GENETIC DIVERSITY, POPULATION STRUCTURE, MATING SYSTEM AND POLLEN FLOW IN ARBUTUS (ARBUTUS MENZIESII PURSH)

by

JACLYN DARLENE BELAND

B.Sc., The University of Guelph, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Forest Sciences)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 2004

© Jaclyn Darlene Beland, 2004
Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it
freely available for reference and study. I further agree that permission for
extensive copying of this thesis for scholarly purposes may be granted by the
head of my department or by his or her representatives. It is understood that
copying or publication of this thesis for financial gain shall not be allowed without
my written permission.

Jaelyn Beland
Name of Author (please print)

28/06/2004
Date (dd/mm/yyyy)

Title of Thesis: Genetic diversity population
structure, mating system and pollen flow in
Arbutus (Arbutus unziesii Pursh)

Degree: Master of Science Year: 2004

Department of Forest Sciences
The University of British Columbia
Vancouver, BC Canada
Abstract

Arbutus (*Arbutus menziesii* Pursh) is the only broadleaved evergreen tree native to Canada. It is a member of four natural plant communities in British Columbia (BC) considered to be at risk as identified by the BC Conservation Data Centre, the main contributing factors being urban encroachment, fire suppression, grazing and exotic invasive species. Very few studies have been conducted on this species, and no data is available on pollination biology or population genetics. Amplified fragment length polymorphisms (AFLPs) were used to conduct the first genetic examination of *A. menziesii* in BC. The study included 10 populations spread throughout the geographic range of arbutus in BC as well as one population from Washington State. Genetic diversity estimates were low (mean $H$ within populations = 0.094) relative to long-lived perennials on average (0.25); there were no significant differences among populations. Significant partitioning of genetic variation among populations was detected ($F_{ST} = 0.15$, $\phi_{ST} = 0.16$). This estimate was comparable to the average in long-lived perennials and frugivores-dispersed species ($F_{ST} = 0.19$ and 0.16, respectively). Jackknifed estimates of $F_{ST}$, a dendrogram of Reynolds' coancestry coefficient for all populations and a principal components analysis all suggested that the Gold River, BC population seemed to differ more from other populations, although this trend was not statistically significant. Isolation by distance was significant based on kinship coefficients ($p < 0.01$): half-sibs were approximately three metres apart. Mating system analysis of the Amelia Island, BC population revealed a high outcrossing rate (0.97), although 10-20% of mating events were attributed to biparental inbreeding. Pollen flow in this population was also investigated using a paternity analysis. Due to the low genetic variability detected in this species as well as the significant biparental inbreeding, the ability to confidently assign paternity was limited. Comparison of paternity assignment at LOD score threshold values of 4 and 5 revealed the characteristic leptokurtic distribution of pollen in many other plant species. The information generated from this investigation is discussed with respect to conservation strategies, and future directions in the study of the genetics of arbutus are suggested.
# Table of Contents

Abstract ................................................................................................................................. ii
Table of Contents .................................................................................................................. iii
List of Tables ........................................................................................................................ v
List of Figures ........................................................................................................................ vi
Acknowledgements ............................................................................................................. viii
Chapter 1: General introduction ........................................................................................... 1
  1.1 Ecology and biology of arbutus ....................................................................................... 1
    1.1.1 Autecology of arbutus ............................................................................................ 1
    1.1.2 Reproductive strategies ......................................................................................... 4
    1.1.3 Community Ecology ............................................................................................. 5
  1.2 Genetic variation in long-lived, woody plant populations .............................................. 8
    1.2.1 The role of genetic markers .................................................................................. 8
    1.2.2 Genetic diversity and population structure ......................................................... 10
    1.2.3 Mating system and paternity analyses .................................................................. 12
    1.2.4 Genetics and conservation of arbutus .................................................................. 14
  1.3 Objectives ..................................................................................................................... 15
Chapter 2: Materials and methods ....................................................................................... 17
  2.1 Field sampling .............................................................................................................. 17
    2.1.1 Sampling arbutus to determine genetic diversity and population structure .......... 17
    2.1.2 Sampling for mating system and paternity analyses ........................................... 18
  2.2 Seed germination and AFLP analysis ......................................................................... 22
    2.2.1 Seed germination ................................................................................................. 22
    2.2.2 DNA extraction .................................................................................................... 22
    2.2.3 AFLP protocol ..................................................................................................... 23
    2.2.4 Scoring ................................................................................................................ 25
  2.3 Statistical analysis ........................................................................................................ 26
    2.3.1 Genetic diversity and population structure ......................................................... 26
    2.3.2 Mating system and paternity analyses ................................................................. 29
Chapter 3: Results .................................................................................................................. 31
  3.1 Genetic diversity and population structure ................................................................. 31
3.2 Mating system and paternity analyses

3.2.1 Seed Germination

3.2.2 Mating system

3.2.3 Paternity analysis

Chapter 4: Discussion

4.1 Arbutus seed germination

4.2 Low genetic diversity within populations and significant structure among them

4.3 Stand level population structure supports mating system estimates

4.4 Paternity: further support for stand level genetic structure and mating system estimates

Chapter 5: Conclusions and future directions

References

Appendix A
List of Tables

Table 2.1 Name, location, sample size and tissue type collected for populations used in the analysis of genetic diversity and population structure. ................................................................. 17

Table 2.2 Description of populations sampled for the study of paternity. ........................................ 19

Table 2.3 Primer combinations used to amplify AFLPs in arbutus ................................................. 24

Table 3.1 Population-level genetic diversity. Standard errors of the estimate for $H$ are based on the variation within a single dataset. ................................................................. 31

Table 3.2 Estimates of $F_{ST}$ with corresponding populations removed and over all populations. Standard errors and 95% confidence intervals were based on jackknifing over loci and 1000 permutations, respectively. ................................................................. 36

Table 3.3 Analysis of molecular variance, with significance calculated using 1000 permutations. ............................................................................................................................ 36

Table 3.4 Distance classes relative to the kinship coefficient $F_{IJ}$ .................................................. 38

Table 3.5 Estimates of multilocus outcrossing rate ($t_m$), mean single-locus outcrossing rate ($t_s$), biparental inbreeding ($t_m - t_s$), mean parental inbreeding coefficient ($F$), correlation of selfing among loci ($r_s$), and correlation of paternity among siblings ($r_p$) ................................................................. 39

Table 3.6 Paternity assignment of 285 progeny using LOD score thresholds. ................................. 40

Table 4.1 Mean estimates of within-population genetic diversity ($H$) and among-population differentiation for three life history traits. The number of studies (N) within each group is given. ............................................................................................... 43

Table 4.2 Mean estimates of within-population genetic diversity ($H$) and among-population differentiation in different mating systems. The number of studies (N) within each group is given. ............................................................................................... 47
List of Figures

Figure 1.1 Geographic range of *Arbutus menziesii* (modified from McDonald & Tappeiner, 1990). ................................................................. 2

Figure 1.2 Expected percent areal cover of arbutus in BC with respect to both its overall and stand-level presence (map courtesy of A. Hamann). ................................................................. 3

Figure 2.1 Populations collected for the study of genetic diversity and population structure. Full population reference name for each acronym is given in Table 2.1 (base map by Generic Mapping Tools (GMT), www.aquarius.geomar.de/omc/). ................................................................. 18

Figure 2.2 Populations collected for the study of paternity (modified from Fisheries and Oceans Canada, A) 1985 and B) 1997 respectively). ................................................................. 20

Figure 2.3 A) Location of samples on Amelia Island. The section indicated with a filled rectangle corresponds to the sampled area (modified from Fisheries and Oceans Canada, 1997). B) Plot of individuals sampled on Amelia Island. Arrows indicate where trees were not sampled due to the inability to reach the fruit. ................................................................. 21

Figure 2.4 An example of AFLP profiles amplified with the same primer combination 
(PstI+AG/MseI+CG) using DNA isolated from vegetative bud tissue (A) and seedling DNA (B). Arrows indicate bands present in the seedling profile that do not amplify when vegetative bud tissue is used ................................................................. 26

Figure 3.1 Scatter plot of geographic variables with respect to average heterozygosity within populations. A) latitude; B) longitude ................................................................. 32

Figure 3.2 UPGMA dendrogram illustrating the relationship among populations with respect to Reynolds’ et al. (1983) coancestry distance. The percentage of similar replicates supporting each node was obtained from 1000 bootstraps. The full population reference name for each acronym is given in Table 3.1 ................................................................. 33

Figure 3.3 a) The first two principal components of the genetic variation in populations of arbutus, explaining 14% of the total variation in the data. The full population reference name for each acronym is given in Table 3.1 ................................................................. 34

Figure 3.3 b) The first two principal components of the genetic variation in populations of arbutus, illustrating clusters apparent for GR, QB and the MH, MD and GI groupings. The full population reference name for each acronym is given in Table 3.1 ................................................................. 35

Figure 3.4 Relationship between kinship coefficient and the logarithm of physical distance showing a significant negative correlation ($R^2 = 0.84$, $p < 0.01$). ................................................................. 37

Figure 3.5 Distribution of mating events over distance as determined by paternity analysis at two LOD score threshold values, 4 (A) and 5 (B). ................................................................. 41
Figure 4.1 Map illustrating the location of the Muchalat Inlet. The Gold River population is indicated with a circle (modified from Fisheries and Oceans Canada, 2002).
Acknowledgements

I would like to acknowledge several people who were instrumental in the completion of this thesis. First and foremost I would like to thank my best friend and partner in life, Marty Megens. Without his support and understanding in my moments of frustration and self-doubt the completion of this work would not have been as positive an experience. I dedicate this thesis to him. Also in this vein, I would like to thank my friends and family for their encouragement. I would like to express my gratitude to my supervisor, Dr. Yousry El-Kassaby, and the members of my committee, Drs Sally Aitken, Carol Ritland and Kermit Ritland, for being open to my ideas and giving me valuable advice that kept me on the right path. In particular Carol, Allyson Miscampbell, and Cherdsak Liewlaksaneeyanawin were essential to my completion of the labwork component of this research project. Kermit helped immensely in clarifying the finer details of data analysis. Allyson, Cherdsak and Jennifer Wilkin were also of excellent aid in the field. Jodie Krakowski was an invaluable field assistant and editor of this thesis, and I thank her for all of her thoughtful input, encouragement, and for her friendship. My labmates in the genetics group, Hugh Wellman, Yanik Bérubé, Dilara Ally, Cherdsak Liewlaksaneeyanawin, Charles (Chin-Lin) Chen, Makiko Mimura, Washington Gapare, Andrew Bower, and in particular Jennifer Wilkin, created a positive, fun learning environment that I could not have done without. This work was funded by a Natural Science and Engineering Research Council of Canada grant to K. Ritland and by the Johnson Family Biotech Endowment to Y. El-Kassaby.
Chapter 1: General introduction

1.1 Ecology and biology of arbutus
1.1.1 Autecology of arbutus

*Arbutus menziesii* Pursh, a member of the family Ericaceae, is one of approximately 10 species of trees and shrubs in this genus, most of which are native to the western hemisphere (Luteyn, 2002). *A. menziesii* is the only species native to Canada (Farrar, 1995) and is one of three that are endemic to North America (Roy, 1974). The species was named after its discoverer, Archibald Menzies (1754-1842), a Scottish naturalist and physician who accompanied Captain Vancouver on his travels throughout the Pacific Northwest. It is also the only native broadleaved evergreen tree in Canada (dropping its leaves during their second growing season) (Farrar, 1995), and is the least frost resistant tree species in Canada (Krajina, 1969).

*A. menziesii* is found along the Pacific coast, from San Diego County, California (33°N) to south coastal British Columbia (BC) extending to its northernmost limit at around 50°N. It is most commonly found on western slopes at low elevations in BC to increasingly higher elevations in the Coast ranges of northern California; however, populations do occur in the Klamath, southern Cascade, and Sierra Nevada mountains (Figure 1.1; McDonald and Tappeiner, 1990).

Although McDonald and Tappeiner (1990) note its expansive latitudinal range, its distribution in BC is limited to areas in and around the Strait of Georgia, with some additional populations found along drainages on both Vancouver Island and the mainland. This distribution roughly overlaps the Coastal Douglas-fir (CDF) biogeoclimatic zone. In fact, approximately half of all arbutus trees in the province are in this zone, while the other half are dispersed within drier subzones of the coastal western hemlock (CWH) zone (Hamann et al., 2002). Figure 1.2 illustrates the expected percent areal cover of arbutus in BC with respect to both its overall and stand-level presence.

Situated in the rainshadow of Vancouver Island and the Olympic mountains, the CDF zone is characterized by warm, dry summers and mild, wet winters. In this temperate Mediterranean climate, temperatures rarely fall below freezing during winter months, and the mean annual precipitation can vary from 66 to 152cm (Krajina, 1969). Across the entire
geographic range for arbutus, extreme temperatures can range anywhere from -21 to 46°C and mean annual rainfall from 46 to 422cm (Tarrant, 1958, c.f. McDonald and Tappeiner, 1990).

Figure 1.1 Geographic range of *Arbutus menziesii* (modified from McDonald & Tappeiner, 1990).
Although arbutus has been identified on many different soil types, in its northern extent it is generally found on nutrient poor deposits of well-drained sands, gravels, and tills with low water retention capacity. Arbutus is drought tolerant (Morrow and Mooney, 1974), and has deep, spreading lateral root system that can penetrate up to 3.5 m of fractured bedrock (Niemec et al., 1995). It can persist on a variety of different terrain, typically on southern and western aspects. For example, in southern California it is usually found growing in cool canyons, whereas in BC it is most common on rocky slopes and outcrops. While a shrubby appearance generally results from poor growing conditions, arbutus can grow to a medium sized tree up to 34 m in height. The oldest have been dated to 200 to 250 years old, and some extremely large individuals have been estimated to be 400 to 500 years old (McDonald and Tappeiner, 1990). Arbutus has low shade tolerance, but this varies with latitude (McDonald and Tappeiner, 1990): in BC it has a very low shade tolerance in the driest sites of the CDF and CWH, while elsewhere it has been referred to as shade intolerant (Krajina, 1969).
1.1.2 Reproductive strategies

Most arbutus trees produce flowers and fruit every year over most of its geographic range. Its small, perfect, white urn-shaped flowers bloom from approximately March to June on dense terminal racemes. The stamens attach below the superior ovary and are shorter than the white, five-lobed corolla from which the style protrudes (Luteyn, 2002). Dead fruit stalks can be persistent and new branchlets originate from buds at the base of the fruit stalk. During abundant flowering years, most branchlets may only produce reproductive buds, causing branch dieback to the next main branch with vegetative buds (Hunt et al., 1992).

No formal pollination biology studies have been conducted on arbutus. Although members of the Ericaceae characteristically have a nectary disc at the base of the corolla (Woodland, 1991), this has yet to be proven for arbutus (Gurung et al., 1999). In general, superior-ovaried Ericaceous species are bee-pollinated while those with an inferior ovary are hummingbird-pollinated (Luteyn, 2002). Bee-pollinated flowers are usually blue or yellow, however some flowers have an ultra-violet spectrum that attracts bees (Meeuse, 1961). It is possible that this is the case for arbutus since bees are attracted to arbutus blossoms (Gurung et al., 1999) and apiarists produce arbutus honey. Rufous and Anna’s hummingbirds (Selasphorus rufus and Calypte anna, respectively) have also been observed feeding on arbutus flowers, implying a nectar reward, and could therefore also serve as pollinators (Gurung et al., 1999).

The arbutus fruit is a rough glandular berry eight to 12 mm in diameter, which turns from green to a bright red when it ripens, generally in mid-September to mid-November. The earliest fruit bearing age is three to five years and berries can contain from two to 37 seeds, with an average of 20 (Roy, 1974; McDonald and Tappeiner, 1990). In general, good crops occur every couple of years and light crops approximately every ten (McDonald and Tappeiner, 1990; McDonald, 1992; Niemec et al., 1995), although in the southern parts of the range heavy crops may be produced every 10 years (Niemec et al., 1995).

Arbutus seedlings are not abundant. Seeds are dispersed by a variety of frugivores including birds, deer, and rodents, while they also often fall to the ground beneath the tree and remain there (gravity dispersed). Germination occurs the following spring and with up to 90% success. Establishment is most common in the partial shade of disturbed areas, on bare mineral soil, or in semi-open forests (McDonald and Tappeiner, 1990). Seedlings emerge within a month, but survival after the first growing season is quite low: from 0 to 10 percent (Tappeiner et
Subsequent growth is 2.5 to 3.0 cm over two years (Tappeiner et al., 1986). Studies by Pelton (1962) and Tappeiner et al. (1986) determined a number of factors contributing to seedling mortality including low soil moisture, litterfall, damping-off fungi, invertebrate browsing (mainly slugs), deer browsing and frost.

Arbutus reproduction is primarily vegetative. Epicormic sprouts are produced from adventitious buds near the root collar, which are often associated with an extensive root burl. Relatively frequent fire, typical of arbutus habitat, readily destroys the thin-barked stems. Disturbance maintains populations since sprouts emerge vigorously following fire or cutting, and grow very rapidly, benefiting from water and nutrients provided by the roots of the mother tree (McDonald and Tappeiner, 1990).

1.1.3 Community Ecology

Arbutus is a member of four natural plant communities categorized by the BC Conservation Data Centre (CDC) (MSRM, 2003) as at risk. These include: 1) arbutus – hairy manzanita (A. menziesii – Arctostaphylos columbiana), 2) Douglas-fir – arbutus (Pseudotsuga menziesii – A. menziesii), 3) Garry oak – arbutus (Quercus garryana – A. menziesii), 4) Douglas-fir – lodgepole pine – arbutus (P. menziesii – Pinus contorta – A. menziesii). The first three are considered to be endangered or threatened in BC; the last is categorized as being of special concern. The Sensitive Ecosystems Inventory (SEI) (MSRM, 2001) of eastern Vancouver Island and the Gulf Islands has also identified arbutus as a dominant tree species in the coastal bluff and woodlands ecosystems of this area. The main factors contributing to the uncertain future of these communities are urban encroachment, fire suppression, grazing and exotic invasive species.

The ability of arbutus to reproduce by sprouting makes it a competitive species during secondary succession. Older stands tend to reflect a successional subclimax (McDonald and Tappeiner, 1990). Arbutus often grows in groves since shoots that sprout near the root collar are able to grow along rock crevices. Clumps regularly reach three metres in height and two to three metres in diameter within three years of dieback (Tappeiner et al., 1984). These seemingly young individuals could in fact be quite old since the main stem can die back several times, followed by rounds of sprouting (Hunt et al., 1992).

Pure stands of arbutus are rare, usually resulting from a severe disturbance of competing Douglas-fir. The importance of frequent fires to sustain arbutus populations was demonstrated in
a fourteen-year survey by Hunter (1997) of two old-growth Douglas-fir – tanoak (*Lithocarpus densiflorus*) forests in California in which arbutus ranked third in basal area. For the duration of the study, arbutus populations experienced no recruitment and had the highest mortality rate (8.4%/yr), most of which died due to lack of light. Tanoak was the only species colonizing canopy gaps, suggesting that in the absence of fire the low shade tolerance of arbutus will lead to its decline when growing within a forest including more shade tolerant species. This phenomenon has also been observed in a study of gap dynamics in an old-growth forest in coastal California that had not experienced fire for 60-80 years (Hunter *et al.*, 1999). Here, although arbutus was creating 35% of the gaps recorded it was colonizing only 4%. Arbutus recruitment may also be low in old-growth forests since germination requires bare mineral soil and in the absence of fire the litter layer would be thick in undisturbed areas.

Although successful Douglas-fir regeneration often occurs under canopies of arbutus (Hunter and Barbour, 2001), ways in which to quantify, reduce, or eliminate sprouting of arbutus in Douglas-fir clear-cuts has been the primary focus of some studies (Harrington *et al.*, 1984; Tappeiner *et al.*, 1984; Hughes *et al.*, 1990). Douglas-fir regeneration can be shaded out by vigorously sprouting arbutus. However, sprouting clumps of arbutus also provide temporary shelter for young conifer seedlings in hot, dry conditions (Minore, 1985). Arbutus' high drought tolerance gives it a competitive advantage over Douglas-fir under water stress conditions (Pabst *et al.*, 1990; Wang *et al.*, 1995; Zwieniecki and Newton, 1996). The influence of arbutus litter on Douglas-fir regeneration has also been of interest to foresters. Del Moral and Cates (1971) demonstrated in the laboratory that arbutus litter produced allelopathic compounds that inhibited germination of Douglas-fir seedlings; however subsequent research groups could not confirm this finding (Tinnin and Kirkpatrick, 1985; Minore, 1987).

Arbutus also has a symbiotic relationship with arbutoid mycorrhizal fungi which create sheaths around the root tips and colonize root cells, altering root morphology by creating a short, stubby appearance. This association increases a plant’s access to soil nutrients while the fungus receives carbon from the plant. Drought and pathogen resistance are also potential benefits for the plant. A study by Elliot *et al.* (2001) found soils near unhealthy arbutus to have higher net nitrification rates than soils near healthy arbutus. It was concluded that this could be due to reduced mycorrhizal colonization of the roots when nitrogen in the soil is increased. Fungi which form arbutoid mycorrhizae also have ectomycorrhizal associations with other tree species
(Molina and Trappe, 1982), where sheaths are formed only around the root tips. Lack of host specificity by particular ectomycorrhizal fungi creates the potential for woody understory plants left behind following conifer harvest, such as arbutus, to act as a reservoir of beneficial fungi accessible to conifer regeneration (Molina and Trappe, 1982; Acsai and Largent, 1983; Massicotte et al., 1999; Hagerman et al., 2001). In fact, Simard et al. (1997) demonstrated that in the field carbon can be transferred between Douglas-fir and birch (Betula papyrifera) seedlings since ectomycorrhizal fungi of these species can be connected by a common mycelial network. It is also possible that this occurs between arbutus and Douglas-fir (Trudell et al., 1999).

A variety of wildlife forages on and nests in arbutus trees. The fruits are a food source for several species of birds, which in turn disperse the seed. Some studies have found that populations of the American robin (Turdus migratorius) (Raphael, 1999), the varied thrush (Ixoreus naevius) (Hagar, 1960; Raphael, 1999) and the band-tailed pigeon (Columba fasciata) (Chappell and Giglio, 1999) fluctuate in synchrony with the abundance of arbutus fruit crops. Deer also browse the berries and young shoots.

Arbutus is also utilized by a number of cavity-nesting birds. In a study of a Douglas-fir–tanoak–arbutus forest, arbutus was the preferred species of the primary cavity nesters the red-breasted sapsucker (Sphyrapicus ruber) and the hairy woodpecker (Picoides villosus). Cavity nests were found in arbutus more frequently than expected based on tree abundance and size distributions. It was suggested that frequent heartwood decay might explain this preference for arbutus (Raphael, 1999). These nests often provide habitat for secondary cavity nesters such as small owl species, raccoons (Procyon lotor), and northern flying squirrels (Glaucomys sabrinus) (Gurung et al., 1999).

During the past few years concern has been rising in some urban centres that arbutus populations have begun to decline. Tree ring patterns suggest that the growth of arbutus is highly sensitive to environmental conditions (Ettl, 1999). It is particularly susceptible to damage caused by freezing and the brittle branches are easily broken under the weight of heavy snow or by winds.

Since arbutus has low shade tolerance in BC and Washington, self-pruning is common within stands. These tall, spindly trees are susceptible to sunscald when exposed resulting in periderm cracks. This kind of injury, whether it stems from mechanical damage, freezing, or
sunscald, creates optimal conditions for the invasion of pathogens (Adams et al., 1999; Bressette and Hamilton, 1999; Elliot, 1999a,b; Hunt, 1999). Although 83 fungi and 15 chewing insects are associated with arbutus (Lowe, 1977; Ginns, 1986; c.f. Hunt et al., 1992) only three agents are responsible for the most severe impacts. *Natrassia mangiferae* (formerly *Hendersonula toruloidea* Natrass (Sutton and Dyko, 1989)) is common in oceanside populations and is commonly named arbutus canker. This pathogen invades existing wounds and attacks the cambium, creating a sunken canker. Attempting to contain the infection, the tree usually forms a callused ridge along the margin. Cankers can result in the dieback of branches by girdling, and a large canker on the main stem can kill the tree (Hunt et al., 1992; Elliot, 1999a,b; Hunt, 1999). When the arbutus canker spreads rapidly the margin may be smooth, which can result in confusion with cankers caused by *Fusicoccum aesculi*. This fungus moves from the branch tips inwards and results in black, cracked stems resembling burned wood. The most common root disease associated with arbutus is *Phytophthora cactorum* root rot. The invasion of this fungus is often indicated by cankers at the base of the trunk that appear water-soaked (Elliot, 1999a,b). Trees growing in water-logged soils are the most susceptible (Hunt et al., 1992; Elliot, 1999a,b; Hunt, 1999). These conditions are common in urban environments since soils are often compacted and lawns copiously watered. This inhospitable root environment may result in profuse flowering, creating additional dieback.

1.2 Genetic variation in long-lived, woody plant populations

1.2.1 The role of genetic markers

Genetic markers have greatly facilitated the study of genetic variation in plant populations for decades. Although this avenue of study began with morphological markers, the advent of protein-based markers (i.e. allozymes) as neutral, co-dominant genetic markers saw the field of evolutionary ecology expand at an unprecedented rate. Allozymes have been assayed in many plant species and comprehensive reviews of the literature have identified trends in levels of genetic diversity and population genetic structure corresponding to different life history traits (Hamrick and Godt, 1996).

Although allozymes are still used to study genetic variation in plants, the innovation of the polymerase chain reaction (PCR) technique (Saiki et al., 1985; Mullis and Faloona, 1987) facilitated the development of several types of DNA-based markers: randomly amplified
polymorphic DNA (RAPD) (Williams et al., 1990, 1993), amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995) and microsatellites (also referred to as simple sequence repeats (SSRs)) (Morgante and Olivieri, 1993). These marker types have advantage over allozymes since they are able to produce many highly variable loci.

Microsatellites (SSRs) are the marker of choice for many plant researchers since they are co-dominant, stable, and highly variable. Identification of these chromosomal regions and designing of primers for amplification is an expensive procedure, therefore substantial funding and justification for their development are usually necessary. Both RAPDs and AFLPs are easily applied to new species since DNA sequence knowledge is not needed. These markers have the disadvantage of dominant expression (homozygous dominant and heterozygous genotypes are indiscernible at a locus) however this reduced amount of information can be compensated for by large numbers of loci distributed throughout the genome (Ritland and Ritland, 2000; Mariette et al., 2002a). Reproducibility of results among these three markers was investigated by a network of European laboratories (Jones et al., 1997): RAPD profiles were difficult to reproduce, AFLPs were highly conserved, and there were slight discrepancies between laboratories in the sizing of SSR alleles.

The suitability of a DNA marker to any study depends on the species and research goals. There are reports of studies where estimates of genetic diversity in natural populations using different markers in the same species vary widely. This sparked the recent effort by Mariette et al. (2002a) to compare estimates of within-population diversity under different evolutionary scenarios using AFLPs and microsatellites by way of simulation. Microsatellites exhibited more variable results over different evolutionary scenarios, while AFLPs technique produced less variation among its predictions. For example, when migration rates were high microsatellites were the best predictors of genomic heterogeneity; however they were poor when migration rates were low. For AFLPs, predictions of genomic heterogeneity were fairly consistent over various migration rates. In a comparative study of RAPD, AFLP, and microsatellite markers for the estimation of within- and among-population genetic diversity, Nybom (2004) found the three markers yielded similar results with respect to among-population diversity estimates while microsatellite markers yielded values almost three times higher for within-population diversity. Moreover, estimates of diversity summarized over different life history traits using microsatellite and RAPD markers gave similar patterns to those surveyed in allozymes (Hamrick and Godt,
RAPD-based estimates of within-population genetic differentiation were directly comparable to those calculated from allozymes.

In summary, one must be cautious when comparing statistics derived from different types of genetic markers as the results are not always directly equivalent. Given the objectives of this study, (see section 1.3) estimation of all population parameters will be based on AFLPs.

1.2.2 Genetic diversity and population structure

Insight into levels of genetic diversity and population structuring provides important baseline information since genetic variation is the basis for evolution and reflects a species' history and evolutionary dynamics. Genetic variation in traits that confer fitness allow a species to adapt to changing environments. As evolutionary processes such as genetic drift, mutation, and selection affect populations, substructuring can also occur (genetic differentiation among subpopulations). Levels of variation in selectively neutral markers like AFLPs indicate potential for variation in these adaptive traits. Although most studies of plant species using AFLPs have been on herbaceous plants because they are often used as model systems, it is increasingly being used to study genetic variability in tree species (Van der Merwe et al., 2000; Zawko et al., 2001; Mariette et al., 2002b; Rivera-Ocasio et al., 2002; Cavers et al., 2003; Peakall et al., 2003; Tang et al., 2003).

Literature surveys of marker data in plant species with a range of life history traits indicate that woody plants have more genetic diversity within species and within populations than non-woody plants with similar life history traits. Angiosperms are more heterozygous than gymnosperms with respect to within species heterozygosity ($H_e$) and long-lived perennials with ingestion as a seed dispersal mechanism have significantly higher within-species heterozygosity than those that are dispersed by gravity or wind (Hamrick et al., 1992; Hamrick and Godt, 1996; Nybom, 2004). Since arbutus is thought to be both bee- and hummingbird-pollinated, the comparison of animal-pollinated versus wind-pollinated species is also relevant: genetic diversity tends to be greater in the former (Hamrick et al., 1992). One might therefore expect arbutus to be more genetically diverse relative to plant species in general.

How genetic diversity is partitioned among populations is also paramount in understanding a species' historical and evolutionary dynamics. In particular, genetic
Differentiation among populations can aid in determining historical range expansion. Many researchers have studied post-Pleistocene range expansion in temperate tree species of the Pacific Northwest (reviewed by Soltis et al., 1997). As the spatial dynamics of range expansion are largely governed by seed dispersal (Ouborg et al., 1999), a “leading edge hypothesis” (Hewitt, 1993, c.f. Soltis et al., 1997) where northward long-distance seed dispersal events from southern refugia form a leading edge of migration may explain genetic differentiation in populations. Since individuals at the leading edge would be more likely to recolonize a newly deglaciated niche, these populations would have undergone a succession of bottlenecks. A series of founder events such as this can leave behind colonization footprints characterized by a reduction in genetic variation (Hartl and Clark, 1997; Austerlitz et al., 2000). Data for several plant species exhibit a trend showing lower levels of genetic variability in northern populations (north of the Oregon-California state border) of plants in the Pacific Northwest (Soltis et al., 1997). It is important to note, however, that not all plant species in the Pacific Northwest exhibit this trend (e.g. Sitka spruce, see Gapare, 2003).

Studies of genetic variability using a variety of marker systems report lower levels of genetic differentiation among populations of long-lived woody species than for short-lived or annual plants regardless of the seed dispersal mechanism (Nybom 2004; Hamrick and Godt, 1996). This difference is usually explained by pollen flow being higher in trees. By way of simulation Austerlitz et al. (2000) refuted this showing that it is in fact due to differences in life cycle. They hypothesize that founder effects in tree species are markedly less pronounced since reproductive maturity is delayed compared to annual plants. This results in population growth due to incoming seed flow rather than via offspring of the primary colonizers, thus increasing diversity and decreasing population differentiation. The first generation of offspring is therefore a result of mating among a more genetically variable population and the founder effect is reduced. This may also be the case for arbutus since it does not become reproductively mature until at least three to five years of age. No information exists to date on seed flow in arbutus, therefore colonization rates of new niches are unknown. This would affect the patterns of mating among primary colonizers.

Spatial genetic structure within a stand of trees is also an important aspect of population dynamics since together with modes of pollen and seed dispersal it affects mating patterns and the action of natural selection (Epperson, 1992). When pollen and seed dispersal distances are
limited, a plant population is expected to develop distinctive patch structure (spatial autocorrelation of genotypes) and genetic isolation by distance. This familial clumping has been shown for a number of both coniferous and angiosperm tree species (see Epperson, 1992 for a review).

1.2.3 Mating system and paternity analyses

The mating system of a plant species has a direct impact on the population genetic composition since it influences how genetic diversity is partitioned within and among populations (Wright, 1978; Barrett, 2003). Outcrossing species maintain more genetic diversity within populations than those that are predominantly selfing, while interpopulation differences are reduced in outcrossing species. Many species have mixed mating systems, a combination of selfing and outcrossing, in which the partitioning of variation can be more complex. Mating patterns are influenced by both intrinsic and extrinsic factors, including post-zygotic inbreeding depression, and pollinator type and abundance. These interrelated factors constitute a plant breeding system (Sage et al., in press) subject to selective pressures which impact a species' evolutionary potential (Charlesworth and Charlesworth, 1981).

A recent survey of the literature by O'Connell (2003) gave a mean outcrossing estimate of 0.887 for angiosperm trees. Mixed mating is supported by empirical evidence as surveys of outcrossing rates across animal-pollinated species reveal a continuous distribution (Barrett and Eckert, 1990; Vogler and Kalisz, 2001, c.f. Barrett, 2003). Estimates of shared paternity among siblings (correlated mating) and mating among relatives (biparental inbreeding) (Ritland, 1989, 2002) provide additional measures of the mating system to advance our understanding of genetic differentiation within and among populations.

As mentioned previously, the spatial scale of gene exchange between individuals is governed by pollen dispersal. It therefore plays an important role in the structure and maintenance of genetic variation. Estimates of pollen dispersal can be obtained by assigning paternity to progeny in natural populations of plants (reviewed by Jones and Arden, 2003). This can be difficult since many factors can affect pollen dispersal and pollination success, including plant abundance and density, pollinator behavior, environmental conditions, phenology, pollen viability and genetic incompatibility systems. Many pollen dispersal studies result in leptokurtic
curves for dispersal distance, indicating most pollen does not move far from its source, but a very small proportion can move a great distance. (Sage et al., in press)

Paternity exclusion analyses can measure gene flow directly using molecular makers and comparing multilocus genotypes for each maternal plant in a population with that of its progeny. Gene flow can then be interpreted as the fraction of seeds sired by individuals outside of the study population (Ellstrand, 1992; Chase et al., 1996; Krauss, 1999, 2000a). Paternity analyses are not always so straightforward, however, as the number of polymorphic genetic markers available for a species many not be sufficient to resolve all relationships. Meagher (1986) first developed a maximum likelihood method to assign paternity to the most-likely male using isozyme data and many studies have used his approach (Meagher, 1991; Adams et al., 1992; Devlin et al., 1992; Kaufman et al., 1998; Apsit et al., 2001). Methods for determining statistical confidence for likelihood-based paternity assignment using co-dominant markers (Marshall et al., 1998) and maximum likelihood methods for determining parentage when using dominant genetic markers have since been developed (Gerber et al., 2000).

No formal pollination biology data is available for arbutus, making it difficult to predict patterns of gene flow influenced by pollen dispersal. General knowledge of the foraging habits and distances for both bumblebees (Bombus spp.) and hummingbirds, the probable pollinators of arbutus, will clarify the definition of ‘isolated’ populations and interpretation of paternity analysis.

It has been generally accepted that bumblebees forage close (less than 50m) to their nests and concentrate their efforts in patches, returning to the same patch and moving mostly from flower to flower. Bumblebees are central-place foragers and optimal foraging models predict they will forage in a manner allowing net energy return in the form of nectar to maximize success. Since flying is energetically costly it follows that bumblebees should minimize their travel from the hive to nectar sources (Dramstad, 1996). A mark-reobservation study by Dramstad (1996) provided evidence that bumblebees often fly from 360 to 600m to forage. Foraging habitats may also be species-specific. A comparison of three Bombus species sharing a nest site with equal food availability found all B. muscorum workers within 500m of their nest, 43% of B. terrestris foraged within 500m and 25% as far as 1500 to 1750m away, and 78% of B. lapidarius foraged within 500m and approximately 9% from 1001 to 1500m away from the nest (Walther-Hellwig and Frankl, 2000). A more recent study using microsatellites (Chapman et al.,
2003) has shown median foraging distances for *B. terrestris* and *B. pascuorum* to be as far as 0.62-2.8 km and 0.51-2.3 km, respectively.

Pollen dispersal by hummingbirds is more extensive in non-territorial species that forage on scattered forest plants than along forest edges pollinated by territorial species (Linhart, 1973, *c.f.* Webb and Bawa, 1983). A comparison of the flight patterns of broad-tailed (*Selasphorus platycercus*) and rufous (*S. rufus*) hummingbirds in patches of *Delphinium* species found the more territorial rufous hummingbirds to move shorter distances between inflorescences than the broad-tailed species. Territorial behaviour was found to have a greater impact on foraging habit than the spatial arrangement of nectar resources (Kotliar, 1992). Webb and Bawa (1983) reported pollen dispersal distances for *Malvaviscus arboreus*, a patchily distributed Costa Rican shrub (or small tree) with hermaphroditic flowers pollinated by hummingbirds. Dispersal distance followed a leptokurtic curve, with pollen rarely transported further than 100 m to a maximum of 225.5 m.

Pollen carryover by bumblebees and the rufous hummingbird is both density- and species-dependent (Ellstrand, 1992). Its pattern has been described as erratic to a steep, exponential decay (Lertzman, 1981). In general, however, pollen and gene flow studies indicate highly leptokurtic distributions (Ellstrand, 1992; Williams, 2001). This trend is the expected dispersal pattern in arbutus.

### 1.2.4 Genetics and conservation of arbutus

Arbutus is a diploid species with *n=13* (Stebbins and Major, 1965, *c.f.* Sorensen, 1995). Phylogenetic studies based on morphological characters and molecular evidence show strong support for the Arbutoideae subfamily forming a distinct monophyletic clade within the family Ericaceae (Kron *et al.*, 2002). Phylogenetic resolution of the Arbutoideae itself indicates that *Arbutus* species from western North America are genetically distinct from those endemic to the Mediterranean basin (Hileman *et al.*, 2001).

To date, no other genetic information exists for *A. menziesii* aside from these data. As previously noted, knowledge of genetic differentiation within and among populations, mating system and pollination biology is important to our understanding of evolutionary processes that a species may undergo. It follows then, that these data are also valuable from a conservation standpoint. Maintaining of appropriate levels of genetic diversity for species to ensure long-term
persistence is important to any strategy to conserve that ecosystem's biodiversity. As noted earlier (see section 1.1.3), the ecosystems in which arbutus thrives in British Columbia are endangered or threatened, or at the very least of special concern. Although the range of arbutus stretches from BC to Mexico (Figure 1.1) this does not devalue it in ecosystems of BC since the biodiversity of any ecosystem can be defined as “the source of biological resources” (Wood, 1997). These biological resources create the economic, recreational, aesthetic and cultural resources of our surrounding environment, to which we are so tightly linked and dependent upon (Wood, 1997).

Although arbutus is not used in commercial forestry, the aesthetic appeal and quality of its high density red wood makes it popular for furniture, flooring and paneling, while its dramatic root burls are used to make items such as tables and clocks (Niemiec et al., 1995). The Coast Salish First Nations of the CDF zone on Vancouver Island use its bark and leaves for many different medicinal purposes (Turner and Hebda, 1990).

1.3 Objectives

As there are no data available to date on the population genetics or pollination biology of arbutus, the general aim of this project will be to elucidate information on these two broad topics. The results can then be incorporated into potential conservation strategies for arbutus. My specific objectives and expectations are as follows:

1. To characterize levels of genetic diversity and population structure in arbutus. I expect this species to have high levels of within population diversity and low genetic differentiation among populations, as this is the case for most long-lived woody species.

2. To describe patterns of mating. Due to the floral structure of arbutus, a survey of the literature on outcrossing rates in angiosperm trees, and reports of it being bee-pollinated and potentially hummingbird-pollinated, I hypothesize that arbutus will have a mixed mating system and be primarily outcrossing.
3. To determine patterns of gene flow by pollen within populations using a paternity analysis. Since the behavior of pollinators of arbutus have been documented to yield leptokurtic distributions of pollen dispersal, and flowers bloom on dense terminal racemes, I anticipate that the greatest proportion of pollen will not move far from its source.
Chapter 2: Materials and methods

2.1 Field sampling
2.1.1 Sampling arbutus to determine genetic diversity and population structure

Young leaves or vegetative bud tissues were randomly sampled from a total of 269 individuals spread across 11 locations within the northern range of arbutus, in BC and Washington State, to represent its geographic distribution (Table 2.1, Figure 2.1). All sampled individuals within each population were spatially mapped. Locations were 135 km apart on average (range: 16 to 373 km). The average distance between trees within a population was 35 m (range: two to 100 m). Since arbutus has a continuous yet patchy distribution and is bee- and potentially hummingbird-pollinated this sampling strategy is expected to encompass a good representation of the populations sampled. The health and size of a tree did not influence sampling decisions and tissue was collected from only one stem when multiple stems were clumped together to avoid sampling an individual more than once. All samples were placed on ice and transported back to the lab. The materials were stored at -20°C until DNA was extracted. Once all samples were isolated, field materials were placed into vials for permanent storage at -80°C.

Table 2.1 Name, location, sample size and tissue type collected for populations used in the analysis of genetic diversity and population structure.

<table>
<thead>
<tr>
<th>Population (acronym)</th>
<th>Latitude (°N)</th>
<th>Longitude (°W)</th>
<th>N</th>
<th>Tissue Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortes Island (CI)</td>
<td>50.073</td>
<td>124.982</td>
<td>30</td>
<td>vegetative bud</td>
</tr>
<tr>
<td>Powell Lake (PL)</td>
<td>49.887</td>
<td>124.544</td>
<td>30</td>
<td>vegetative bud</td>
</tr>
<tr>
<td>Gold River (GR)</td>
<td>49.677</td>
<td>126.103</td>
<td>30</td>
<td>vegetative bud</td>
</tr>
<tr>
<td>Horseshoe Bay (HB)</td>
<td>49.362</td>
<td>123.277</td>
<td>26</td>
<td>leaf</td>
</tr>
<tr>
<td>Qualicum Beach (QB)</td>
<td>49.357</td>
<td>124.429</td>
<td>18</td>
<td>leaf</td>
</tr>
<tr>
<td>Amelia Island (AI)</td>
<td>49.303</td>
<td>124.153</td>
<td>30</td>
<td>leaf</td>
</tr>
<tr>
<td>Port Alberni (PA)</td>
<td>49.247</td>
<td>124.839</td>
<td>30</td>
<td>vegetative bud</td>
</tr>
<tr>
<td>Galiano Island (GI)</td>
<td>48.961</td>
<td>123.516</td>
<td>30</td>
<td>leaf</td>
</tr>
<tr>
<td>Malahat Rise (MH)</td>
<td>48.576</td>
<td>123.526</td>
<td>30</td>
<td>leaf</td>
</tr>
<tr>
<td>Mount Douglas Park (MD)</td>
<td>48.497</td>
<td>123.342</td>
<td>30</td>
<td>leaf</td>
</tr>
<tr>
<td>Vashon Island (VI)</td>
<td>47.381</td>
<td>122.405</td>
<td>30</td>
<td>leaf</td>
</tr>
</tbody>
</table>

1 number of individuals sampled
2 in Washington State
2.7.2 Sampling for mating system and paternity analyses

As a paternity analysis was initially the primary goal, effort was made to locate isolated populations of arbutus from which all individuals in the population could be sampled. All open-pollinated siblings collected were then also used to study the mating system of arbutus. Given the results of a literature survey (see section 1.2.3) on the foraging habits of the pollinators of arbutus, bees and hummingbirds, I defined an "isolated" population as being approximately 500 m from individuals outside the study population. Since arbutus grows in a patchy, continuous...
fashion it proved difficult to identify such an isolated population on Vancouver Island. An attempt was therefore made to locate Gulf Islands with no more than approximately 30 reproductive individuals. Five candidate populations were identified in the summer of 2002 from a boat with binoculars (Table 2.2, Figure 2.2).

Table 2.2 Description of populations sampled for the study of paternity.

<table>
<thead>
<tr>
<th>Population</th>
<th>Closest adjacent population (m)</th>
<th>N(^1) germinable seed</th>
<th>N leaf tissue and fruit samples collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Island</td>
<td>440</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Ragged Islets</td>
<td>600</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Reef</td>
<td>440</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Ada Islands</td>
<td>450</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Amelia Island</td>
<td>600</td>
<td>80</td>
<td>40</td>
</tr>
</tbody>
</table>

\(^1\)N = number of individuals

Returning to the islands to collect fruit and leaf tissue in the fall proved exceedingly difficult because of the weather. When sampled, the actual number of reproductive individuals on the islands differed from initial estimates. It was not possible to collect the fruit from some trees due to inaccessibility. On many of the smaller islands (most notably Ada Islands) the fruit on the trees had become hard, black, filled with mould, and the seeds had dried up. There were no germinable seeds. Consequently, Ada Islands had only one individual bearing germinable seed, rendering it unusable for this study. The fruit collected from the smaller islands appeared of poorer quality than those collected on Amelia Island as they were relatively tough, a dull shade of red and the seeds were smaller. This may have been caused by the environment: vegetation on the small islands was more windblown, probably experienced more salt spray and there was little soil, and therefore relatively less available water and nutrients. All leaves and fruit collected were stored on ice until returning from the field. As with the samples collected for genetic diversity and population structure estimates, the materials were stored at -20°C until DNA was extracted. Surplus samples were then placed into vials for permanent storage at -80°C. The fruit collected was stored at 4°C until the seeds were removed.
Figure 2.2 Populations collected for the study of paternity (modified from Fisheries and Oceans Canada, A) 1985 and B) 1997 respectively).
I initially expected that the islands with few individuals (all islands but Amelia Island) could be used to test the ability to discern the paternity of progeny, while Amelia Island would be the main study population. Unfortunately, what appeared to be 40 reproductive individuals on Amelia Island from the boat in the summer turned out to be approximately 80 when a ground survey was conducted in the fall. Most of the trees were situated along the east and west sides of the island on cliffs along the rocky shoreline. Due to weather and logistical constraints, only 40 individuals were collected along the west side of the island. Some trees along the sampling path were not collected since fruit was too high (Figure 2.3).

It was hypothesized that although not all individuals were collected, identification of paternity would be possible since it is more likely that successful pollination occurred among individuals close to one another rather than via pollen flow from the eastern shoreline of the island as there was a largely arbutus-free ridge in the centre of the island.

**Figure 2.3** A) Location of samples on Amelia Island. The section indicated with a filled rectangle corresponds to the sampled area (modified from Fisheries and Oceans Canada, 1997). B) Plot of individuals sampled on Amelia Island. Arrows indicate where trees were not sampled due to the inability to reach the fruit.
2.2 Seed germination and AFLP analysis

2.2.1 Seed germination

Seed was extracted from the fruit manually by squashing the berry and using forceps to pull out seed. In the fruit from all islands, but less so on Amelia, the fruit was permeated with mould or fungus. Sticky pulp adhered to the seed coat so seeds were soaked in water for approximately one minute to facilitate the removal of pulp from the seed coat. They were then left to dry in paper envelopes at room temperature for one day before being sealed in plastic bags and stored at 4°C until use.

Arbutus seeds must be cold-stratified for at least 30 to 60 days for successful germination (Maleike and Hummel, 1999). To begin this process, 20 to 30 seeds from each open-pollinated family were imbibed for 24 hours, then placed on filter paper in a Petri dish using sterile technique. Autoclaved deionized distilled water was then used to dampen the filter paper. This was followed by a fungicide treatment with No-Damp (Plant-Prod, Brampton, ON). All seeds were then stored at 4°C and Petri dishes periodically monitored to ensure the filter paper remained damp. Some dishes became visibly infected with fungus during cold-stratification. In these cases, sterile technique was used to reintroduce the seeds into the same environment as described above. At the end of the cold-stratification period the Petri dishes were placed on the bench top at room temperature for seed germination.

2.2.2 DNA extraction

DNA (deoxyribonucleic acid) was extracted from all samples using modifications of the CTAB (cetyltrimethyl ammonium bromide) method (Doyle and Doyle, 1990). Approximately 0.5 g of leaf tissue was ground in liquid nitrogen and incubated at 65°C for one hour in 20 mL of CTAB isolation buffer (1.5% CTAB, 0.1M Tris base (pH 8.0), 0.02 M EDTA (pH 8.0), 1.4 M NaCl, 0.2% β-mercaptoethanol). Following centrifugation the supernatant was incubated at 37°C for 45 minutes with 0.2 mg of RNase A. Each sample was then mixed with an equal volume of chloroform:isoamylalcohol (24:1) for 30 minutes to remove the CTAB-protein complex. Following centrifugation the DNA was precipitated from the supernatant for one hour at −20°C using 2/3 by volume of cold isopropanol. The DNA pellet was then washed twice with ice-cold 70% ethanol, then air-dried, and finally dissolved in from 50-200 μL of sterilized deionized distilled water (sddH₂O) depending on the size of the pellet.
DNA extractions from the vegetative buds (approximately 1/2 bud) and 15 progeny per parent (open-pollinated family) were ground in 1.5 mL tubes and incubated in 1 mL of CTAB isolation buffer using only 10 μg of RNase A per sample. Seedling isolations were left overnight in isopropanol at -20°C to maximize yield. Vegetative buds were dissolved in 50 μL of sddH₂O. Since little DNA was extracted from individual seedlings, pellets were dissolved in 15 μL sddH₂O. The quality of all samples was evaluated by spectrophotometry.

2.2.3 AFLP protocol

A study by Mariette et al. (2002a) suggests that when using AFLPs, at least 100-200 loci (including both polymorphic and monomorphic) should be used for an accurate estimation of genetic diversity. To attain this, the aim was to use roughly four primer pairs (approximately 50 scorable bands per primer set). With respect to the number of loci needed to resolve paternity, Gerber et al. (2000) suggest those with allele frequencies of the presence band between 0.1 and 0.4 are the most informative and that about 100-200 loci should be adequate for studies of parentage (although this is obviously species-dependent).

The AFLP protocol (Vos et al., 1995) used in this study was modified by the Genetic Data Centre (GDC) at UBC (Vancouver, BC), adapted from Cervera et al. (2000). The restriction enzyme PstI was used as the rare cutter and Msel as the frequent DNA cutter (Roche, Laval, QB). The other key enzymes used in the protocol, Taq DNA polymerase and T4 DNA ligase, were obtained from Roche and Invitrogen (Burlington, ON), respectively. PCR (polymerase chain reaction) was run in an MJ Research PTC-100 thermal cycler (Watertown, MA) and stored at -20°C until electrophoresis was performed. All samples were loaded on 0.4 mm thick 6% (Long Ranger™) polyacrylamide gels 25 cm in length. DNA fragments were separated using a LiCor 4200 automated sequencer (LiCor Inc., Lincoln, NE).

Approximately 60 primer combinations were tested for genetic diversity analysis using one individual from each of five populations and 500 ng of DNA per reaction. The top 12 primers based on clarity, resolution of loci and approximate level of polymorphism were screened using five individuals per population for three populations spread across the geographic range. Four final primer combinations were then chosen (see Table 2.3) and tested for repeatability using three individuals. DNA from each individual was subjected to the entire
protocol three times. The primer combinations were also tested for tissue differences between leaves and vegetative buds by comparing AFLP profiles for each tissue in six individuals.

Since the AFLP profiles in each population were scored prior to paternity analysis it was evident that more primers should be screened to potentially increase the number of polymorphic bands. To effectively increase the probability that paternity could be resolved, 30 additional primers were tested and three final primer combinations chosen (see Table 2.3). The data gathered for progeny using these final primers were also used for mating system analysis. For all AFLP combinations the primers used in the first round of selective PCR (pre-amplification) were the same and the $P_srl$ primers tailed so that the fragments could be visualized on the gel using an M13 infrared labeled primer (LiCor Inc.).

<table>
<thead>
<tr>
<th>Primer combinations</th>
<th>$P_srl$ primer (tailed)</th>
<th>$MseI$ primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic diversity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-amplification</td>
<td>$P_srl$ + A</td>
<td>$MseI$ + C</td>
</tr>
<tr>
<td>final amplifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>combination 1</td>
<td>$P_srl$ + AG</td>
<td>$MseI$ + CG</td>
</tr>
<tr>
<td>combination 2</td>
<td>$P_srl$ + AA</td>
<td>$MseI$ + CG</td>
</tr>
<tr>
<td>combination 3</td>
<td>$P_srl$ + AA</td>
<td>$MseI$ + CAG</td>
</tr>
<tr>
<td>combination 4</td>
<td>$P_srl$ + AG</td>
<td>$MseI$ + CGA</td>
</tr>
<tr>
<td>Paternity and mating system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-amplification</td>
<td>$P_srl$ + A</td>
<td>$MseI$ + C</td>
</tr>
<tr>
<td>final amplifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>combination 1</td>
<td>$P_srl$ + AG</td>
<td>$MseI$ + CG</td>
</tr>
<tr>
<td>combination 2</td>
<td>$P_srl$ + AG</td>
<td>$MseI$ + CC</td>
</tr>
<tr>
<td>combination 3</td>
<td>$P_srl$ + AA</td>
<td>$MseI$ + CAC</td>
</tr>
</tbody>
</table>

Since the seedling isolations yielded very low concentrations of DNA, it was not possible to standardize the amount of DNA added to each AFLP reaction at 500 ng as was done with the leaf and vegetative bud tissues. Different amounts of seedling DNA as well as a standard amount of 100 ng were tested to determine the relative intensities and repeatability of the AFLP profiles across this range and the minimum amount that could be used. Based on this, it was decided that no less than 80 ng of DNA could be used per AFLP reaction, and that a range of amounts could be used since the profile intensities remained constant and repeatable relative to those samples standardized at 100 ng.
2.2.4 Scoring

AFLP profiles were scored for presence (1) or absence (0) of bands at loci using SAGA version 2.1 (LiCor Inc.). All bands visualized were scored as present regardless of their relative intensities. Only loci deemed to be reliably scorable (based on clarity and ease of scoring) across all gels for all populations were scored. Individuals that were missing at least three bands at a given locus within a primer combination that were otherwise monomorphic across all populations were discarded. This was most likely caused by poor DNA quality in these samples, as AFLPs are very sensitive. Repeatability testing identified one locus in each of two primers (PstI + AG/Msel+ CG and PstI + AG/Msel+ CGA) used that were not reliable and therefore not scored. Inspection for tissue differences revealed seven loci across two primer pairs (PstI + AG/Msel + CG and PstI + AA/Msel + CAG) that differed in state between leaves and vegetative buds and were therefore also excluded.

All AFLPs in the populations collected for the analysis of genetic diversity and population structure were scored prior to beginning to work with the seedling DNA. I began by screening the seedlings with the primer combination (PstI+AG/Msel+CG) that was also used to analyze populations for genetic diversity and population structure. This is when I discovered that the profiles contained numerous loci that had not been detected in any other collected population or in any of the maternal trees from which the seeds were collected (see Figure 2.4). It seemed exceedingly unlikely that these loci could actually be amplified arbutus DNA if they had not been detected anywhere else. Occasionally the anomalous profiles also appeared to be consistent within families. Given the sensitivity of this technique to DNA quality, and the wide range of conditions among families with respect to both their health and fungal contamination (see Results), only loci that were present in the parents across all primers were scored. Although this could technically exclude new immigrants to Amelia Island the probability of this occurring would be very small since only one out of all loci scored in the populations analyzed for genetic diversity was detected solely in one population. In addition, samples that had strikingly different profiles than those of the parent population were discarded since my confidence in their authenticity was low.
Figure 2.4 An example of AFLP profiles amplified with the same primer combination (PstI+AG/MseI+CG) using DNA isolated from vegetative bud tissue (A) and seedling DNA (B). Arrows indicate bands present in the seedling profile that do not amplify when vegetative bud tissue is used.

2.3 Statistical analysis

2.3.1 Genetic diversity and population structure

Measures of genetic variation using neutral markers are based on allele frequencies, commonly calculated under the Hardy-Weinberg equilibrium (HWE) model and its assumptions, as was the case here. The program TFPGA version 1.3 (Miller, 1997) was used to estimate allele frequencies using the method of Lynch and Milligan (1994) designed for dominant markers. In addition to HWE, the authors assume that AFLPs produce two alleles per locus and that the banding profiles can be interpreted in an unambiguous fashion. It is suggested that loci with a null homozygote (band absence) frequency less than 3/N be excluded from the analysis to avoid bias in estimation; however, this too can lead to a biased estimate of gene frequency (Krauss, 2000b; Mariette et al., 2002b). Since the level of polymorphism detected in AFLP profiles of arbutus was low, removing loci according to the 3/N criterion would have a greater impact on the result than it would if the profiles had been highly polymorphic. This is because the proportion of loci to be removed would be far greater when there are few polymorphic loci to begin with. Given these somewhat contradictory suggestions and the nature of the data in this study, all final loci scored were included in the allele frequency estimations.
Population allele frequencies were then used to estimate a number of commonly reported genetic statistics using the same program. These included: Nei’s (1978) unbiased average heterozygosity \((H)\), the percent of loci polymorphic (PLP) including all loci scored, and Reynolds’ et al. (1983) coancestry distance \((D)\) among populations. This particular distance measure is designed to measure genetic divergence among populations due solely to genetic drift and is therefore an appropriate measure for short-term evolution. Populations are assumed to be non-inbred, finite in size, in HWE at each locus and in linkage equilibrium at every pair of loci. Student’s t-Test was used to test for significant differences in average heterozygosity among populations with a probability of type I error \((\alpha)\) of 0.05. Regression analysis was used to determine if average heterozygosity was correlated with either latitude or longitude.

Polymorphic loci were also classified across populations as being either widespread or localized. This terminology is usually associated with the sampling of alleles and linked to categories of alleles considered to be common or rare (Brown and Hardner, 2000). Since this study used dominant markers, rather than considering the dominant (presence band) and recessive (absence band, or null homozygote) alleles, loci were categorized to avoid the counterintuitive notion of categorizing the absence of an allele (the null homozygote). The frequencies at which loci were considered to be widespread versus localized was somewhat subjective. In this case, loci with a frequency of occurrence < 0.5 were considered to be localized while those present in populations at frequencies \(\geq 0.5\) were deemed widespread following Gapare’s (2003) classification. This analysis is not meant to be used for ex situ sampling, but rather as a tool with which to further describe the populations in this study.

A UPGMA (unweighted pair-group method with arithmetic averaging) cluster analysis was used to generate a dendrogram illustrating the relationship of genetic distance among populations (using TFPGA). Support for the branching pattern was determined by 1000 bootstrap permutations of the data. A regression of pairwise population genetic and physical distances was conducted to determine if there was a significant relationship. A Mantel test (Mantel, 1967) was also done in this program (one matrix randomized 1000 times) to determine if there was a significant association between genetic and geographic distance. A principal components analysis (PCA) using SAS version 8.2 (SAS Institute, Cary, NC) was carried out for comparison with the UPGMA cluster analysis.
Using Weir and Cockerham’s (1984) methods, \( F_{ST} \) was estimated with TFPGA on a population basis, as well as jackknifing over populations, to determine the relative contributions of each population to the overall estimate (El-Kassaby and Yanchuk, 1995). \( F_{ST} \) was introduced in Wright’s infinite island model (1951). The model assumes that a species is divided into an infinite number of equal-sized randomly mating islands that exchange migrants at a constant rate. It is also assumes that populations have reached an equilibrium between gene flow and genetic drift, and that there is no mutation or selection. Neigel (2002) provides an excellent summary of the impact of relaxing these assumptions under different scenarios and the subsequent effect on estimates of \( F_{ST} \), hence this will not be explored here.

\( F_{ST} \) is calculated from allele frequencies and is often referred to as a standardized measure of the variance in allele frequencies among populations relative to the total genetic variation. Weir and Cockerham (1984) define \( F_{ST} \) (denoted \( \theta \)) as the correlation of genes among populations (coancestry). In other words, it is a description of how genetic variation is partitioned across versus within populations and therefore is a relative measure of population divergence. It became ubiquitous in the literature with the advent of allozymes and continues to be widely used, making it a useful measure to compare population structure in different species.

Jackknifing over loci was conducted to obtain variance estimates and bootstrapping (1000 replications) to generate a 95% confidence interval. Significance of population differentiation was also assessed statistically using a Markov Chain Monte Carlo approach to approximate Fisher’s exact test (1000 dememorization steps, 20 batches, 2000 permutations/batch). With diploid dominant data, this analyzes each locus to determine if there are significant differences in band frequencies among populations, assuming loci are independently sorting (Raymond and Rousset, 1995).

Population-level differentiation was also assessed by an analysis of molecular variance (AMOVA) (Excoffier et al., 1992) using the program Arlequin version 2.000 (Schneider et al., 2000) so it could be compared with the value calculated based on allele frequencies. This method uses a pairwise Euclidean squared distance matrix (number of mismatches/total # of bands, between every individual in the analysis) to partition the total variance into covariance components for the desired level of hierarchical analysis. These components can then be used to compute \( \phi \)-statistics analogous to Wright’s \( F \)-statistics. Significance of the fixation index \( \phi_{ST} \) was tested using a hierarchical non-parametric permutation approach (1000 permutations). It is
important to note that data from dominant molecular markers is treated as being haplotypic (one allele per locus, present or absent) for any analysis using Arlequin. (Schneider et al., 2000).

Spatial genetic structure of within stands was explored, using the program SPAGeDi version 1.1 (Hardy and Vekemans, 2003), by analyzing how pairwise genetic relatedness varied between individuals with respect to physical distance using regression analysis. Hardy’s (2003) kinship coefficient adapted to dominant markers was used. Kinship coefficients can be defined as the probability of identity-by-descent (IBD) of two randomly chosen genes at a locus in two different individuals. More specifically, the kinship coefficient used in this study, $F_{ij}$ (i.e. the kinship coefficient between individuals $i$ and $j$), is interpreted as a ratio of differences in probability of IBD between homologous genes and defined as $F_{ij} = Q_{ij} - \bar{Q} / 1 - \bar{Q}$, where $Q_{ij}$ is the probability of IBD for random genes between $i$ and $j$, and $\bar{Q}$ the probability of IBD between random genes within a ‘reference population’ (the sample population). This measure does not assume genotypes are in HWE proportions but requires knowledge of the inbreeding coefficient, estimated according to the methods in the following section.

Given the large standard errors often associated with pairwise coefficients, comparisons within all populations were pooled to increase the sample size within a distance interval. For example, 10 pairwise comparisons in a particular distance class in each of two populations would be pooled for a total of 20 observations. Interpopulation comparisons were avoided by setting the distance between each population at one kilometre. Distance intervals were chosen by considering both the number of pairwise comparisons in a class as well as total distance over which kinship estimates were to be made, with the latter taking precedence. The estimates were jackknifed over loci to determine the associated standard error and 1000 random permutations were performed to test for significance. Regression analysis of the kinship coefficients over the logarithm of physical distance was performed since theoretical models of isolation-by-distance predict that this relationship is approximately linear in a two-dimensional space (Rousset, 1997).

2.3.2 Mating system and paternity analyses

Mating system statistics were generated using a modified version of the maximum likelihood based program MLTR version 2.4 (Ritland, 2002) for use with dominant markers, excluding loci with a null homozygote frequency greater than 0.95 or less than 0.05. The
assumptions of the original model of Ritland and Jain (1981) on which this program is based are as follows: all maternal genotypes have the same outcrossing rate to a homogenous pollen pool, loci are in linkage equilibrium as well as HWE, and there is no mutation or selection between the time of mating and the progeny assay. Estimates of the following parameters were obtained: the multilocus \( t_m \) and mean single locus \( t_s \) population outcrossing rates, the mean single locus inbreeding coefficient of maternal parents \( F \), and the correlation of outcrossed paternity within progeny arrays \( r_p \). This latter measure estimates the probability that a given pair of progeny in an array are full sibs. A measure of biparental inbreeding was calculated as the difference between the multilocus and mean single locus outcrossing rates \( t_m - t_s \). An estimate of the correlation of selfing \( r_s \) gave an approximation of the fraction of selfing due to uniparental inbreeding. This also provided an additional measure of biparental inbreeding as the remaining fraction \( 1 - r_s \) could be attributed to this type of mating. The Newton-Raphson (NR) numerical method was chosen to solve the maximum likelihood equation, pollen gene frequencies were constrained to equal ovule frequencies, and 100 bootstraps were performed (with the progeny array as the sampling unit) to determine the standard errors of the estimates. All significance tests were Student's t-Tests with a probability of type I error of 0.05.

The recently developed program FaMoZ (Gerber et al., 2003), designed to accommodate dominant markers, was used to assign paternity with a categorical maximum likelihood technique, which assigns the most likely father from a pool of non-excluded fathers. This method assumes that the loci used in the analysis are not linked. To assign paternity, a log of the odds ratio (LOD) score was calculated for each possible father. Paternity was evaluated and compared at two LOD score threshold values, 4 and 5. A LOD score of 4 corresponds to the father being 10 000 times more likely to be the true father than a randomly chosen one while a LOD score of 5 would make a chosen father a 100 000 times more likely than a randomly chosen one from the paternal pool. The difference in LOD scores between the first and second most-likely fathers is also provided. A scoring error of 1% and the inbreeding coefficient estimated as described in the mating system analysis were incorporated into calculations. The distance between pairs of mates was graphed to illustrate pollen flow.
Chapter 3: Results

3.1 Genetic diversity and population structure

Table 3.1 summarizes the within-population genetic differentiation based on all 247 final loci scored, using allele frequencies estimated by the Lynch and Milligan (1994) method for dominant markers. There were a total of 12 different localized loci (frequency < 0.5) found across populations with an average of four per population. The percentage of loci that were polymorphic ranged from 25.1 to 33.2 (mean = 28.4) (Table 3.1). The populations collected on Galiano Island and the Malahat Rise had the highest average heterozygosity ($H = 0.106 \pm 0.011$, mean $\pm$ SE, in both populations) while those from Powell Lake and Gold River had the lowest ($H = 0.083 \pm 0.010$ and $H = 0.083 \pm 0.011$, respectively) (Table 3.1); however, heterozygosity was not significantly different between any pair of populations (paired t-test, $\alpha = 0.05$). The overall average heterozygosity within populations was 0.094 (Table 3.1). There were also no significant geographic trends for heterozygosity (versus latitude: $R^2 = 0.0504$, $p = 0.51$; versus longitude: $R^2 = 0.1661$, $p = 0.21$) (Figure 3.1).

Table 3.1 Population-level genetic diversity. Standard errors of the estimate for $H$ are based on the variation within a single dataset.

<table>
<thead>
<tr>
<th>Population (acronym)</th>
<th>N (^{1})</th>
<th>No. of polymorphic loci</th>
<th>PLP (^{2})</th>
<th>$H^\circ$ (SE)</th>
<th>No. of localized loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortes Island (CI)</td>
<td>30</td>
<td>64</td>
<td>25.9</td>
<td>0.089 (0.011)</td>
<td>0</td>
</tr>
<tr>
<td>Powell Lake (PL)</td>
<td>30</td>
<td>62</td>
<td>25.1</td>
<td>0.083 (0.010)</td>
<td>2</td>
</tr>
<tr>
<td>Gold River (GR)</td>
<td>27</td>
<td>62</td>
<td>25.1</td>
<td>0.083 (0.011)</td>
<td>1</td>
</tr>
<tr>
<td>Horseshoe Bay (HB)</td>
<td>16</td>
<td>69</td>
<td>27.9</td>
<td>0.094 (0.011)</td>
<td>4</td>
</tr>
<tr>
<td>Qualicum Beach (QB)</td>
<td>11</td>
<td>63</td>
<td>25.5</td>
<td>0.102 (0.012)</td>
<td>2</td>
</tr>
<tr>
<td>Amelia Island (AI)</td>
<td>28</td>
<td>78</td>
<td>31.6</td>
<td>0.097 (0.011)</td>
<td>5</td>
</tr>
<tr>
<td>Port Alberni (PA)</td>
<td>28</td>
<td>65</td>
<td>26.3</td>
<td>0.087 (0.011)</td>
<td>1</td>
</tr>
<tr>
<td>Galiano Island (GI)</td>
<td>22</td>
<td>82</td>
<td>33.2</td>
<td>0.106 (0.011)</td>
<td>10</td>
</tr>
<tr>
<td>Malahat Rise (MH)</td>
<td>24</td>
<td>79</td>
<td>32.0</td>
<td>0.106 (0.011)</td>
<td>7</td>
</tr>
<tr>
<td>Mount Douglas Park (MD)</td>
<td>24</td>
<td>76</td>
<td>30.8</td>
<td>0.097 (0.011)</td>
<td>9</td>
</tr>
<tr>
<td>Vashon Island (VI)</td>
<td>29</td>
<td>72</td>
<td>29.1</td>
<td>0.087 (0.010)</td>
<td>4</td>
</tr>
<tr>
<td>Overall average</td>
<td>24</td>
<td>70</td>
<td>28.4</td>
<td>0.094 (0.011)</td>
<td>4</td>
</tr>
</tbody>
</table>

\(^{1}\)number of individuals  
\(^{2}\)percentage of polymorphic loci  
\(^{3}\)Nei's (1978) unbiased average heterozygosity
Figure 3.1 Scatter plot of geographic variables with respect to average heterozygosity within populations. A) latitude; B) longitude.

The UPGMA dendrogram (Figure 3.2) using Reynolds’ et al. (1983) coancestry distance ($D$) somewhat supported the clustering of all populations except Gold River (61% of bootstrap replicates). With Gold River forming its own monophyletic clade, there was strong support (90% of bootstrap replicates) that Qualicum Beach differed from the remaining populations.
Two other clusters were also weakly supported: the Malahat Rise, Mount Douglas, and Galiano Island populations (67% of replicates) and Cortes Island, Port Alberni, and Amelia Island populations (56% of replicates). While Gold River differed the most from all populations, the difference between Gold River and Qualicum Beach was the greatest ($D = 0.411$). A Mantel test conducted on the genetic and physical distance matrices found no significant association ($r = 0.239$, $p = 0.18$).

![UPGMA dendrogram illustrating the relationship among populations with respect to Reynolds' et al. (1983) coancestry distance. The percentage of similar replicates supporting each node was obtained from 1000 bootstraps. The full population reference name for each acronym is given in Table 3.1.](image)

Figure 3.2 UPGMA dendrogram illustrating the relationship among populations with respect to Reynolds' et al. (1983) coancestry distance. The percentage of similar replicates supporting each node was obtained from 1000 bootstraps. The full population reference name for each acronym is given in Table 3.1.

The principal components analysis supported the genetic differentiation of the Gold River and Qualicum Beach populations from all other populations, as well as the clustering of the Malahat Rise, Mount Douglas and Galiano Island populations. There was no detectable grouping of any other populations. The first two principal components explained only 14% of the variation in the data (see Figures 3.3 a, b).
Figure 3.3 a) The first two principal components of the genetic variation in populations of arbutus, explaining 14% of the total variation in the data. The full population reference name for each acronym is given in Table 3.1.
Figure 3.3 b) The first two principal components of the genetic variation in populations of arbutus, illustrating clusters apparent for GR, QB and the MH, MD and GI groupings. The full population reference name for each acronym is given in Table 3.1.
The $F_{ST}$ estimate ($0.149 \pm 0.019$, mean $\pm$ SE) indicated moderate levels of genetic differentiation among populations (Table 3.2). When $F_{ST}$ was estimated by jackknifing over populations, Gold River was shown to contribute the most to the overall estimate as the value for $F_{ST}$ was lowest with this population removed. $F_{ST}$ also decreased when Qualicum Beach was removed, although to a lesser extent. There was no significant difference between the estimate of $F_{ST}$ including all populations and estimates with any population removed (Table 3.2).

**Table 3.2** Estimates of $F_{ST}$ with corresponding populations removed and over all populations. Standard errors and 95% confidence intervals were based on jackknifing over loci and 1000 permutations, respectively.

<table>
<thead>
<tr>
<th>Population</th>
<th>$F_{ST}$ (SE)</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>0.156 (0.018)</td>
<td>[0.193, 0.123]</td>
</tr>
<tr>
<td>PL</td>
<td>0.158 (0.019)</td>
<td>[0.200, 0.123]</td>
</tr>
<tr>
<td>GR</td>
<td>0.125 (0.018)</td>
<td>[0.163, 0.096]</td>
</tr>
<tr>
<td>HB</td>
<td>0.152 (0.019)</td>
<td>[0.195, 0.119]</td>
</tr>
<tr>
<td>QB</td>
<td>0.141 (0.019)</td>
<td>[0.180, 0.109]</td>
</tr>
<tr>
<td>AI</td>
<td>0.157 (0.020)</td>
<td>[0.200, 0.122]</td>
</tr>
<tr>
<td>PA</td>
<td>0.153 (0.018)</td>
<td>[0.191, 0.121]</td>
</tr>
<tr>
<td>GI</td>
<td>0.158 (0.020)</td>
<td>[0.201, 0.126]</td>
</tr>
<tr>
<td>MH</td>
<td>0.159 (0.020)</td>
<td>[0.204, 0.124]</td>
</tr>
<tr>
<td>MD</td>
<td>0.155 (0.020)</td>
<td>[0.199, 0.123]</td>
</tr>
<tr>
<td>VI</td>
<td>0.151 (0.020)</td>
<td>[0.192, 0.118]</td>
</tr>
<tr>
<td>All Populations</td>
<td>0.149 (0.019)</td>
<td>[0.192, 0.119]</td>
</tr>
</tbody>
</table>

1Full population reference name for each acronym is given in Table 3.1.

An analysis of molecular variance (AMOVA) similarly found the $F_{ST}$ analog $\phi_{ST}$ described 16% of the genetic variation as being among populations ($p < 0.00001$) (Table 3.3). The approximation of Fisher's exact test of the presence and absence band frequencies at loci also indicated significant-among population diversity ($p < 0.0001$).

**Table 3.3** Analysis of molecular variance, with significance calculated using 1000 permutations.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>variance component</th>
<th>% variation</th>
<th>$\phi$-statistic</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>10</td>
<td>2.219</td>
<td>15.95</td>
<td>$\phi_{ST} = 0.160$</td>
<td>$&lt; 0.00001$</td>
</tr>
<tr>
<td>Within populations</td>
<td>258</td>
<td>11.688</td>
<td>84.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>268</td>
<td>13.907</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spatial population genetic structure at the stand level as described by a regression analysis of the kinship coefficient ($F_{ij}$) on the logarithm of physical distance showed a significant negative correlation ($R^2 = 0.84, p < 0.01$) (Figure 3.4). Individuals within populations, pooled over sample locations, were compared within seven distance classes (Figure 3.4, Table 3.4). Estimates of kinship were significant in all classes except the largest, which included pairwise comparisons of trees 126 to 200 m apart (Table 3.4). An extrapolation of the regression line suggests that individuals approximately three metres apart have an average kinship coefficient of half-sibs, 0.125 (confidence interval: [0.186, 0.066], $\alpha = 0.05$; Figure 3.4).

![Figure 3.4](image)

**Figure 3.4** Relationship between kinship coefficient and the logarithm of physical distance showing a significant negative correlation ($R^2 = 0.84, p < 0.01$).
Table 3.4 Distance classes relative to the kinship coefficient $F_{xy}$.

<table>
<thead>
<tr>
<th>Distance class</th>
<th>Mean distance (m)</th>
<th>Mean log$_{10}$-distance (m)</th>
<th>N$^1$</th>
<th>% partic$^2$</th>
<th>$F_{xy}$ (SE)$^3$</th>
<th>p-value$^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-20</td>
<td>10</td>
<td>1.01</td>
<td>862</td>
<td>88.5</td>
<td>0.0919 (0.010)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>21-35</td>
<td>27</td>
<td>1.44</td>
<td>631</td>
<td>84.4</td>
<td>0.0793 (0.010)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>36-50</td>
<td>42</td>
<td>1.62</td>
<td>434</td>
<td>81.4</td>
<td>0.0718 (0.012)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>51-75</td>
<td>62</td>
<td>1.79</td>
<td>640</td>
<td>70.3</td>
<td>0.0670 (0.012)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>76-100</td>
<td>87</td>
<td>1.94</td>
<td>279</td>
<td>47.2</td>
<td>0.0455 (0.013)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>101-125</td>
<td>110</td>
<td>2.04</td>
<td>104</td>
<td>26.4</td>
<td>0.0536 (0.017)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>126-200</td>
<td>145</td>
<td>2.16</td>
<td>38</td>
<td>7.8</td>
<td>0.0228 (0.024)</td>
<td>0.1708</td>
</tr>
</tbody>
</table>

$^1$Number of pairwise comparisons within each distance class.
$^2$Proportion of all individuals represented at least once within each class.
$^3$Standard errors of the estimate were found by jackknifing over loci.
$^4$Significance was assessed by 1000 permutations of the data.

3.2 Mating system and paternity analyses

3.2.1 Seed Germination

Percent germination overall was quite variable, but relatively consistent within families. In some instances the radicle had already emerged from the seed coat during cold-stratification, while in most cases it took three to five days for the seeds to germinate. Some families germinated very slowly while in others only the radicle and hypocotyl would emerge. In these cases, many of the germinants began to degrade in physical condition relatively quickly. In addition, signs of fungal contamination were evident in some dishes but absent from others.

Both germinant vigour and possible fungal contamination may have affected DNA quality (see Discussion). This may explain why, as described previously in section 2.2.4, the AFLP profiles in the progeny contained numerous loci that had not been detected in any other collected population or in any of the maternal trees from which the seeds were collected (see Figure 2.4). As a result, only a portion of the progeny from Amelia Island were used for mating system and paternity analysis and none of the progeny collected from the other islands.
3.2.2 Mating system

Table 3.5 summarizes the mating system analysis of arbutus based on 44 polymorphic loci in 21 open-pollinated families. This included 285 progeny with a mean of 14 individuals within a family (range: nine to 15). The mean single-locus \( (t_s) \) and multilocus \( (t_m) \) outcrossing rates were high: 0.869 and 0.970, respectively. The mean single-locus inbreeding coefficient of the maternal parents \( (F) \), 0.047, did not differ significantly from zero. The proportion of matings due to biparental inbreeding, 0.101, as estimated by the difference between the multilocus and single-locus outcrossing rates \( (t_m - t_s) \) was significantly different from zero. The correlation of selfing \( (r_s) \) among loci indicated that approximately 80% of selfing events were true selfing events (uniparental), and the remaining 20% were due to biparental breeding. The correlation of outcrossed paternity \( (r_p) \), 0.094, was significant but low, indicating that most open-pollinated sibs do not share fathers.

<table>
<thead>
<tr>
<th>Mating system parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_m )</td>
<td>0.970</td>
<td>0.018</td>
</tr>
<tr>
<td>( t_s )</td>
<td>0.869</td>
<td>0.032</td>
</tr>
<tr>
<td>( t_m - t_s )</td>
<td>0.101</td>
<td>0.027</td>
</tr>
<tr>
<td>( F )</td>
<td>0.047</td>
<td>0.042</td>
</tr>
<tr>
<td>( r_s )</td>
<td>0.793</td>
<td>0.188</td>
</tr>
<tr>
<td>( r_p )</td>
<td>0.094</td>
<td>0.014</td>
</tr>
</tbody>
</table>

\(^1\)Standard errors were obtained from 100 bootstrap estimates.

3.2.3 Paternity analysis

The resolution of 63 polymorphic loci for determining paternity in the 285 progeny analyzed was limited resulting in my inability to assign statistical confidence. Appendix A presents the raw data summarized herein. LOD score thresholds of 4 and 5 assigned paternity to 49 and 22 progeny, respectively. At a LOD score threshold of 4, 44 of the 49 mating pairs identified also had likely second fathers. The mean difference in the LOD scores between the first and second most likely fathers \( (\Delta) \) was 2.8 meaning that the second father was 600 times
less likely than the first father (Table 3.6). At the LOD score threshold of 5, 19 of the 22 mating pairs identified had second most likely fathers with Δ being 3.3 or 2200 times less likely (Table 3.6). The number of selfing events detected at the low threshold was nine over four individuals, while four selfing events over two individuals were identified at the high threshold.

Figure 3.5 illustrates the frequency of mating events with respect to distance between mating pairs, assigned at the two LOD score threshold values. The mean distance between mates, and therefore the mean pollen flow distance, was 29 and 22 m at the threshold values of 4 and 5, respectively (Table 3.6). The distance at which the greatest proportion of mating events took place was similar at the two threshold comparison values (LOD score 4 threshold: 20 m; LOD score threshold value 5: 18 m) (Table 3.6, Figure 3.5).

Table 3.6 Paternity assignment of 285 progeny using LOD score thresholds.

<table>
<thead>
<tr>
<th></th>
<th>LOD score threshold value</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>N progeny with identified fathers</td>
<td>44</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Mean distance between mates (m)</td>
<td>29</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Distance with greatest proportion of mating events (m)</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>N progeny with a second likely father (mean Δ)</td>
<td>40 (2.8)</td>
<td>19 (3.3)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.5 Distribution of mating events over distance as determined by paternity analysis at two LOD score threshold values, 4 (A) and 5 (B).
Chapter 4: Discussion

4.1 Arbutus seed germination

Based on observations during seed germination and DNA extraction there was no consistent connection between the clarity of AFLP profiles and the physical condition of the germinants. Generally, when seeds did not germinate readily and germinants appeared to be declining in vigour the DNA profile was less clear. Some individuals germinated may not have developed regardless of the growing conditions due to inbreeding depression. While germination success has been documented to be 90% in nature, only 0-10% survive to the next growing season (Tappeiner et al., 1986). This could be due to environmental conditions, however the biparental inbreeding detected in this study suggests that inbreeding depression, through the accumulation of deleterious recessive genotypes, may have been the cause of germinant deterioration.

During programmed cell death following the degradation of the nuclear envelope, DNA breaks into fragments (Alberts et al., 1998). The additional loci detected in the profiles of unhealthy germinants may have been a function of this process. More likely, the numerous loci detected in the germinants that were not in any population sampled for genetic diversity may indicate fungal contamination since there was fungal growth in some Petri dishes. However, there was no clear correlation between the observed fungi growth and the quality of the AFLP profiles, although damping-off fungi contribute to seedling mortality in nature (Pelton, 1962; Tappeiner et al., 1986). Great care should be taken in the extraction of arbutus seed from its fruit. Careful monitoring of the physical condition of germinants and the elimination of any possible fungus are critical when using this marker type, as DNA quality is paramount to genotyping success.

4.2 Low genetic diversity within populations and significant structure among them

Nei's (1978) unbiased average heterozygosity within populations was low (0.094 ± 0.011, mean ± SE) compared to most long-lived woody perennials with gravity and/or frugivores-dispersed seeds. Estimates for these life history traits based on RAPDs are given in Table 4.1 for comparison. Similarly low allozyme-based estimates have also been reported in Acacia species.
(reviewed by Moran et al., 1989). The level of locus polymorphism was also low in arbutus compared to other tree species analyzed using AFLPs (Van der Merwe et al., 2000; Zawko et al., 2001; Rivera-Ocasio et al., 2002; Cavers et al., 2003; Tang et al., 2003).

Table 4.1 Mean estimates of within-population genetic diversity ($H$) and among-population differentiation for three life history traits. The number of studies (N) within each group is given.

<table>
<thead>
<tr>
<th>Life history trait</th>
<th>N</th>
<th>$H^2$</th>
<th>N</th>
<th>$\phi_{ST}$</th>
<th>N</th>
<th>$G_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>long-lived perennial</td>
<td>37</td>
<td>0.25</td>
<td>60</td>
<td>0.25</td>
<td>24</td>
<td>0.19</td>
</tr>
<tr>
<td>gravity dispersed</td>
<td>24</td>
<td>0.19</td>
<td>46</td>
<td>0.45</td>
<td>26</td>
<td>0.32</td>
</tr>
<tr>
<td>frugivore dispersed</td>
<td>22</td>
<td>0.24</td>
<td>32</td>
<td>0.27</td>
<td>11</td>
<td>0.16</td>
</tr>
<tr>
<td>present study</td>
<td>0.094</td>
<td>0.16</td>
<td>0.15 ($F_{ST}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Nybom (2004) based on RAPDs
2based on polymorphic loci using Lynch and Milligan (1994)
3equivalent to $F_{ST}$ when there are two alleles at a locus, as is the case with dominant DNA markers

There were no significant differences in average heterozygosity between any pair of populations, and no significant correlation between genetic and physical distance among populations. The lack of any significant relationship between heterozygosity and either latitude or longitude may be due to sampling only the northernmost portion of the species' range. Several studies on post-Pleistocene range expansion of plants in the Pacific Northwest report lower levels of genetic variability north of the Oregon-California state border (Soltis et al., 1997). This pattern would not have been detected in this study because samples were only collected as far south as Puget Sound. Nonetheless, genetic diversity shown in this study corresponds with the observed trend in some plants of the Pacific Northwest. It is important to note, however, that some species whose ranges span portions of the Pacific Northwest simply have low genetic variability in general (e.g., western redcedar (Thuja plicata), see O’Connell, 2003), and this could be the case for arbutus. Genetic analysis over the entire range is needed to make any further conclusions regarding the post-Pleistocene range expansion of arbutus.

Analyses of among-population genetic structure estimated by $\phi_{ST}$ and $F_{ST}$ were similar and significant (0.16 and 0.15, respectively, see Tables 3.2 and 3.3). The estimate of $\phi_{ST}$ was low compared to other long-lived perennial species with seed dispersal by gravity or frugivory (Table 4.1). The estimate of $F_{ST}$ for arbutus was comparable to the mean RAPD-based estimates for long-lived perennials and species with seed dispersed by frugivory. A greater proportion of
genetic variation partitioned among populations (higher $F_{ST}$) is expected when seeds are dispersed by gravity (Table 4.1). Although arbutus seed is dispersed by both frugivores and gravity, this suggests that the former may be the dominant mode of dispersal. No certainty can be attributed to this comparison, however, since genetic divergence among populations is strongly influenced by migration. It takes very few migrants coming into a population to rapidly decrease $F_{ST}$ (Hartl and Clark, 1997); therefore it is also possible that little migration among populations occurs via seed flow in arbutus and frugivores dispersal homogenizes the patterns caused by gravity dispersal.

The discrepancy between $\phi_{ST}$ and $F_{ST}$ was also noted by Nybom (2004). She explains that when considering studies that report both parameters while simultaneously using AFLPs and RAPDs, the results are nearly identical, as is the case here. It follows then that the comparison of the results in this study with the RAPD-based mean estimates of $\phi_{ST}$ are probably misleading. Since the estimate of $F_{ST}$ for arbutus is similar to that for long-lived perennials, the amount of gene flow among populations in this species is likely comparable to other tree species on average.

Estimating $F_{ST}$ while jackknifing over populations did not reveal any significant difference upon removal of any population. Partitioning of genetic diversity among populations was most notably decreased when the population from Gold River was excluded, and somewhat decreased when Qualicum Beach was removed (Table 3.2). The segregation of these two populations was also apparent in both the UPGMA dendrogram of Reynolds' et al. (1983) coancestry distance as well as in the PCA. However, the separation of the Gold River samples apart from other populations was more distinct than that of Qualicum Beach in both the UPGMA dendrogram and the PCA (Figure 3.2 and Figure 3.3, respectively). In the former, the genetic distance between Qualicum Beach and all other populations was slightly lower than the genetic difference between Gold River and all other populations.

Gold River was the most geographically disjunct population sampled and one of the two populations with the lowest heterozygosity (Table 3.1). All other populations were within the continuous, patchy range of arbutus (Figures 1.1 and 2.1). There were no arbutus trees observed, nor have there been anecdotal reports of them, in the mountainous region between Campbell River and Gold River. The closest sampled population to the one sampled in Gold River was 92
km away on Cortes Island. A long distance seed dispersal event may have resulted in the colonization of Gold River recently enough that a resulting founder effect was detected.

The Gold River population falls within the driest maritime variant of the coastal western hemlock (CWH) biogeoclimatic zone (CWHxm2). Samples were collected on a steep embankment at the end of the Muchalat Inlet (Figure 4.1). Climatic conditions differ here compared to most parts of this CWH very wet, maritime region. The inlet runs through steep-sided fjords, which could result in a localized increase in diffuse solar radiation within the inlet, much like in a valley, and hence warmer, drier conditions on well-drained areas allow arbutus to establish. The presence of water from the Pacific Ocean would also moderate the climate. There may be several other populations of arbutus along other inlets in this region as individuals have also been noted up Tlupana Inlet (Figure 4.1).

Figure 4.1 Map illustrating the location of the Muchalat Inlet. The Gold River population is indicated with a circle (modified from Fisheries and Oceans Canada, 2002).
An explanation as to why the population in Qualicum Beach appeared differed from other sampled populations is not so straightforward. It may be due to its small sample size (QB: $n=11$, mean $n=24$; Table 3.1) and its location within the community of Qualicum Beach. Although it is not probable that this small cluster of trees was planted since nursery production of this species has begun quite recently (Gonzalez, 1999; Winters and Hummel, 1999), perhaps there was some selection for stature or health by humans. This population had somewhat higher heterozygosity than the average and is located within the continuous portion of the species' range (Table 3.1 and Figures 1.1 and 2.1, respectively).

One other group of populations also appeared to cluster in the analyses (Figures 3.2 and 3.3). This included populations from the Malahat Rise, Mount Douglas, and Galiano Island. These three populations are close geographically (Figure 2.1) and all exhibited higher heterozygosity and numbers of localized loci than on average across populations, although this trend was not significant (Table 3.1).

### 4.3 Stand level population structure supports mating system estimates

The patchy genetic structure found in this study (Figure 3.4) has been shown for both coniferous and angiosperm tree species. The scale and magnitude of genetic structure among species varies widely since species have a variety of dispersal mechanisms and reproductive strategies (see Epperson, 1992 for a review). Comparison of $F_{ST}$ estimates for arbutus (this study) and Nybom (2004) for species dispersed by frugivory and gravity suggests the former may be the more dominant dispersal mode in arbutus (Table 4.1). Again, however, it is possible that little migration occurs among populations via seed flow since frugivore dispersal homogenizes the patterns caused by gravity dispersal. Many fruits were observed beneath and around trees in the field so seed is also gravity dispersed. Species feeding on arbutus berries may excise racemes or drop a significant amount in and around a tree during foraging. Open-pollinated progeny from the same mother tree establishing close to its source may account for the genetic isolation by distance pattern and the general distribution in groves (Tappeiner et al., 1984).

Arbutus was highly outcrossing ($t_m=0.970$). This was hypothesized *a priori* due to adaptations for outcrossing in floral morphology. Although the flower is perfect, the stigma and
anther are spatially separated. A review of the literature by O'Connell (2003) indicated angiosperm trees have a $t_m$ of 0.887. Primarily outcrossing species, as shown for arbutus in this study, usually maintain high levels of genetic diversity within populations. Arbutus does not seem to conform to this trend since heterozygosity was lower than expected for outcrossing species (Table 4.2). The proportion of genetic variation partitioned among populations in this study ($F_{ST}$) is lower than might be expected for outcrossing species (Table 4.2) although it was similar to the estimate for long-lived perennials (Table 4.1). Table 4.2 presents mean estimates for within- and among-population genetic differentiation with for different mating systems. Estimates for $\phi_{ST}$ are given but should be interpreted with caution as previously mentioned.

<table>
<thead>
<tr>
<th>Mating system</th>
<th>N</th>
<th>$H^2$</th>
<th>$\phi_{ST}$</th>
<th>N</th>
<th>$G_{ST}$^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>selfing</td>
<td>10</td>
<td>0.12</td>
<td>14</td>
<td>0.65</td>
<td>6</td>
</tr>
<tr>
<td>mixed mating</td>
<td>8</td>
<td>0.18</td>
<td>18</td>
<td>0.40</td>
<td>6</td>
</tr>
<tr>
<td>outcrossing</td>
<td>38</td>
<td>0.27</td>
<td>73</td>
<td>0.27</td>
<td>31</td>
</tr>
<tr>
<td>present study</td>
<td></td>
<td>0.094</td>
<td>0.16</td>
<td></td>
<td>0.15 ($F_{ST}$)</td>
</tr>
</tbody>
</table>

1Nybom (2004) based on RAPDs
2based on polymorphic loci using Lynch and Milligan (1994)
3equivalent to $F_{ST}$ when there are two alleles at a locus, as is the case with dominant DNA markers

The estimate of average heterozygosity for arbutus resembles that of species that self-pollinate. Interestingly, the mating system analysis indicated significant biparental inbreeding ($r_s=0.793$ and $t_m-t_s=0.101$). These estimates are in close agreement, indicating approximately 10-20% of mating events are consanguineous. This finding supports the genetic structure observed at the stand level.

Since the arbutus raceme has numerous flowers, and both bees and hummingbirds tend to forage in a patchy manner, pollen carry-over from neighbouring trees could fertilize several flowers on a stalk. Pollen from several adjacent trees would probably pollinate an individual tree since, according to the correlation of outcrossed paternity among siblings, most do not share fathers ($r_p=0.094$). Regression analysis of the kinship coefficient on distance indicated that these neighbours might be closely related. It follows then, that this seed is dispersed close to the mother tree, resulting in a feedback loop that would reinforce the stand level population structure.
observed in this study. This may also explain the presence of significant among-population genetic differentiation despite the high outcrossing. Due to significant levels of biparental inbreeding and the variability in seedling vigour observed during germination, inbreeding depression may occur in arbutus. The detrimental effects of inbreeding depression in arbutus may be limited since it is primarily outcrossing and reflects similar $F_{ST}$ ($\phi_{ST}$) to other long-lived perennials and lower $F_{ST}$ ($\phi_{ST}$) than species that are outcrossing or have mixed mating (Tables 4.1 and 4.2).

Another possibility that may explain the extent of consanguineous mating is a 'leaky' self-incompatibility (SI) system. This phenomenon has been described in *Arabidopsis lyrata* and other Brassicaceae (Mable et al., 2003). It involves compatible mating between two individuals whose expressed S-locus alleles are the same, which should theoretically result in incompatibility. Arbutus exhibited little or no detectable selfing, suggesting the possible presence of an SI system; perhaps the observed biparental inbreeding is due to this 'leaky' SI phenomenon. If so, the frequency of 'leaky' matings in the population studied here may be accelerated by receiving pollen from relatives since the stand level analysis of population structure suggested that individuals at three metres are half-sibs. There is currently no data on self-incompatibility in arbutus.

Pollination biology and mating system studies in other Ericaceae report a wide variety of results. Generally, outcrossing is the predominant form of mating while there are varying levels of selfing (e.g., *Calluna vulgaris*, see Mahy and Jacquemart, 1999). This family contains 130 genera and 2,700 species (Judd et al., 2002) that vary in life form from herbaceous plants to trees. Phylogenetic relationships within the family are complicated and controversial (see Kron et al., 1997). Therefore, general conclusions about mating system and pollination biology at a family level based on the relatively few studies that exist are not justified.

### 4.4 Paternity: further support for stand level genetic structure and mating system estimates

Pollen dispersal plays an important role in the maintenance of genetic variation as it directly affects the spatial scale of gene exchange between individuals. Although I was unable to assign statistical confidence to paternity assignments, the two LOD score thresholds at which relationships were evaluated gave similar results, providing support for their accuracy. Despite this, since the LOD score difference between first and second most likely pollen parents was not
great (see Appendix A) statistical confidence is likely low. Regardless, I believe the results obtained provide important insight into the pollen distribution trend in arbutus.

Pollen distributions were leptokurtic, as is the case in most pollen flow studies (Williams, 2001) (Table 3.6, Figure 3.5). According to the regression analysis of kinship over distance (Figure 3.4), individuals at the distance where the greatest proportion of mating events took place were closely related. This concurs with the biparental inbreeding found in this population. It should be noted, however, that since pollination is density dependent (Ellstrand, 1992), the distance at which the greatest proportion of matings take place in a population can be expected to vary, and so would the degree of relatedness among mates as suggested by this study.

Paternity assignment was difficult because of the low genetic diversity in this species. The lack of heterogeneity resulted in a limited ability to differentiate among possible fathers despite a large number of marker loci. In addition, the presence of biparental inbreeding resulted in higher relatedness among individuals as compared to a species that mates with unrelated individuals. These difficulties were unforeseeable since this was the first investigation into the population genetics and pollen flow in arbutus. Future attempts to resolve paternity in this species would require the screening of several more AFLP primers to obtain sufficient polymorphism. If the development of microsatellite makers in arbutus were warranted in the future, they might provide the much needed marker variability for paternity analysis in this species.
Chapter 5: Conclusions and future directions

The populations of arbutus sampled in this study exhibited low genetic diversity and no geographic trends. Significant differences among populations were detected. $F_{ST}$ was comparable to mean estimates for long-lived perennials and species with seed dispersed by frugivores. The population collected in Gold River was most distinct genetically. For this reason, it should definitely be a candidate if conservation strategies were developed for this species. There may be other very localized regions in this geographic area with similar ecology that could also provide habitat for arbutus. Other plant species with disjunct populations in these locations may also be genetically different from their continuous populations and if identified should be included in population genetic studies. There are currently few protected areas that encompass this particular very dry variant of the CWH zone.

A range-wide study is needed to determine whether the patterns of genetic diversity observed in this study are representative of arbutus throughout its distribution. Sampling should at the very least encompass populations south of the Oregon-California state border to include populations that were not glaciated during the Pleistocene. Population genetic analyses of other species of this genus may also prove interesting: no data exists to date and phylogenetic analyses suggest that the species in the Mediterranean region and those on the west coast of North America are genetically divergent (Hileman et al., 2001).

The estimated outcrossing rate for arbutus was relatively high, which is often a feature of high within-population genetic diversity. Mating system estimates in this study were based only on one population; therefore additional populations should be assayed to confirm this finding. Despite this, stand level genetic structure, coupled with significant biparental inbreeding may explain why diversity estimates were low. According to these results, it follows that both in situ and ex situ conservation efforts should concentrate on maintaining genetic diversity by focusing on individuals that are well beyond three metres apart. Further attempts to directly study gene flow using paternity analyses should use microsatellite markers so that sufficient variability might be available. Future studies to address optimal sampling strategies for conservation should consider the impact of stand density on the mating system of arbutus as this could affect patch structure.
Investigation into whether there is a self-incompatibility system in arbutus may elucidate more fine-scale mating system dynamics. Determining the number of different fathers that fertilize ovules within a fruit would also be useful since high relatedness at this level may contribute to spatial autocorrelation of genotypes since seed from the same fruit are likely to be deposited in close proximity. Field observations of frugivore foraging on arbutus would aid in our understanding of how far seed is dispersed.

Arbutus habitat is at risk in British Columbia (MSRM, 2003). Primary threats are urban encroachment, fire suppression, grazing and exotic invasive species. Its growth is highly sensitive to environmental conditions (Ettl, 1999); therefore future climate change could adversely affect these populations already at risk due to more severe summer drought. Understanding the significance of genetic variability with respect to the reproductive, ecological and evolutionary dynamics of a species is important in developing effective conservation strategies. This thesis is an important first step in understanding these dynamics in arbutus. The population-level information identified here using molecular markers provides us with valuable information that will guide conservation of this species in British Columbia and throughout its geographic range.
References


Elliot, M. 1999a. The decline of Pacific madrone (Arbutus menziesii Pursh) in urban and natural environments: its causes and management. Thesis (M.Sc.). College of Forest Resources, University of Washington


http://www-sci.pac.dfo-mpo.gc.ca/osap/projects/bcinlets/muchalat_inlet_e.htm


Hardy, O. J. and X. Vekemans. 2002. SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618-620.


### Appendix A

Paternity analysis data at the LOD score threshold values of (i) 4 and (ii) 5.

#### i) LOD score threshold of 4

<table>
<thead>
<tr>
<th>Maternal tree</th>
<th>Progeny #</th>
<th>1st most likely father</th>
<th>1st LOD score</th>
<th>1st Distance (m)</th>
<th>2nd most likely father</th>
<th>2nd LOD score</th>
<th>2nd Distance (m)</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1_02</td>
<td>31</td>
<td>4.40</td>
<td>102.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1_09</td>
<td>1</td>
<td>4.71</td>
<td>0.0</td>
<td>30</td>
<td>3.06</td>
<td>98.7</td>
<td>1.65</td>
</tr>
<tr>
<td>2</td>
<td>2_01</td>
<td>12</td>
<td>4.60</td>
<td>60.1</td>
<td>27</td>
<td>2.37</td>
<td>73.6</td>
<td>2.23</td>
</tr>
<tr>
<td>2</td>
<td>2_06</td>
<td>27</td>
<td>4.55</td>
<td>73.6</td>
<td>12</td>
<td>3.11</td>
<td>60.1</td>
<td>1.44</td>
</tr>
<tr>
<td>3</td>
<td>3_08</td>
<td>2</td>
<td>6.97</td>
<td>3.4</td>
<td>38</td>
<td>2.21</td>
<td>119.2</td>
<td>4.76</td>
</tr>
<tr>
<td>5</td>
<td>5_18</td>
<td>17</td>
<td>4.64</td>
<td>54.2</td>
<td>12</td>
<td>1.90</td>
<td>50.7</td>
<td>2.74</td>
</tr>
<tr>
<td>5</td>
<td>5_20</td>
<td>12</td>
<td>4.21</td>
<td>50.7</td>
<td>38</td>
<td>2.75</td>
<td>106.2</td>
<td>1.46</td>
</tr>
<tr>
<td>14</td>
<td>14_02</td>
<td>17</td>
<td>8.57</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14_09</td>
<td>22</td>
<td>5.37</td>
<td>17.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14_11</td>
<td>21</td>
<td>4.94</td>
<td>16.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14_12</td>
<td>29</td>
<td>5.92</td>
<td>23.3</td>
<td>27</td>
<td>3.93</td>
<td>22.4</td>
<td>1.99</td>
</tr>
<tr>
<td>14</td>
<td>14_15</td>
<td>27</td>
<td>4.70</td>
<td>22.4</td>
<td>18</td>
<td>0.29</td>
<td>11.0</td>
<td>4.41</td>
</tr>
<tr>
<td>15</td>
<td>15_04</td>
<td>36</td>
<td>7.13</td>
<td>66.4</td>
<td>10</td>
<td>5.01</td>
<td>3.3</td>
<td>2.12</td>
</tr>
<tr>
<td>15</td>
<td>15_05</td>
<td>14</td>
<td>6.39</td>
<td>2.0</td>
<td>2</td>
<td>2.16</td>
<td>59.4</td>
<td>4.23</td>
</tr>
<tr>
<td>15</td>
<td>15_06</td>
<td>5</td>
<td>4.22</td>
<td>50.3</td>
<td>36</td>
<td>3.00</td>
<td>66.4</td>
<td>1.22</td>
</tr>
<tr>
<td>15</td>
<td>15_09</td>
<td>20</td>
<td>4.12</td>
<td>18.1</td>
<td>24</td>
<td>1.60</td>
<td>24.7</td>
<td>2.52</td>
</tr>
<tr>
<td>15</td>
<td>15_15</td>
<td>22</td>
<td>4.58</td>
<td>19.0</td>
<td>19</td>
<td>2.65</td>
<td>15.3</td>
<td>1.93</td>
</tr>
<tr>
<td>16</td>
<td>16_03</td>
<td>17</td>
<td>4.59</td>
<td>2.1</td>
<td>29</td>
<td>0.70</td>
<td>18.9</td>
<td>3.89</td>
</tr>
<tr>
<td>16</td>
<td>16_07</td>
<td>17</td>
<td>5.30</td>
<td>2.1</td>
<td>3</td>
<td>0.88</td>
<td>61.0</td>
<td>4.42</td>
</tr>
<tr>
<td>17</td>
<td>17_07</td>
<td>12</td>
<td>5.32</td>
<td>6.5</td>
<td>22</td>
<td>2.27</td>
<td>10.7</td>
<td>3.05</td>
</tr>
<tr>
<td>17</td>
<td>17_08</td>
<td>12</td>
<td>4.18</td>
<td>6.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>18_06</td>
<td>30</td>
<td>4.54</td>
<td>31.5</td>
<td>20</td>
<td>1.34</td>
<td>7.0</td>
<td>3.20</td>
</tr>
<tr>
<td>19</td>
<td>19_07</td>
<td>15</td>
<td>4.39</td>
<td>15.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>19_14</td>
<td>23</td>
<td>6.99</td>
<td>6.4</td>
<td>37</td>
<td>1.76</td>
<td>52.7</td>
<td>5.23</td>
</tr>
<tr>
<td>23</td>
<td>23_05</td>
<td>17</td>
<td>4.21</td>
<td>12.8</td>
<td>3</td>
<td>2.36</td>
<td>75.1</td>
<td>1.85</td>
</tr>
<tr>
<td>23</td>
<td>23_12</td>
<td>16</td>
<td>4.13</td>
<td>14.7</td>
<td>19</td>
<td>1.44</td>
<td>6.4</td>
<td>2.69</td>
</tr>
<tr>
<td>27</td>
<td>27_10</td>
<td>27</td>
<td>6.72</td>
<td>0.0</td>
<td>30</td>
<td>1.81</td>
<td>22.6</td>
<td>4.91</td>
</tr>
</tbody>
</table>

64
### Table of Data

<table>
<thead>
<tr>
<th>Maternal tree</th>
<th>Progeny #</th>
<th>1st most likely father</th>
<th>1LOD score</th>
<th>Distance (m)</th>
<th>2nd most likely father</th>
<th>2LOD score</th>
<th>Distance (m)</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>28.04</td>
<td>25</td>
<td>6.00</td>
<td>10.1</td>
<td>11</td>
<td>0.54</td>
<td>21.0</td>
<td>5.46</td>
</tr>
<tr>
<td>28</td>
<td>28.06</td>
<td>28</td>
<td>4.39</td>
<td>0.0</td>
<td>2</td>
<td>2.12</td>
<td>73.8</td>
<td>2.27</td>
</tr>
<tr>
<td>30</td>
<td>30.02</td>
<td>29</td>
<td>10.60</td>
<td>17.8</td>
<td>25</td>
<td>6.36</td>
<td>17.9</td>
<td>4.24</td>
</tr>
<tr>
<td>30</td>
<td>30.07</td>
<td>17</td>
<td>4.25</td>
<td>34.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>30.08</td>
<td>29</td>
<td>6.12</td>
<td>17.8</td>
<td>25</td>
<td>1.89</td>
<td>17.9</td>
<td>4.23</td>
</tr>
<tr>
<td>30</td>
<td>30.13</td>
<td>29</td>
<td>8.71</td>
<td>17.8</td>
<td>25</td>
<td>5.47</td>
<td>17.9</td>
<td>3.24</td>
</tr>
<tr>
<td>30</td>
<td>30.14</td>
<td>25</td>
<td>5.62</td>
<td>17.9</td>
<td>29</td>
<td>0.09</td>
<td>17.8</td>
<td>5.53</td>
</tr>
<tr>
<td>31</td>
<td>31.08</td>
<td>29</td>
<td>5.85</td>
<td>21.1</td>
<td>25</td>
<td>3.97</td>
<td>21.5</td>
<td>1.88</td>
</tr>
<tr>
<td>31</td>
<td>31.11</td>
<td>17</td>
<td>5.29</td>
<td>37.9</td>
<td>26</td>
<td>3.47</td>
<td>25.7</td>
<td>1.82</td>
</tr>
<tr>
<td>31</td>
<td>31.12</td>
<td>12</td>
<td>4.67</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>31.13</td>
<td>3</td>
<td>4.60</td>
<td>98.8</td>
<td>14</td>
<td>1.00</td>
<td>44.4</td>
<td>3.60</td>
</tr>
<tr>
<td>31</td>
<td>31.19</td>
<td>26</td>
<td>4.41</td>
<td>25.7</td>
<td>17</td>
<td>3.33</td>
<td>37.9</td>
<td>1.08</td>
</tr>
<tr>
<td>34</td>
<td>34.09</td>
<td>17</td>
<td>4.97</td>
<td>55.9</td>
<td>11</td>
<td>2.83</td>
<td>60.6</td>
<td>2.14</td>
</tr>
<tr>
<td>36</td>
<td>36.11</td>
<td>3</td>
<td>4.72</td>
<td>110.5</td>
<td>12</td>
<td>2.02</td>
<td>64.1</td>
<td>2.70</td>
</tr>
<tr>
<td>36</td>
<td>36.13</td>
<td>3</td>
<td>5.30</td>
<td>110.5</td>
<td>12</td>
<td>1.55</td>
<td>64.1</td>
<td>3.75</td>
</tr>
<tr>
<td>37</td>
<td>37.01</td>
<td>37</td>
<td>4.32</td>
<td>0.0</td>
<td>34</td>
<td>2.74</td>
<td>9.9</td>
<td>1.58</td>
</tr>
<tr>
<td>37</td>
<td>37.02</td>
<td>37</td>
<td>5.14</td>
<td>0.0</td>
<td>34</td>
<td>4.72</td>
<td>9.9</td>
<td>0.42</td>
</tr>
<tr>
<td>37</td>
<td>37.03</td>
<td>37</td>
<td>6.48</td>
<td>0.0</td>
<td>34</td>
<td>4.89</td>
<td>9.9</td>
<td>1.59</td>
</tr>
<tr>
<td>37</td>
<td>37.05</td>
<td>37</td>
<td>5.08</td>
<td>0.0</td>
<td>34</td>
<td>4.45</td>
<td>9.9</td>
<td>0.63</td>
</tr>
<tr>
<td>37</td>
<td>37.06</td>
<td>37</td>
<td>4.89</td>
<td>0.0</td>
<td>34</td>
<td>3.22</td>
<td>9.9</td>
<td>1.67</td>
</tr>
<tr>
<td>37</td>
<td>37.09</td>
<td>37</td>
<td>4.12</td>
<td>0.0</td>
<td>34</td>
<td>2.34</td>
<td>9.9</td>
<td>1.78</td>
</tr>
<tr>
<td>38</td>
<td>38.10</td>
<td>5</td>
<td>5.33</td>
<td>106.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Values correspond to most likely father.
2 Values correspond to second most likely father.
## ii) LOD score threshold of 5

<table>
<thead>
<tr>
<th>Maternal tree</th>
<th>Progeny #</th>
<th>Most likely father</th>
<th>1LOD score</th>
<th>1Distance (m)</th>
<th>2nd most likely father</th>
<th>2LOD score</th>
<th>2Distance (m)</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>03_08</td>
<td>2</td>
<td>6.97</td>
<td>3.4</td>
<td>38</td>
<td>2.21</td>
<td>119.2</td>
<td>4.76</td>
</tr>
<tr>
<td>14</td>
<td>14_02</td>
<td>17</td>
<td>8.57</td>
<td>8.1</td>
<td>27</td>
<td>5.01</td>
<td>3</td>
<td>2.12</td>
</tr>
<tr>
<td>14</td>
<td>14_09</td>
<td>22</td>
<td>5.37</td>
<td>17.8</td>
<td>10</td>
<td>2.16</td>
<td>59.4</td>
<td>4.23</td>
</tr>
<tr>
<td>14</td>
<td>14_12</td>
<td>29</td>
<td>5.92</td>
<td>23.3</td>
<td>27</td>
<td>3.93</td>
<td>22.4</td>
<td>1.99</td>
</tr>
<tr>
<td>15</td>
<td>15_04</td>
<td>36</td>
<td>7.13</td>
<td>66.4</td>
<td>10</td>
<td>5.01</td>
<td>3</td>
<td>2.12</td>
</tr>
<tr>
<td>15</td>
<td>15_05</td>
<td>14</td>
<td>6.39</td>
<td>2</td>
<td>3</td>
<td>0.88</td>
<td>61.0</td>
<td>4.42</td>
</tr>
<tr>
<td>16</td>
<td>16_07</td>
<td>17</td>
<td>5.30</td>
<td>2.1</td>
<td>2</td>
<td>2.27</td>
<td>10.7</td>
<td>3.05</td>
</tr>
<tr>
<td>17</td>
<td>17_07</td>
<td>12</td>
<td>5.32</td>
<td>6.5</td>
<td>22</td>
<td>1.76</td>
<td>52.7</td>
<td>5.23</td>
</tr>
<tr>
<td>19</td>
<td>19_14</td>
<td>23</td>
<td>6.99</td>
<td>6.4</td>
<td>37</td>
<td>1.81</td>
<td>22.6</td>
<td>4.91</td>
</tr>
<tr>
<td>27</td>
<td>27_10</td>
<td>27</td>
<td>6.72</td>
<td>0</td>
<td>30</td>
<td>1.81</td>
<td>22.6</td>
<td>4.91</td>
</tr>
<tr>
<td>28</td>
<td>28_04</td>
<td>25</td>
<td>6.00</td>
<td>10.1</td>
<td>11</td>
<td>0.54</td>
<td>21.0</td>
<td>5.46</td>
</tr>
<tr>
<td>30</td>
<td>30_02</td>
<td>29</td>
<td>10.60</td>
<td>17.8</td>
<td>25</td>
<td>6.36</td>
<td>17.9</td>
<td>4.24</td>
</tr>
<tr>
<td>30</td>
<td>30_08</td>
<td>29</td>
<td>6.12</td>
<td>17.8</td>
<td>25</td>
<td>1.89</td>
<td>17.9</td>
<td>4.23</td>
</tr>
<tr>
<td>30</td>
<td>30_13</td>
<td>29</td>
<td>8.71</td>
<td>17.8</td>
<td>25</td>
<td>5.47</td>
<td>17.9</td>
<td>3.24</td>
</tr>
<tr>
<td>30</td>
<td>30_14</td>
<td>25</td>
<td>5.62</td>
<td>17.9</td>
<td>29</td>
<td>0.09</td>
<td>17.8</td>
<td>5.53</td>
</tr>
<tr>
<td>31</td>
<td>31_08</td>
<td>29</td>
<td>5.85</td>
<td>21.1</td>
<td>25</td>
<td>3.97</td>
<td>21.5</td>
<td>1.88</td>
</tr>
<tr>
<td>31</td>
<td>31_11</td>
<td>17</td>
<td>5.29</td>
<td>37.9</td>
<td>26</td>
<td>3.47</td>
<td>25.7</td>
<td>1.82</td>
</tr>
<tr>
<td>36</td>
<td>36_13</td>
<td>3</td>
<td>5.30</td>
<td>110.5</td>
<td>12</td>
<td>1.55</td>
<td>64.1</td>
<td>3.75</td>
</tr>
<tr>
<td>37</td>
<td>37_02</td>
<td>37</td>
<td>5.14</td>
<td>0</td>
<td>34</td>
<td>4.72</td>
<td>9.9</td>
<td>0.42</td>
</tr>
<tr>
<td>37</td>
<td>37_03</td>
<td>37</td>
<td>6.48</td>
<td>0</td>
<td>34</td>
<td>4.89</td>
<td>9.9</td>
<td>1.59</td>
</tr>
<tr>
<td>37</td>
<td>37_05</td>
<td>37</td>
<td>5.08</td>
<td>0</td>
<td>34</td>
<td>4.45</td>
<td>9.9</td>
<td>0.63</td>
</tr>
<tr>
<td>38</td>
<td>38_10</td>
<td>5</td>
<td>5.33</td>
<td>106.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Values correspond to most likely father.

2Values correspond to second most likely father.