 1 reveals significant sub-structuring in hookworm populations." *Electrophoresis* 23 p27-34.
- Gilbert, M.T.P, W Moore, L Melchior, and M Worobey. (2007) "DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage." *PloS ONE* 2, no. 3 (2007): e272. doi:10.1371/journal.pone.0000272.
- Grapputo, A, S. Boman, L. Lindstrom, A. Lyytinen, and J. Mappes. (2005) "The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations." *Molecular Ecology* 14: p4207-4219.
- Hardie, D.C., G.J. Baker, and D.R. Marshall. (1995) "Field screening of *Pisum* accessions to evaluate their susceptibility to the pea weevil (Coleoptera: Bruchidae)." *Euphytica* 84: p155-161.

- Hardie, D.C., and S.L. Clement. (2001) "Development of bioassays to evaluate wild pea germplasm for resistance to pea weevil (Coleoptera: Bruchidae)." *Crop Protection* 20, no. 6: p517-522.
- Higgins, D, J Thompson, T Gibson, J. D. Thompson, D. G. Higgins, and T. J. Gibson. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Research* 22: p4673-4680.
- Hu, M, N.B Chilton, X Zhu, and R.B Gasser. (2002a) "Single-strand conformation polymorphism-based analysis of mitochondrial cytochrome c oxidase subunit 1 reveals significant substructuring in hookworm populations." *Electrophoresis* 23: p27-34.
- Hu, M., J Hoglung, N.B. Chilton, X. Zhu, and R.B Gasser. (2002b) "Mutation scanning analysis of mitochondrial cytochrome c oxidase subunit 1 reveals limited gene flow among bovine lungworm subpopulations in Sweden." *Electrophoresis* 23: p3357–3363.
- Huelsenbeck, J. P., F Ronquist, R Nielsen, and J.P Bollback. (2001) "Bayesian inference of phylogeny and its impact on evolutionary biology." *Science* 294: p2310-2314.
- Hunter, M.D. (2002) "Ecological causes of pest outbreaks." *Encyclopedia of Pest management*, p214-217.
- Huson, DH. (1998) "SplitsTree: analyzing and visualizing evolutionary data." *Bioinformatics* 14: p68–73.
- Huson, DH, and D Bryant. (2006) "Application of phylogenetic networks in evolutionary studies." *Molecular Biology and Evolution* 23: p254–267.
- Ito, K, *et al.* (2011) "Analysis of genetic variation and phylogeny of the predatory bug, *Pilophorus typicus*, in Japan using mitochondrial gene sequences." *Journal of Insect Science* 11: Article 18.
- Johnson, K.P, R.G Mole, C.C Witt, R.C Faucett, and J.D. Wecstein.(2001) "Phylogenetic relationships in the louse genus *Penenirmus* based on nuclear (EF-1 α) and mitochondrial (CO1) DNA sequences." *Systematic Entomology* 26: p491-497.
- Jukes, TH, and CR Cantor. (1969) "Evolution of protein molecules." Edited by In Munro HN. *Mammalian Protein Metabolism*, (Academic Press) p21-132.

Kergoat, G.J, P Delobel, and A Delobel. (2007a) "Phylogenetic relationships of a new species of seed-beetle infesting *Cercis siliquastrum* L. in China and in Europe (Coleoptera: Chrysomelidae: Bruchinae: Bruchini)." *Ann. soc. entomol. Fr. (n.s.)* 43 (3): p265-271.

Kergoat, G.J., S. Jean-François, A. Delobel, T. Midori, and K. Anton. (2007b) "Defining the limits of taxonomic conservatism in host-plant use for phytophagous insects: Molecular systematics and evolution of host-plant associations in the seed-beetle genus *Bruchus* Linnaeus (Coleoptera: Chrysomelidae: Bruchinae)." *Molecular Phylogenetics and Evolution* 43: p251-269.

Kissi, E. (2000) "The Politics of Famine in U.S. Relations with Ethiopia, 1950-1970." *The International Journal of African Historical Studies* 33, no. 1: p113-131.

Lachowska, D, M Holecova, and M Rozek. (1998) "Chromosome Number in Nine Beetle Species (Coleoptera: Bruchidae, Apionidae, Curculionidae)." *Folia Biologica* 46, no. 1-2: p41-46.

Leff, B, N Ramankutty, and J.A Foley. (2004) "Geographic distribution of major crops across the world." *Global Biogeochemical Cycles* 18: p1-27.

Liu, J., R.E Berry, and M.S Blouin. (1999) "Molecular differentiation and phylogeny of Entomopathogenic nematodes (Rhabditida: Heterorhabditidae) based on ND4 gene sequence of mitochondrial DNA." *Journal of Parasitology* 85: p709-715.

Lucigen® Corporation. "EconoTaq plus green 2x Master Mix." www.Lucigen.com. Lucigen Corporation. 2006. <http://lucigen.com/store/docs/manuals/MA037-EconoTaq-PLUSGREEN.pdf> (accessed 2011).

Messiaen, C.M., A.A. Seif, M. Jarso, and G Keneni. (2006) *Pisum sativum* L. Record from Protabase. Edited by M. & Belay, G Brink. PROTA (Plant Resources of Tropical Africa / Ressources végétales de l'Afrique tropicale), Wageningen, Netherlands. 2006. <http://database.prota.org/search.htm> (accessed 2011).

Montesino, M, and L Prieto. (2012) "Capillary Electrophoresis of Big-Dye Terminator Sequencing Reactions for Human mtDNA Control Region Haplotyping in the Identification of Human Remains." *Methods in Molecular Biology* 830: p267-281.

NCBI. (2012) NCBI (National Center for Biotechnology Information). 15 February 2012. <ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt> (accessed 2012).

- Nei, M, and S Kumar. (2000) "Molecular Evolution and Phylogenetics". New York: Oxford University Press..
- Nei M and W. Li.(1979) "Mathematical model for studying genetic variation in terms of restriction endonucleases." Proceedings of the National Academy of Sciences. 76, no. 10: p5269-5273.
- Nicol, D, Armstrong K.F, S.D Wratten, C.M Cameron, C Frampton, and B Fenton.(1997) "Genetic variation in an introduced aphid pest (*Metopolophium dirhodum*) in New Zealand and relation to individuals from Europe." Molecular Ecology 6: p255-265.
- Okasha, S.(2008) "Population Genetics." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta.
[\url{http://plato.stanford.edu/archives/fall2008/entries/population-genetics/}}](http://plato.stanford.edu/archives/fall2008/entries/population-genetics/)
- Okumuş, İ. and Y. Çiftci. (2003) "Fish Population Genetics and Molecular Markers: II- Molecular Markers and Their Applications in Fisheries and Aquaculture." Turkish Journal of Fisheries and Aquatic Sciences 3: p51-79.
- Posada, D. and K.A Crandall. (2001): "Intraspecific gene genealogies: trees grafting into networks." TRENDS in Ecology & Evolution 16, no. 1: p37-44.
- Rambaut, Andrew. (2009) FigTree. 21 12 2009.
<http://tree.bio.ed.ac.uk/software/figtree/> (accessed 2011).
- Ratnasingham, S., and P. D. N Hebert.(2007) "BOLD : The Barcode of Life Data System." <http://www.barcodeoflife.org/content/about/what-dna-barcoding> (accessed 2011).
- Roche.(2007) "Roche High Pure PCR Template Preparation Kit". Mannheim: Roche Diagnostics GmbH, Roche Applied Science, 2007.
- Ronquist, F, and J.P Huelsenbeck. (2003) "MRBAYES 3: Bayesian phylogenetic inference under mixed models." Bioinformatics 19: p1572-1574.
- Rzhetsky, A, and M Nei. (1993) "Theoretical foundation of the minimum-evolution method of phylogenetic inference." Molecular Biology and Evolution 10, no. 5: p1073-1095.
- Saeb, A.T.S.(2006) "Phylogenetic and population genetic studies on some insect and plant associated nematodes." Dissertation, 2006:
<http://www.ohiolink.edu/etd/send-pdf.cgi/Saeb%20Amr.pdf?osu1158348092>.

Saitou, N, and M Nei. (1987) "The neighbor-joining method: A new method for reconstructing phylogenetic trees." *Molecular Biology and Evolution* 4: p406-425.

Sanger, F, S Nicklen, and AR Coulson. (1977) "DNA sequencing with chain-terminating inhibitors." *Proceedings of the National Academy of Sciences of the United States of America*: p5463-5467.

Schoville, S.D., M.Stuckey, and G.K. Roderick. (2011) "Pleistocene origin and population history of a neoendemic alpine butterfly." *Molecular Ecology* 20: p1233–1247.

Sc tterer, C. (2004) "The evolution of molecular markers - just a matter of fashion?" *Nature Reviews: Genetics* 5: p63-69.

Simmons, RB, and SJ. Weller. (2001) "Utility and evolution of cytochrome b in insects." *Molecular Phylogenetics and Evolution* 20, no. 2: p196-210.

Sneath, P, and R Sokal. (1973) "Numerical Taxonomy: The Principles and Practice of Numerical Classification." San Francisco: W.H. Freeman.

Stoeckle, M, and J.H Asubel.(2003) The Rockefeller University. 2003. phe.rockefeller.edu/barcode/docs/Barcode_of_Life.doc (accessed 2011).

Tamura, K, D. Peterson, N. Peterson, G. Stecher, M. Nei, and and S. Kumar. (2011) "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." *Molecular Biology and Evolution*, no. 28: p2731-2739.

Templeton, AR, KA Crandall, and CF Sing. (1992): "A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation." *Genetics* 132: p619–633.

Waterhouse, D.F, and D.P.A Sands.(2001) "Classical Biological Control of Arthropods in Australia." Vol. 77, in *ACIAR MONOGRAPH SERIES*, by Canberra Australian Centre for International Agricultural Research.

Werle, E, C. Schneider, M. Renner, M. V lker, and W.W.Fiehn. (1994) "Convenient single-step, one tube purification of PCR." *Nucleic Acids Research* 22, no. 20: p4354-4355.

White, T.C.R. (1984) "The abundance of invertebrate herbivores in relation to the availability of nitrogen in stressed food plants." *Oecologia* 63: p90-105.

- Williams, C. Lenney, S.L Goldson, D.B Baird, and D.W Bullock. (1994) "Geographical origin of an introduced insect pest, *Listronotus bonariensis* (Kuschel), determined by RAPD analysis." *Heredity* 72: p412-419.
- Williams, T. L., and M. E. Bernard. (2003) "An Investigation of Phylogenetic Likelihood." *Proceedings of the 3rd IEEE Symposium on Bioinformatics and BioEngineering*. Washington, DC, USA: IEEE Computer Society, 2003. 79.
- Wiser, Mark F. Tulane University. (2008).
<http://www.tulane.edu/~wiser/protozoology/notes/tree.html> (accessed 2011).
- Wright, Sewall. (1951) "The genetical structure of populations." *Annals of Eugenics* 15: p 323-354.
- Xu, Haigen, *et al.* (2006) "The distribution and economic losses of alien species invasion to China." *Biological Invasions* 8: p1495–1500.