Morphological Differentiation of *Alnus* Pollen from Western North America

by

Laura May
B.Sc., Thompson Rivers University, 2009
B.A., Thompson Rivers University, 2009

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTERS OF SCIENCE

in the Department of Biology

Supervisory Committee

Morphological Differentiation of *Alnus* Pollen from Western North America

by

Laura May
B.Sc., Thompson Rivers University, 2009
BA, Thompson Rivers University, 2009

Supervisory Committee

Dr. Terri Lacourse (Department of Biology)
Supervisor

Dr. Patrick von Aderkas (Department of Biology)
Departmental Member

Dr. Dan Smith (Department of Geography)
Outside Member

Abstract

Supervisory Committee

Dr. Terri Lacourse (Department of Biology)
Supervisor

Dr. Patrick von Aderkas (Department of Biology)
Departmental Member

Dr. Dan Smith (Department of Geography)
Outside Member

Increasing the taxonomic resolution of fossil pollen identification is important for accurate paleoecological reconstructions. Here, an attempt is made to identify the critical morphological features that will permit differentiation of *Alnus* pollen in fossil records. Palynologists working in the Pacific Northwest often distinguish alder pollen into two morphotypes. However, no definitive method outlining the validity of species level identifications has been devised to date. To test and validate species-level identifications, the pollen morphology of the three main alder species (*Alnus viridis* subsp. *sinuata, Alnus incana* subsp. *tenuifolia* and *Alnus rubra*) that occur in westernNorth America is examined with the goal of identifying morphological characteristics with which to distinguish the pollen of these species in fossil records. Modern pollen samples were collected from 27-35 individual plants from across the range of each of the three alder species. Pollen grains (n=30) from each individual plant were examined using light microscopy at 1000× magnification under oil immersion. For each individual pollen grain, six quantitative traits (pollen grain diameter, exine thickness, arci width, and annulus height, width and area), and three qualitative traits (pore protrusion, grain shape

and arci strength) were measured. In total, 21,390 alder pollen were examined from 93 separate collections. In addition, the number of pores was determined for 200 pollen grains from each individual plant. Statistically significant differences between species were found for all quantitative traits when traits were compared via nested ANOVA. However, there is high variability in pollen morphology within each species and pollen morphology is best described as occurring along a morphological continuum. A single morphological trait is insufficient for precise identification of alder pollen to species. CART analysis, when used to derive a multi-trait classification model, is shown to be a useful tool in separating the pollen of *A. rubra* and *A. viridis* subsp. *sinuata* into two separate 'morphotypes,' analogous to species identification. The confounding intermediate morphology of *A. incana* subsp. *tenuifolia* precludes the possibility of distinguishing the pollen of all three species. CART modelling isolates *A. rubra* and *A. viridis* subsp. *sinuata* pollen based on annulus width, arci strength, diameter and exine thickness, traits that support the differences used by palynologists for separating alder pollen into 'morphotypes.' Sensitivity analysis shows clearly that the common practice of using small sample sizes (e.g. n=7 and n=15) for identifying critical morphological traits for pollen identification produces misleading and erroneous results. Regional differences in pollen morphology were also assessed by splitting the dataset into regions. Classification accuracy is diminished from over 70% to less than 20% when a CART model derived from pollen grains from one region is used to classify grains from a different region. This research underscores the importance of using large sample sizes from across species' ranges when attempting to determine the diagnostic morphological features for accurate pollen identification.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank the members of my committee, Dr. Dan Smith and Dr. Patrick von Aderkas for the time they took to help me develop my research ideas, and for providing great feedback regarding my project, and Dr. Laura Cowen for agreeing to participate in my thesis defense as the external examiner. I would also like to thank the University of Victoria, University of British Columbia (Beaty Biodiversity Museum) and Royal British Columbia Museum herbaria, as well as Dr. Rolf Mathewes at Simon Fraser University for providing my sample material. Most importantly, I would like to thank my supervisor Dr. Terri Lacourse who is an amazing scientist and mentor, and who provided insight and motivation throughout the MSc process. Lastly, I would like to thank the Canadian Association of Palynologists (CAP) for helping to support my research by awarding me the 2011 CAP Student Research Award, as well as the National Science and Engineering Research Council (NSERC) for awarding me an Alexander Graham Bell (CGSM) Scholarship in 2009. This research was also supported by NSERC and Canadian Foundation of Innovation research grants through my supervisor Dr. Terri Lacourse.

Dedication

For Aaron,
my mom & dad,
and Odin

Introduction

The fossil pollen record is multivariate and multi-scale in nature, and can therefore be used to reconstruct multiple aspects of the environment simultaneously (Huntley et al. 1993; Huntley 2001). However, as a result of differential pollen production, dispersal and preservation, fossil pollen records are unavoidably biased toward certain plant taxa i.e., wind-pollinated trees and shrubs (Prentice 1988). Pollen records also suffer from low taxonomic resolution due to the difficulty in identifying many pollen types beyond the family or genus level (Birks 1993; Seppä and Bennett 2003). Given the large ecological differences between species within genera and between genera within an individual plant family, low taxonomic resolution poses a problem in paleoecological studies. Moreover, the prevalence of important autoecological differences between species within one genus means that grouping pollen types by genus in reconstructions of past vegetation masks species composition changes through time, as well as differential responses of congeneric species to changes in climatic and environmental conditions (Finkelstein et al. 2006). Low taxonomic resolution is particularly a problem for studies focussing on species specific ecologies and/or interspecific interactions through time (Flenley 2003). Lack of taxonomic resolution in paleoecological studies also inhibits correlations between paleo-vegetation reconstructions and modern plant survey data (Finkelstein et al. 2006). Work aimed at increasing taxonomic resolution is needed, as it is these questions that are of increasing importance to the field of paleoecology (Walker 1990; Birks 1993; Seppä and Bennett 2003; Finkelstein et al. 2006; Payne et al. 2011). While recent studies have concentrated

on enhancing taxonomic resolution for *Picea*, *Betula* and *Pinus* pollen (Lindbladh et al. 2002; Clegg et al. 2005; Barton et al. 2011), the pollen of *Alnus*, a genus consisting of tree and shrub species of particular ecological importance in the Pacific Northwest, has not been formally differentiated beyond the genus level.

Ecology of Pacific Northwest Alders

Alder species, trees and shrubs in the family Betulaceae, are important components of ecosystems and plant communities. Species in this genus are able to fix atmospheric nitrogen via a symbiotic association with the actinomycete *Frankia*, found within nodules on the roots of the plants with which they are associated (Flora of North America 1993). Due to their ability to fix nitrogen, alder are important early seral species on landscapes undergoing plant community succession (Chapin et al. 1994; Bormann and Sidle 1990; Titus 2009). Moreover, alder are important indicator species for forest fire and ecosystem disturbance regimes (Lantz et al. 2010). As principal components in modern plant communities, it is likely that alder played a similarly important role in plant succession and ecosystem dynamics throughout the late Quaternary period (e.g. Hu et al. 2001; Lacourse 2005).

Given the importance of alder in plant communities and ecosystems, it would be advantageous to be able to distinguish between alder species in the fossil pollen record. This is especially true for the alder species that occur in the Pacific Northwest: *Alnus rubra* Bong. (red alder), *Alnus incana* subsp. *tenuifolia* Nutt. (mountain alder), and *Alnus viridis* subsp. *sinuata* Regel (green alder). These three species are ecologically disparate, such that grouping them into a single taxonomic unit results in a substantial loss of ecological information.

Alnus rubra, a tree species, grows to 28 m tall and is the largest alder in North America (Flora of North America 1993). It is the most widely distributed broadleaf tree species on the Pacific coast, often forming extensive stands along the coastline, stream banks and in low-lying flood plains (Flora of North America 1993; Xie 2008). *Alnus rubra* is restricted to the Pacific coastal fog belt and does not grow more than 200 km from the coast (Fig. 1.1 [C], Thompson et al. 1999). This species reaches maximum elevations of 300 m (Furlow 1979; Flora of North America 1993; Douglas et al. 1998). *Alnus rubra* is particularly adapted to the coastal zone as it can withstand flooding as well as brackish water (USDA Plants Database 2011), but is restricted from colonizing inland areas as it requires a minimum of 180 frost free days per year and extensive precipitation during the growing season in order to survive (Thompson et al. 1999; USDA Plants Database 2011). Moreover, *A. rubra* has a low tolerance to shading (Niinemets and Valladares 2006).

Flowering in *A. rubra* occurs early in the year (late-winter or early spring), in comparison to *A. incana* subsp. *tenuifolia* and *A. viridis* subsp. *sinuata*, but all three species produce prolific seed crops. *Alnus rubra* is also capable of vegetative reproduction via re-sprouting (USDA Plants Database 2011). *Alnus rubra* can be differentiated from *A. incana* subsp. *tenuifolia* and *A. viridis* subsp. *sinuata* based on its growth form, size and strongly revolute leaf margins (Flora of North America 1993).

Alnus incana subsp. *tenuifolia* and *A. viridis* subsp. *sinuata* (mountain and green alder, respectively) both have shrubby growth habits. However, *A. viridis* subsp. *sinuata* is distinct among the alders due to its sessile buds with several imbricate scales. *Alnus*

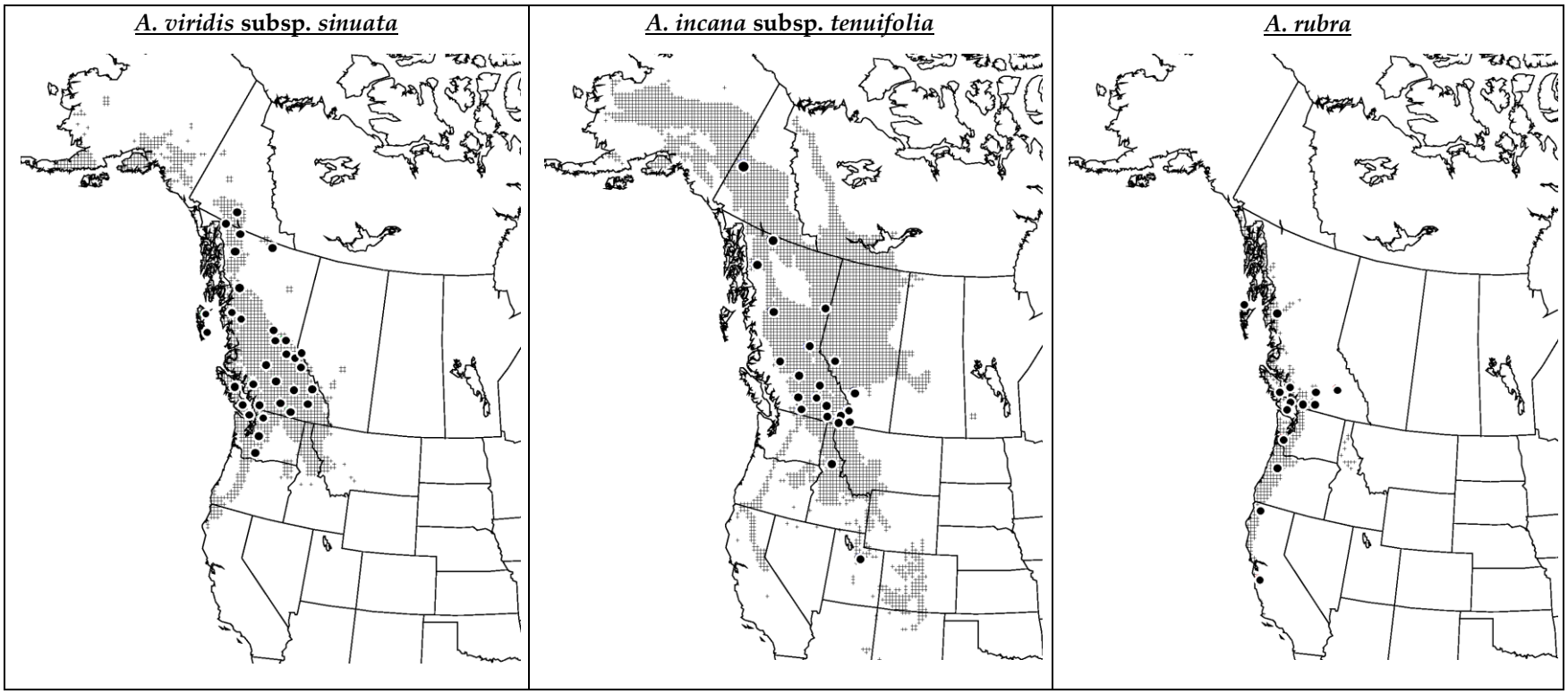


Figure 1.1: Distribution maps of *Alnus viridis* subsp. *sinuata* (green alder), *Alnus incana* subsp. *tenuifolia* (mountain alder) and *Alnus rubra* (red alder). Black circles represent sample locations. Note that some samples cannot be seen due to overlap in sample location. Distribution map source: U.S. Geologic Survey (USGS) Professional Paper #1650 (Thompson et al. 1999) -http://pubs.usgs.gov/pp/p1650-a/pages/hardwoods.html.

viridis subsp. *sinuata* grows to a maximum height of 10 m. It reaches a maximum elevation of 2500 m and is most prevalent along stream banks, lakeshores, coastlines and on rocky slopes, preferring areas with a high water table (Flora of North America 1993). This species ranges from central Alaska and the Yukon to north-western California and into western Alberta and Idaho (Fig. 1.1 [A], Thompson et al. 1999). Like *A. rubra*, *A. viridis* subsp. *sinuata* can re-sprout vegetatively, but is more shade tolerant than *A. rubra* and can persist under conifer stands (Niinemets and Valladares 2006; USDA Plants Database 2011).

Alnus incana subsp. *tenuifolia* grows at the highest elevations (to 3000 m) of the three alder species that occur in the Pacific Northwest (Flora of North America 1993). It grows to 12 m tall and is the most widely distributed alder species in western North America (Fig. 1.1 [B], Thompson et al. 1999). Like *A. rubra* and *A. viridis* subsp. *sinuata*, *A. incana* subsp. *tenuifolia* is commonly found in moist habitats such as along streams and riverbanks. It can also re-sprout vegetatively, but unlike *A. rubra* and *A. viridis* subsp. *sinuata*, it forms large thickets via rhizomes (USDA Plants Database 2011). *Alnus incana* subsp. *tenuifolia* is also shade tolerant and can persist in the forest understory (Douglas et al. 1998; USDA Plants Database 2011). *Alnus rubra*, *A. incana* subsp. *tenuifolia*, and *A. viridis* subsp. *sinuata* may be congeners, but they differ in their environmental tolerances to waterlogging, soil pH, soil texture, soil oxygen levels and fire resistance, and have variable drought tolerance, moisture use and palatability for browse animals (Niinemets and Valladares 2006; USDA Plants Database 2011).

Interestingly, *Alnus* species were not differentiated from birch in early floristic keys. Linnaeus classified alder as a single species within the genus *Betula* (Furlow

1979). Alder species are now differentiated into a single genus based primarily on their woody infructescences (Flora of North America 1993). The genus *Alnus* is further divided into subgenera based on exposure of pistillate catkins in winter, leaf venation and blooming season. *Alnus incana* subsp. *tenuifolia* and *A. rubra* have historically been grouped together along with several other alder species into the subgenus *Alnus*, whereas *A. viridis* subsp. *sinuata* falls into the subgenus *Alnobetula* (Furlow 1979). Originally subgenus classifications were made as a result of morphological and life history traits; however, these divisions have recently been supported by phylogenetic evidence (Navarro et al. 2003; Chen and Li 2004).

Alder Pollen

Despite different ecologies, life history traits and environmental requirements, the pollen morphology of these three western North American alder species is very similar. Each individual alder plant produces an abundance of wind-borne pollen grains from their male catkins. While the production of spores and pollen as a means of propagating new individuals is universal among plants (Brasier 1980), as angiosperms, alders produce pollen via a process exclusive to flowering plants. This angiosperm lineage dates back to the lower Cretaceous (Brasier 1980).

Angiosperm pollen is formed during the process of microsporogenesis, which results in single celled pollen grains within the microsporangia (Raven et al. 2005). During microsporogenesis, four groups of fertile (sporogenous) cells develop in the anther. Each sporogenous cell is surrounded by sterile cells which become the wall of the pollen sac, as well as by nutritive cells called the tapetum, which nourish the developing sporogenous cells. Each of the diploid sporogenous cells divide meiotically into four

haploid microspores. Microspores are simultaneously walled off after their second meiotic division. It is at this point when angiosperm pollen grains develop their outer wall or exine. The exine is made up of sporopollenin, made up of resistant phenylpropanoid polymers and lipidic monomers covalently bonded by ether and ester bonds (Grienenberger et al. 2010). Pollen grains also develop an intine (internal wall) of cellulose and pectin, as well as a pollen coat (Raven et al. 2005). The overall shape of pollen grains is a result of the type of meiotic division the grains undergo. It is the grain shape as well as the features of the outer wall itself (i.e., number and type of apertures, exine structure and exine ornamentation) that allow pollen grains to be identified to a particular plant family, genus and/or species (Fægri and Iversen 1989).

Alder pollen grains are stephanoporate i.e., having three or more pores arranged equatorially. They are sub-circular or pentagonal when viewed on their isopolar axis. The overall shape of the pollen grains is oblate i.e., two flattened sides opposite each other. Pores tend to protrude and each pore is surrounded by exine thickening called an annulus. Thickened, curved bands, called arci, often connect the annuli of each grain. Alder pollen grains appear psilate i.e., no exine ornamentation is visible on the pollen grain surface, when using light microscopy (Richard 1970; Fægri and Iversen 1989); however, alder pollen is scabrate when observed using a scanning electron microscope (Blackmore et al. 2003).

Study Background and Need

The lack of a reliable method for differentiating alder pollen to species in the Pacific Northwest prevents examination of species-specific questions involving this genus and hinders the reconstruction of species level post-glacial histories. Work has

been done on the east coast of North America and in Europe with regards to describing the morphology of alder pollen types that occur in these areas (Richard 1970; Furlow 1979; Mayle et al. 1993; Wittborn et al. 1996; Blackmore et al. 2003); however, the vast majority of palynological studies in the Pacific Northwest simply group all alder species into their genus '*Alnus*' (e.g. Mathewes 1973; Banner et al. 1983; Hebda 1995; Minckley et al. 2008). In recent years, some palynological studies from western North America (e.g. Gavin et al. 2001; Lacourse 2005, 2009) have separated alder pollen into two morphotypes, an '*Alnus rubra*-type' and an '*Alnus viridis*-type,' based on morphological descriptions in eastern North America and European studies (Richard 1970, Mayle et al. 1993) in combination with comparisons between fossil pollen and modern pollen reference collections. The informal and non-quantitative traits used to differentiate thealder morphotypes are thick arci, convex grain shape, larger diameter and visibly protruding annulus for the '*Alnus rubra*-type' and thin arci, concave grain shape, smaller diameter and less protruding pores for the '*Alnus viridis*-type' (T. Lacourse, pers. comm.). Alder pollen has also been suggested as a possible tool for rock strata correlation based on observed temporal shifts in alder pore number from predominantly 4-pored to predominantly 5-pored grains in Alaska (Reinink-Smith 2010). However, as Lindbladh et al. (2002) point out in their study on *Picea* pollen morphology, where the morphological differences between pollen of closely related taxa are slight, judgment-based identification, even when based in extensive experience, may nonetheless result in errors of classification and non-comparable / non-reproducible results. To date, no definitive method for species level identification and/or separation of alder pollen into 'morphotypes' has been devised for the alder species that occur along the west coast of

North America. This is problematic as *Alnus* pollen can account for a large proportion (up to 80%) of fossil pollen assemblages in western North America (e.g., MacDonald and Richie 1986; Hansen and Engstrom 1990; Brown and Hebda 2003; Lacourse 2005; Lacourse et al. 2005).

Research Approach and Objectives

Here, modern alder pollen are assessed with the objective of determining if the pollen of these three western North American alders (*A. viridis* subsp. *sinuata*, *A. incana* subsp. *tenuifolia*, and *A. rubra*) can be reliably and consistently differentiated to species based on morphology. Six quantitative morphological traits and three qualitative traits are examined on a total of 21,390 pollen grains from 93 separate collections. Statistical analyses, including nested ANOVA and classification and regression tree (CART) modelling, are performed in an effort to produce a well-defined technique for identifying alder pollen to species in this region. As is the case in all studies using modern reference pollen to identify morphological traits important in the identification of fossil pollen, temporal stability of pollen morphology is assumed.

For the purpose of this research, pollen was collected with the goal of gathering a minimum of 30 individual plants from across the range of each alder species. Sample size can greatly influence the results of all statistical analyses and insufficient sampling of any population can be confounding to statistical relevance and interpretation (Glover and Mitchell 2002; Whitlock and Schluter 2009). Most studies aimed at identifying diagnostic morphological traits for pollen identification have been based on much smaller sample sizes than n=30 and often samples have been collected from limited regions (e.g., Lindbladh et al. 2002; Clegg et al. 2005; Barton et al. 2011). To test the effectiveness of

the sampling strategy used here to avoid these study design pitfalls, sensitivity analysis is performed on all statistical tests and models. By reducing the alder pollen dataset in size and by splitting the dataset into regional sub-sets and re-analyzing these reduced datasets, the impact of the number of samples and sample location is assessed. Determination of whether sample size and/or location have the potential to change conclusions about species level differentiation of alder pollen may have important implications for other palynological studies, as well as implications for forwarding the discipline-wide goal of improving taxonomic resolution in paleoecological reconstructions (Seppä and Bennett 2003; Finkelstein et al. 2006).

Methods and Materials

Pollen Sample Collection and Preparation

Pollen samples from all three alder species were collected from the University of Victoria Herbarium, University of British Columbia Herbarium and the Royal British Columbia Museum Herbarium. A total of 93 individual alder plants were sampled from these herbaria (Appendix A). Effort was made to sample from across each species distribution (Fig. 1.1) and latitude for each sample was estimated from sample locations noted on each herbarium sheet. Total sample number per species was limited by the number of herbarium sheets that included male catkins, as well as sample identification uncertainty. Sample identification was slightly problematic due to the recent taxonomic changes in this genus; all botanical nomenclature follows the current species names listed in the Flora of North America (Flora of North America 1993). Only herbarium sheets with clear species identifications were sub-sampled for pollen. In total, 35 pollen samples were collected for *A. viridis* subsp. *sinuata*, 27 for *A. incana* subsp. *tenuifolia* and 31 for *A. rubra*.

Male catkins were prepared for light microscopy using the standard pollen acetolysis technique, which removes the external pollen kit and internal cellular components from pollen grains (Fægri and Iversen 1989; Bennett and Willis 2001). Samples were treated first with 10% potassium hydroxide for eight minutes followed by a three minute treatment of acetolysis solution, which is a 9:1 mixture of acetic anhydride and concentrated sulfuric acid. Following acetolysis, samples were washed with glacial acetic acid to prevent further action of the acetolysis mixture. Samples were then treated with two rounds of 95% ethanol. Silicone oil (2000 cs) was added to the samples once

dehydration was complete and all remaining ethanol had evaporated. Pollen grains were not stained. Silicone oil was used as a mounting medium and slides were sealed using clear nail polish. Silicone oil was chosen as the mounting medium because other commonly used mediums such as glycerine cause changes in pollen size and shape (Andersen 1960). Silicone oil is effective for pollen size comparisons and morphological assessments because it remains fluid and pollen size remains constant when immersed in the medium (Andersen 1960; Whitehead 1961; Fægri and Iversen 1989; Mäkelä 1996).

Morphological Measurements

The morphological traits assessed for each pollen grain in this study were chosen based on the informal criteria used currently by palynologists when separating fossil alder pollen into two morphotypes, an '*A. rubra* type' and an '*A. viridis* type,' and on published pollen identification keys and morphological descriptions of alder pollen in eastern North America and Europe (Richard 1970; Furlow 1979; Mayle et al. 1993; Kapp et al. 2000; Blackmore et al. 2003). Five quantitative morphological traits were measured on each pollen grain: arci width, annulus height and width, exine thickness and grain diameter (Fig. 2.1). Exine thickness is a measure of combined endexine and ektexine thickness. Annulus area was derived for each pollen grain based on annulus height and width. For traits where multiple measurements were possible on one grain (e.g., there are up to six arci on any given pollen grain), multiple measurements were taken and then averaged across an individual grain.

Three qualitative morphological traits (arci strength, grain shape and annulus protrusion) were also assessed on each pollen grain. Arci strength was assessed for each grain and assigned a relative rank from 0 (arci not visible) to 5 (very prominent arci).

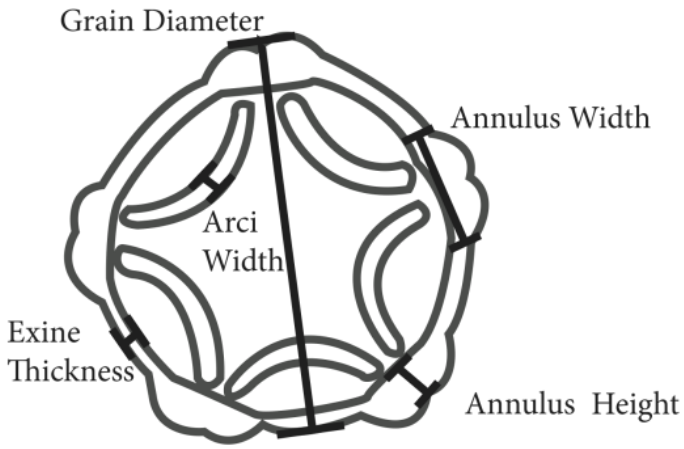


Figure 2.1: Isopolar view of a simplified 5-pored convex alder pollen grain showing the five measured quantitative traits.

The overall protrusion of the annulus from the exine was scored on a scale from 1 (annulus flush with the exine) to 3 (annulus protruding substantially from the exine). Overall grain shape was assessed as concave, convex or mixed (Fig. 2.2). As grain shape is a function of exine concavity between any two pores on each individual grain, threshold parameters were outlined in defining grain shape. For example, a 5-pored grain needed at least four exine segments of similar concavity to be classified as either concave or convex. The 'mixed' category was assigned to grains that did not meet the threshold set for each specific pollen grain pore number encountered (e.g., 3-pored grain threshold: all exine segments of similar concavity or grain classified as mixed; 4-pored grain threshold: ≥ 3 segments of similar concavity or grain classified as mixed). Pore number (i.e., the number of pores on an individual pollen grain) was also counted.

The six quantitative and three qualitative traits were measured or assessed on 30 pollen grains from each alder sample. Pore counts were made on an additional 200 grains

Figure 2.2: Pollen shape classes.

per sample. In total, 21,390 pollen grains were examined. The number of pores on each pollen grain was determined at 400× magnification using a Zeiss Axio Imager M1 compound microscope. All measurements of morphological features were made under oil at 1000× magnification using Zeiss Axiovision 4.7.1 software (Carl Zeiss MicroImaging 2008). The Axiovision software provided a measurement interface, allowing individual morphological traits to be enhanced via the zoom tool and measurements to be made to two decimal places (± 0.02µm) when using the calibrated measurement tool. All measurements, with the exception of pore counts, were made only on individual pollen grains that were lying flat on their isopolar axis.

Statistical Analysis

Before completing statistical tests, all quantitative variables were tested for normality via the use of exploratory QQ plots (Appendix B). The assumption of equal variance was tested using standard F-tests and boxplots to display species variance by trait. Interspecies differences in arci strength, annulus protrusion, grain shape and pore number were tested for significance using Wilcoxon rank-sum tests to compare sample modes. In the case of grain shape, data were first converted into numerical categories. To test the association between pollen morphology and latitude, Pearson's correlation

was used for all quantitative variables. The association between qualitative traits and latitude was assessed using Spearman's rank correlation. All statistical analyses were performed using R (R Development Core Team 2005).

Nested ANOVA

Due to the hierarchical sampling design (i.e., pollen samples are from only one of the three species and pollen grains are from a specific sample), nested ANOVA analysis was performed for each quantitative trait. The nested model allows for partitioning of the total variability (i.e., in each morphological trait) into components explained by each of the nested factors (i.e., between species, between individuals within a species and between pollen grains from each individual sample) by incorporating a series of ANOVA models, each with different error terms. Nested ANOVA was used to test the null hypothesis that the means of each quantitative trait do not vary between all three species. Explained variability is calculated by subtracting the variability that is unexplained by the tested factor from the total variability explained by a reduced model that does not contain the factor of interest (Logan 2010). The amount of variation in the response variable attributable to a given nested factor is noted as percent variance for each full nested model. Where the assumption of balanced nested design was not met, procedures for unbalanced models were used. To test the null hypothesis that means do not differ between specific alder species, pair-wise nested ANOVA models were performed. A Bonferroni correction was applied to each pair-wise model to adjust the p-value for multiple comparisons and decrease the probability of Type I statistical errors. ANOVA model significance was set at $\alpha=0.05$.

To determine the degree of overlap between morphological trait distributions

between species, Mann-Whitney U two-way comparisons were performed between each of the three alder species, for each quantitative trait. As per Clegg et al. (2005), the resulting U statistic was scaled by the multiplier $(2/n_1n_2)$, where n_1 and n_2 are the number of pollen grains included in the dataset for each species. The resulting U statistic (a number between 0-100) gives a quantitative measure of variable distribution overlap, with 0 indicating no overlap in trait distribution and 100 indicating complete overlap.

Classification and Regression Trees (CART)

Determination of a method for identifying alder pollen to species via multiple morphological traits was performed using Classification and Regression Tree analysis (CART). Classification and Regression Trees use recursive partitioning of independent variables to create a binary classification (decision) tree that is conceptually similar to a standard dichotomous identification key (Breiman et al. 1984). CART graphical output consists of a tree encompassing internal binary nodes that coincide with specific splitting variables and threshold values, and terminal nodes that unify data into a specific class (i.e., a species). The probability of correct classification for each specific terminal node is quantified via the number of correctly classified cases within that node. Total model classification error is a function of misclassification across the terminal nodes. The tree that results from CART modelling is pruned to minimize cross-validation error and avoid over-fitting the data. This is done via an assessment of model complexity parameters. Tree nodes that over-fit data are removed until the decision tree is of an optimal size and misclassification cost is minimized (Breiman et al. 1984; Therneau et al. 2009).

A classification tree including all three alder species was grown using all quantitative traits (arci width, grain diameter, exine thickness, and annulus height, width

and area), arci strength and pore protrusion as model inputs. To determine the accuracy of the resulting decision tree, a randomly selected test set of 30% of the data was held in reserve and used to test model predictions. As per Lindbladh et al. (2002), the consistency of splitting variables and threshold values for isolating the pollen of one species from the pooled dataset was assessed by creating binary classification trees for *A. viridis* subsp. *sinuata* and *A. rubra* pollen, respectively. A classification tree was also grown using a 'reduced species dataset' including only data collected from *A. rubra* and *A. viridis* subsp. *sinuata* pollen. Again, a 30% test set was held in reserve, allowing comparison of classification accuracy for *A. rubra* and *A. viridis* subsp. *sinuata* pollen to be made between the full CART model and reduced CART model. Two further CART models were derived for the full dataset and reduced species dataset, but with categorical traits excluded.

CART was chosen as the primary statistical tool for use in this study because previous research has shown CART to be a useful nonparametric method for classifying pollen to species in *Picea* (Lindbladh et al. 2002; Lindbladh et al. 2007) and *Pinus* (Barton et al. 2011). Moreover, CART analysis can provide a more powerful statistical model than the more commonly used discriminate function analysis when comparing morphological traits that overlap between species (Breiman et al. 1984; Lindbladh et al. 2002). CART models can also incorporate rank and ordinal data. This is not true of discriminant function analysis, which assumes that multivariate data is from a normal distribution with common covariance (Breiman et al. 1984). Here, discriminant function analysis is used only to graphically represent the distance between alder samples via input of quantitative trait sample means. CART analysis was performed using the 'Rpart'

package (Therneau et al. 2009) in the R statistical environment (R Development Core Team 2005). A discriminant function graph was created using the classification models procedure in SPSS (SPSS 11.5.0 2002).

Random Forest Analysis

Random Forest Analysis was performed with the goal of determining an unbiased estimate of the generalized (out of the bag - OOB) error rates involved in classifying pollen grains to species, as well as generating an overall ranked list of trait importance in species identification. Supplementing CART modelling with Random Forest analysis is necessary when making assessments of morphological trait importance because CART derived decision trees can be unstable i.e., small changes in the sample used to create the tree can equal changes in splitting variables (Sutton 2005). Random Forest models generate large quantities of bootstrapped trees via random variable sampling and classify data input by combining the results of all generated trees. OOB model error estimation removes the need for cross validation and test sets, as it is derived as an internal function of the Random Forest bootstrapped model. Ranking of morphological trait importance for classification is a function of each trait's Gini coefficient. Gini importance is derived by adding Gini decreases for each variable across all the bootstrapped trees in the Random Forest model. Gini decreases represent the reduction in Gini impurity when a parent node is split into two descendent nodes (Breiman 2001). Random Forest modelling was performed using R ('randomForest' package - Liaw and Wiener 2002) for both the full dataset and the reduced dataset that included only *A. rubra* and *A. viridis* subsp. *sinuata* pollen.

Sensitivity Analysis

Dataset Reductions in Size and by Region

In an effort to assess the influence of sample size on statistical results, the alder pollen dataset was reduced in two distinct ways: 1) overall sample size, where each sample represents one individual plant; and, 2) the number of grains measured per sample. These sample size reductions were performed randomly. The number of samples was reduced to n=15 and n=7 for each alder species. The number of grains measured per sample was also reduced randomly to 20 grains per sample and 10 grains per sample, for all samples across each species. To test the effects of reduced sample size in combination with fewer grains measured per sample, the number of grains measured per sample was again randomly reduced to 20 grains and 10 grains per sample, but with the reduction being applied to re-randomized n=15 and n=7 datasets.

To test whether regional sampling breadth influences statistical results and therefore whether pollen morphology varies by region, *A. rubra* and *A. viridis* subsp. *sinuata* pollen datasets were split into regional subsets. *Alnus rubra* samples were divided into two regional subsets, one composed solely of samples from Vancouver Island, the other composed of samples from the British Columbia mainland. *Alnus viridis* subsp. *sinuata* samples were divided into three regions: 'North' - northern BC / Yukon subset, 'Inland' - inland BC subset and 'Coast' - Vancouver Island/ coastal BC subset.

To test the impact of these random and regional dataset reductions on statistical results, all statistical analyses (nested ANOVA, Mann-Whitney U Tests, CART analysis, and Random Forest Analysis) were re-run on the reduced datasets and regional data subsets. For CART analysis, a 30% test set was again held in reserve for model testing.

However, the pollen grains from samples removed during dataset reduction also serve as a much larger test set for analyzing the accuracy of reduced models in classifying pollen from samples not included as model input. Test sets containing grains from alder samples not used in the model are defined in text and figures as 'Other Sample' - (OS) Test Sets. For number of grains within sample reductions, 'Other Grain' - (OG) test sets were created. Lastly, the test sets for the regional analyses are the pollen grains for each species (*A. viridis* subsp. *sinuata* and *A. rubra*) that were not used in the model, but grouped within the other regional subsets for each species. There are three test sets; 'North' and 'Inland' test sets for *A. viridis* subsp. *sinuata* and a 'Mainland' test set for *A. rubra* pollen. These regional test sets allow model classification accuracy across regions to be assessed.

Results

Pollen Morphology and Variability

All three of the alder species have stephanoporate pollen with three to six annulate pores (Fig. 3.1), with the endexine detached from that of the pores, forming a vestibulum. Alder pollen ranges in diameter from 16.2 µm (*A. viridis* subsp. *sinuata*) to 30.1 µm (*A. rubra*), when measured equatorially (Table 3.1). The annulus ranges from 1.4 µm (*A. viridis* subsp. *sinuata*) to 4.1 µm (*A. viridis* subsp. *sinuata*) in height, and from 4.7 µm (*A. incana* subsp. *tenuifolia*) to 10.19 µm (*A. incana* subsp. *tenuifolia*) in width. Exine thickness ranges from 1.2 µm (*A. viridis* subsp. *sinuata*) to 3.4 µm (*A. incana* subsp. *tenuifolia*). Strength of arci varies in all three alder species and grains with no visible arci occur in all three species. Arci range in width from 1.0 µm (*A. viridis* subsp. *sinuata*) to 2.9 µm (*A. rubra*). For all of the quantitative traits, the smallest mean dimensions by trait occur in *A. viridis* subsp. *sinuata* and the largest occur in *A. rubra*. Mean values for *A. incana* subsp. *tenuifolia* are intermediate across all quantitative traits (Table 3.1).

Alder pollen grains are characterized by a varying degree of natural morphological variability that is apparent across experimental scales. Within sample variability is well illustrated in Figure 3.2, where grain size, pore number and overall grain shape differ between the pollen grains of a single *A. incana* subsp. *tenuifolia* plant. Within species variability is also apparent; this is shown clearly by a linear discriminant function plot of all alder samples (Fig. 3.3). Distances between samples are derived from multi-trait differences in pollen morphology between each individual plant. Morphological dissimilarity between individual samples varies both within a species as

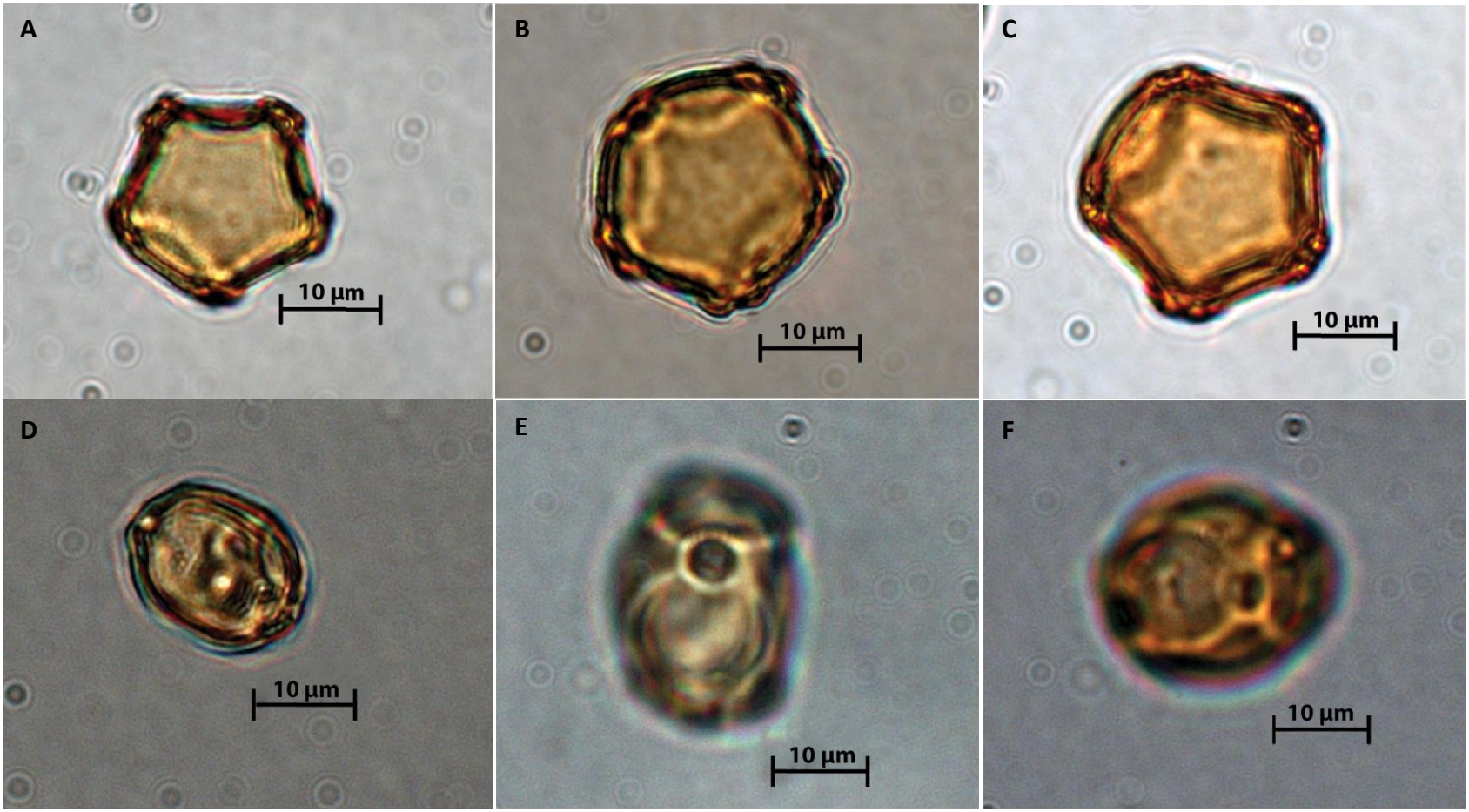


Figure 3.1: Isopolar vs. equatorial views of pollen grains (A and D) *Alnus viridis* subsp. *sinuata*, (B and E) *Alnus incana* subsp. *tenuifolia*, and (C and F) *Alnus rubra*. Photographs were taken at 1000× magnification under oil immersion.

Table 3.1: Summary of quantitative morphological traits for each alder species based on all measured grains.

	Alnus viridis* subsp. *sinuata			***Alnus incana* subsp. *tenuifolia***			***Alnus rubra***		
Trait	**Mean ± SE**	**Min**	**Max**	**Mean ± SE**	**Min**	**Max**	**Mean ± SE**	**Min**	**Max**
Arci Width (µm)	1.57 ± 0.02	0.00	2.69	1.67 ± 0.01	0.00	2.78	1.82 ± 0.01	0.00	2.90
Annulus Width (µm)	7.03 ± 0.02	4.73	9.27	7.51 ± 0.03	4.72	10.19	7.86 ± 0.02	5.84	9.97
Annulus Height (µm)	2.67 ± 0.01	1.51	4.09	2.86 ± 0.01	1.43	4.02	2.88 ± 0.01	1.88	3.99
Annulus Area (μm^2)	18.79 ± 0.12	8.66	37.51	21.67 ± 0.14	9.22	39.45	22.84 ± 0.14	11.80	34.45
Exine Thickness (µm)	1.91 ± 0.01	1.17	3.11	2.04 ± 0.01	1.33	3.24	2.11 ± 0.01	1.31	3.37
Grain Diameter (µm)	22.08 ± 0.06	16.22	28.83	22.95 ± 0.06	16.73	28.28	23.99 ± 0.06	18.39	30.05

Figure 3.2: Pollen diameter, shape and pore number variability within one sample of *Alnus incana* subsp. *tenuifolia* (Sample# RM25).

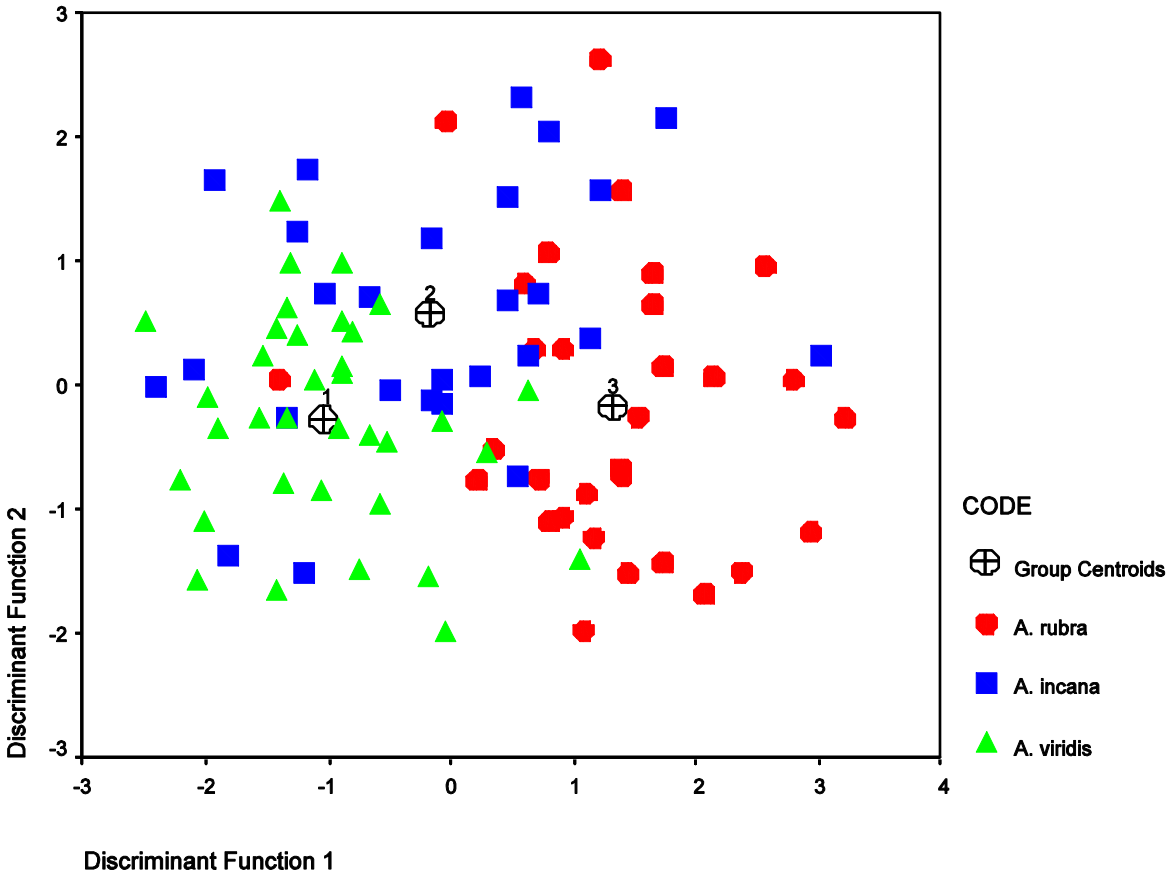


Figure 3.3: Linear discriminant function plot of all alder samples. Axes are determined by the two discriminant functions that best differentiate between the three species groups. The discriminant function model was created using quantitative trait means for each sample. Distance between plotted samples represents differences in pollen morphology between samples. Discriminant function 1 [D = 0.86(annulus width) + 0.65(diameter) + 0.62(arci width) + 0.49(annulus height) + 0.70(annulus area) + 0.46(exine thickness)] accounts for 88% of sample variance.

well as between species. While *A. incana* subsp. *tenuifolia* can generally be described as occupying an intermediate position on the x-axis (Discriminant Function 1) with *A. viridis* subsp. *sinuata* to the left and *A. rubra* to the right (Fig. 3.3), individual samples from each species deviate to varying degrees from this general trend. Intraspecific variability for each quantitative trait (Fig. 3.4) is therefore expected for each of the three alder species and for all measured morphological traits. Moreover, there is extensive overlap in variance distribution between species for all measured traits.

Morphological Trait Comparisons between Species

Qualitative Traits and Pore Number

The qualitative morphological traits vary within each species for each trait (Table 3.2; Fig. 3.5). The number of pores in all three species ranges from 3-6 pores per grain (Fig. 3.5A), with most pollen grains being 4- or 5-pored. Pore protrusion is also variable, with moderate protrusion (class 2) most common in all three species (Fig. 3.5B). The pollen of all three alder species differ significantly in arci strength and grain shape (Table 3.2). Arci are more pronounced in *A. incana* subsp. *tenuifolia* and *A. rubra* than in *A. viridis* subsp. *sinuata* (Fig. 3.5C). Overall grain shape varies in all three species with most grains classified as ‘mixed’ or ‘convex’ (Fig. 3.5D). Intraspecific variability across these categorical traits means that none of the traits on their own are sufficient for distinguishing the pollen of these three alder species.

Quantitative Trait Comparisons

Nested ANOVA models comparing traits between species pairs indicate that there are significant interspecific differences in morphology across all quantitative traits (Table 3.3). The sole exception to this is the non-significant difference in annulus height

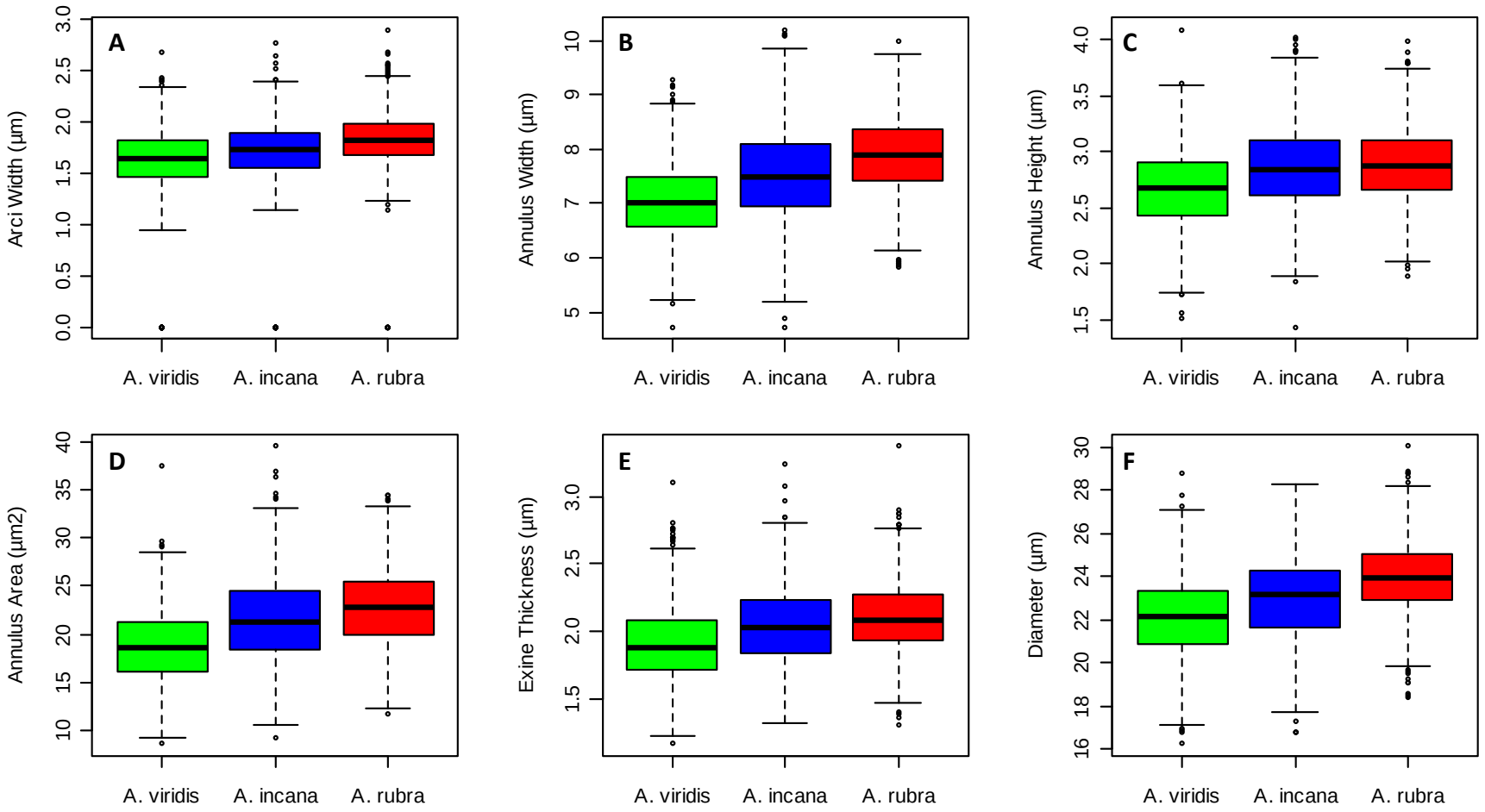


Figure 3.4: Boxplots of species median and within species variability for each quantitative morphological trait: arci width (A), annulus width (B), annulus height (C), annulus area (D), exine thickness (E) and grain diameter (F). Solid lines bisecting each boxplot represent the trait median for that species across all samples. Box edges mark the first and third quartiles. Whiskers extend to the smallest and largest non-extreme data points.

Table 3.2: Wilcoxon rank-sum tests for between species differences across qualitative traits.

	Wilcoxon Rank-Sum Test Comparison of Qualitative Traits between Species		
Trait by Species	**Pair-wise***	**W**	**P-value**
Pore Number			
Alnus viridis subsp. *sinuata*	**V-I**	202	**< 0.001**
Alnus incana subsp. *tenuifolia*	**V-R**	521	0.556
Alnus rubra	**I-R**	196	**< 0.001**
Pore Protrusion			
Alnus viridis subsp. *sinuata*	**V-I**	506	0.443
Alnus incana subsp. *tenuifolia*	**V-R**	516	0.494
Alnus rubra	**I-R**	469	0.167
Grain Shape			
Alnus viridis subsp. *sinuata*	**V-I**	336	**0.006**
Alnus incana subsp. *tenuifolia*	**V-R**	162	**< 0.001**
Alnus rubra	**I-R**	597	**0.002**
Arci Strength			
Alnus viridis subsp. *sinuata*	**V-I**	664	**0.002**
Alnus incana subsp. *tenuifolia*	**V-R**	974	**< 0.001**
Alnus rubra	**I-R**	221	**0.001**

* Pair-wise abbreviations: *A. viridis* subsp. *sinuata* (V), *A. incana* subsp. *tenuifolia* (I) and *A. rubra* (R)

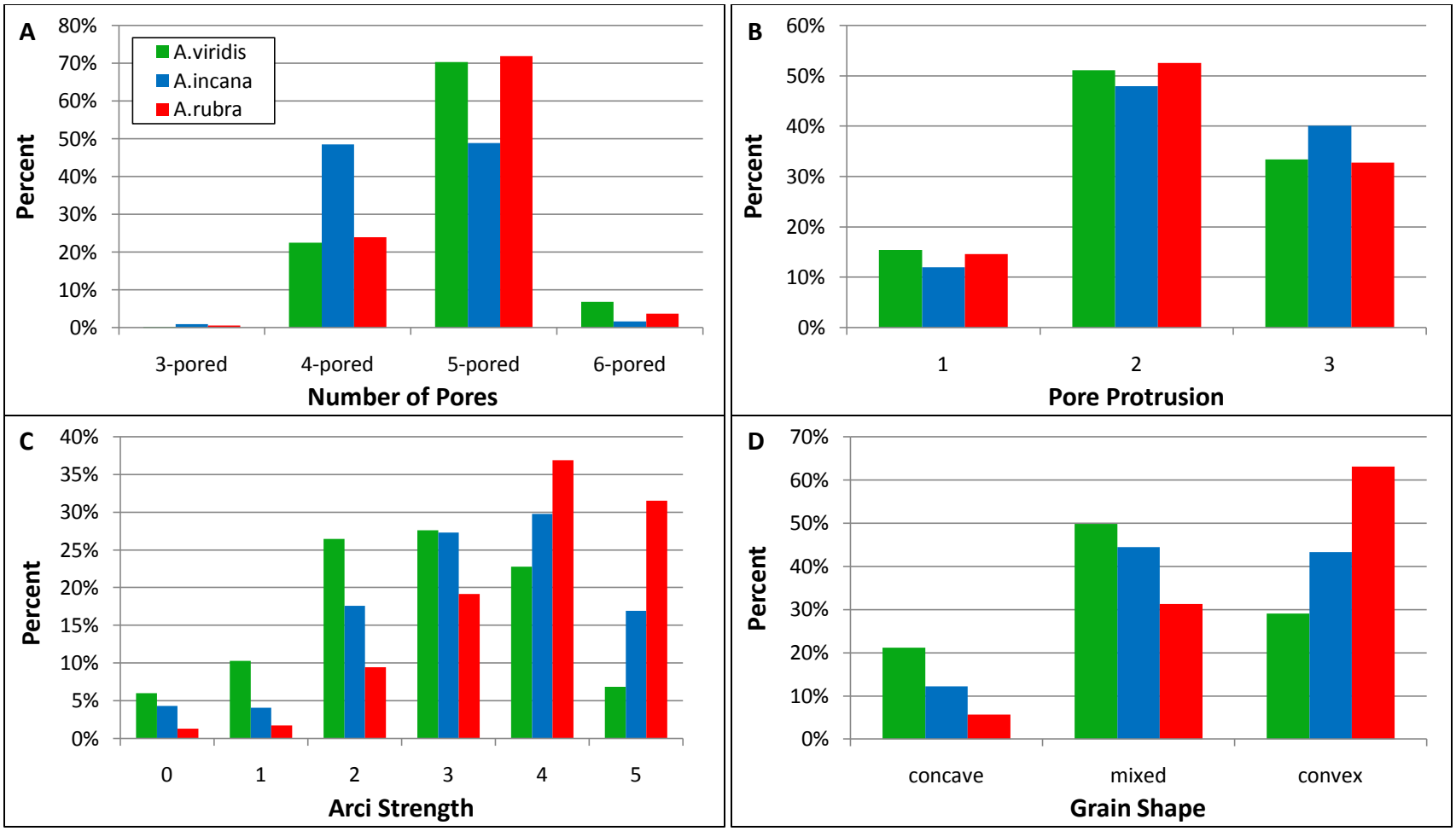


Figure 3.5: Pollen grain pore number (A), pore protrusion (B), arci strength (C), and grain shape (D), shown as percentages (%) using clustered bar graphs to represent species where *Alnus viridis* subsp. *sinuata* is green, *Alnus incana* subsp. *tenuifolia* is blue and *Alnus rubra* is red.

Table 3.3: Summary of nested ANOVA analysis for pair-wise species comparisons: *Alnus viridis* subsp. *sinuata* (V), *Alnus incana* subsp. *tenuifolia* (I), and *Alnus rubra* (R). Scaled U statistic is the percent overlap between species for each quantitative trait. Bolded numbers highlight significant P-values of less than 0.05.

		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic*: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	38.56	**< 0.001**	5.85	**< 0.001**	0.95	1.000	79.2
Alnus incana	**V-R**	233.27	**< 0.001**	6.04	**< 0.001**	1.13	1.000	57.1
Alnus rubra	**I-R**	67.67	**< 0.001**	7.56	**< 0.001**	1.10	1.000	78.9
Annulus Width								
Alnus viridis	**V-I**	283.58	**< 0.001**	16.09	**< 0.001**	1.68	**0.013**	67.1
Alnus incana	**V-R**	990.69	**< 0.001**	9.96	**< 0.001**	1.63	**0.021**	37.2
Alnus rubra	**I-R**	161.52	**< 0.001**	16.34	**< 0.001**	1.73	**0.009**	73.5
Annulus Height								
Alnus viridis	**V-I**	182.87	**< 0.001**	10.42	**< 0.001**	1.08	1.000	72.3
Alnus incana	**V-R**	261.28	**< 0.001**	9.97	**< 0.001**	1.37	0.432	64.6
Alnus rubra	**I-R**	1.56	1.000	9.96	**< 0.001**	1.12	1.000	98.0
Annulus Area								
Alnus viridis	**V-I**	345.53	**< 0.001**	16.87	**< 0.001**	1.43	0.248	65.1
Alnus incana	**V-R**	796.75	**< 0.001**	13.30	**< 0.001**	1.62	**0.023**	44.5
Alnus rubra	**I-R**	49.10	**< 0.001**	16.45	**< 0.001**	1.43	0.307	84.6
Exine Thickness								
Alnus viridis	**V-I**	156.46	**< 0.001**	16.21	**< 0.001**	1.51	0.107	74.5
Alnus incana	**V-R**	403.42	**< 0.001**	13.71	**< 0.001**	1.62	**0.021**	56.7
Alnus rubra	**I-R**	44.91	**< 0.001**	18.46	**< 0.001**	1.35	0.651	86.5
Grain Diameter								
Alnus viridis	**V-I**	162.29	**< 0.001**	23.67	**< 0.001**	1.05	0.360	74.8
Alnus incana	**V-R**	920.39	**< 0.001**	18.91	**< 0.001**	0.87	1.000	42.5
Alnus rubra	**I-R**	231.03	**< 0.001**	20.35	**< 0.001**	1.07	1.000	71.1

*U statistic is derived from Mann-Whitney U two-way comparisons between species. U is then multiplied by $(2/n_1n_2)$ to derive percent overlap.

between *A. incana* subsp. *tenuifolia* and *A. rubra*. There are also significant intraspecific differences (i.e., differences between the pollen morphology of individual plants within each species) for each quantitative trait. Statistical difference in morphology within an individual alder plant (i.e., between grains within samples) is much less common, suggesting that, in general, individual alder plants produce pollen that is morphologically similar. Scaled U statistics indicate extensive overlap (37.2% - 98.0%) in the trait distributions between all three alder species (Table 3.3), with the greatest amount of morphological overlap between *A. rubra* and *A. incana* subsp. *tenuifolia* and the least amount of overlap between *A. rubra* and *A. viridis* subsp. *sinuata*. Nested ANOVA models show that the greatest source of variance occurs between species for all quantitative traits, accounting for 74.4% to 91.4% of model variance (Table 3.4). Variance in pollen morphology within an individual alder plant is of secondary importance, accounting for 6.1% to 21.8% of variance. The greatest interspecific variance occurs in grain diameter (89.1%) and annulus width (91.4%). While there are significant *interspecific* differences in mean values for each trait (Table 3.3), the large amount of *intraspecific* morphological variability as well as *interspecific* overlap in morphology precludes the use of mean values for pollen identification to species. As with the categorical morphological traits, none of the quantitative traits on their own can be used for identifying alder pollen to species.

Multi-Trait Modelling for Identifying Alder Pollen to Species

Full Model 'All Species' Classification

The multi-trait classification model derived using Classification and Regression Tree (CART) analysis provides a binary decision tree for separating the pollen of all three

Table 3.4: Nested ANOVA and variance component analyses comparing the six quantitative traits between all three alder species.

Quantitative Trait by Source of Variance		Nested ANOVA Model		Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.***	**% Variance**
Between Species	2	31.90	15.95	0.4850	74.4
Between Individuals	90	80.93	0.90	0.0250	3.8
Between Grains	2696	382.85	0.14	0.1420	21.8
Annulus Width (µm)					
Between Species	2	371.18	185.59	5.8158	91.4
Between Individuals	90	477.42	5.30	0.1637	2.6
Between Grains	2696	1043.40	0.39	0.3870	6.1
Annulus Height (µm)					
Between Species	2	30.02	15.01	0.4516	77.6
Between Individuals	90	90.16	1.01	0.0303	5.2
Between Grains	2696	268.49	0.10	0.1000	17.2
Annulus Area (μm^2)					
Between Species	2	8635.40	4317.70	133.7310	89.0
Between Individuals	90	15483.00	172.04	5.3603	3.6
Between Grains	2696	30268.00	11.23	11.2270	7.5
Exine Thickness (µm)					
Between Species	2	20.99	10.50	0.3132	80.5
Between Individuals	90	71.49	0.79	0.0247	6.3
Between Grains	2696	136.27	0.05	0.0510	13.1
Grain Diameter (µm)					
Between Species	2	1781.80	890.87	27.3703	89.1
Between Individuals	90	3815.50	42.39	1.3457	4.4
Between Grains	2696	5435.40	2.02	2.0160	6.6

* Var.Comp. = the variance component (i.e., variance explained by each nested factor)

alder species (Fig 3.6). Tree interpretation and pollen grain classification begins at the top of the tree. If the internal splitting variables, located on internal tree nodes, are true for an individual pollen grain, then the right branch is followed. If the criterion is not met, the left branch is followed. Branches terminate in 'end nodes' which classify pollen grains to species and give the probability of correct classification. The CART model derived for all three alder species classifies pollen grains based on annulus width, arci strength and exine thickness. For example, if a pollen grain has a mean annulus width of >7.52 µm and arci strength >3.5, the model will classify the grain as *A. rubra* with 62.5% accuracy (Fig. 3.6). Following similar classification protocol the model also provides pathways for classifying pollen grains as *A. viridis* subsp. *sinuata* and *A. incana* subsp. *tenuifolia*.

CART model classification accuracy is a function of end node probabilities. In the model including all three alder species, end node probabilities range from P=0.440–0.625. Of the 540 *A. incana* subsp. *tenuifolia* grains used to create the decision tree, only 5.5% are classified accurately (Table 3.5). Model accuracy is 89.1% and 61.0% for *A. viridis* subsp. *sinuata* pollen and *A. rubra* pollen, respectively. Total model classification error is 44.6%, which is largely a result of the misclassification of *A. incana* subsp. *tenuifolia* grains, 32.3% of which are classified by the CART model as *A. rubra* and 62.2% as *A. viridis* subsp. *sinuata*. When model classification accuracy is tested using the reserved test-set (Table 3.5), 89.8%, 3.3% and 58.8% of *A. viridis* subsp. *sinuata*, *A. incana* subsp. *tenuifolia* and *A. rubra* pollen grains, respectively, are classified to species accurately. The intermediate morphology of *A. incana* subsp. *tenuifolia* pollen prevents

the model from accurately classifying these grains to species, which in turn inflates overall model classification error.

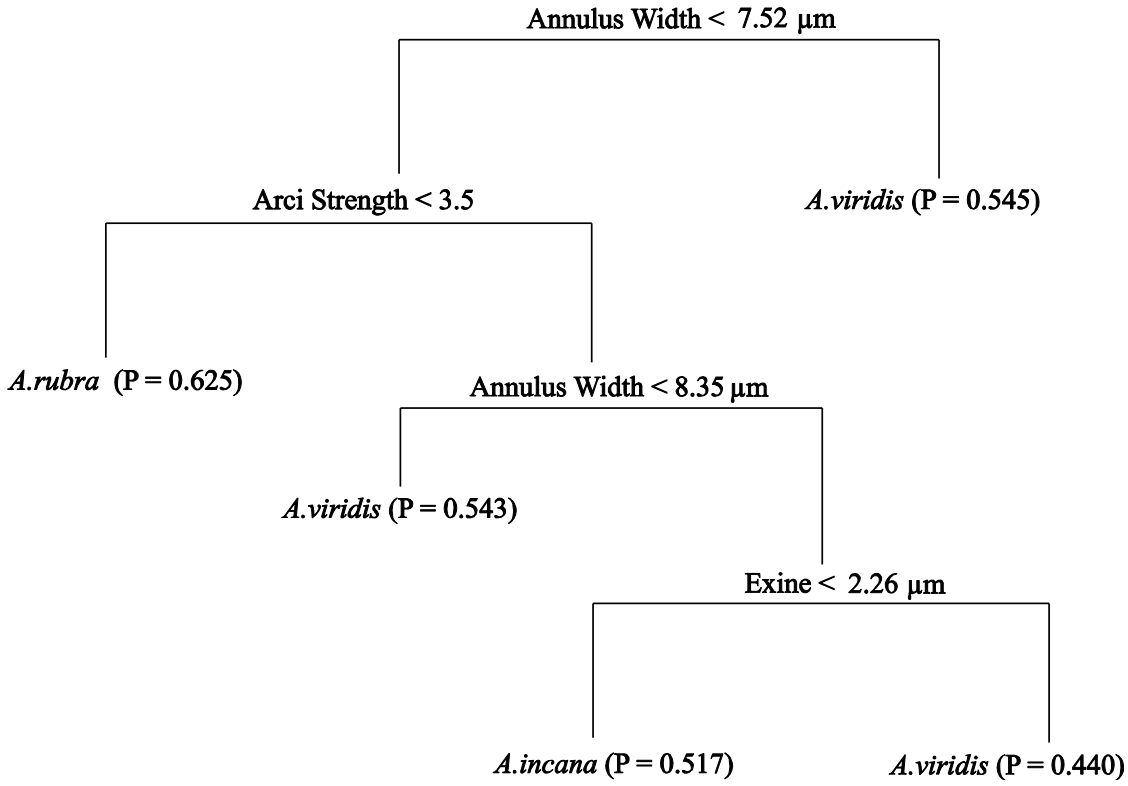


Figure 3.6: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen. All quantitative traits, arci strength and pore protrusion are model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.5: Species classification for full dataset (all species) CART model. Model classification of the reserved 30% test set is also shown. Overall model error is 44.6%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=700)		***A. incana* subsp. *tenuifolia*** (*n*=540)		***A. rubra*** (*n*=620)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
A. viridis	Model	624	89.1	336	62.2	224	36.2
	Test Set	314	89.8	174	64.6	119	38.5
A. incana	Model	10	1.5	30	5.5	18	2.8
	Test Set	9	2.6	9	3.3	8	2.7
A. rubra	Model	66	9.4	174	32.3	378	61.0
	Test Set	27	7.6	87	32.1	183	58.8

These trends of misclassification and high model error are even more pronounced if the CART model is built using only quantitative traits. The resulting decision tree classifies pollen solely on the basis of annulus width and terminates in only two end nodes, classifying pollen grains as either *A. rubra* or *A. viridis* subsp. *sinuata* (Fig. 3.7).

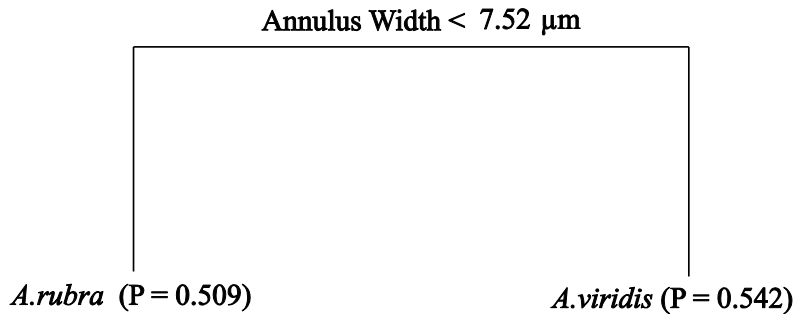


Figure 3.7: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen. Only quantitative traits are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Thus, in this model, classification error for *A. incana* subsp. *tenuifolia* is 100%. Of the *A. incana* subsp. *tenuifolia* grains used to make the model, approximately half are classified as *A. viridis* subsp. *sinuata*, the other half as *A. rubra* (Table 3.6). Overall model error is high (46.7%) due to the erroneous splitting of *A. incana* subsp. *tenuifolia* grains into two different species.

Binary Species Classification Trees

To determine which traits are most important in differentiating the pollen of an individual species from the combined datasets of all other species, binary 'species vs. other' CART trees are used. The binary CART model for the classification of *A. viridis* subsp. *sinuata* distinguishes the pollen of this species based on annulus width, arci

Table 3.6: Species classification for full dataset (all species included but no categorical traits included) CART model. Model classification of the reserved 30% test set is also shown. Total model error is 46.7%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=700)		***A. incana* subsp. *tenuifolia*** (*n*=540)		***A. rubra*** (*n*=620)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
A. viridis	Model	553	79.0	261	48.3	181	29.2
	Test Set	250	71.4	153	56.8	84	27.1
A. incana	Model	0	0.0	0	0.0	18	0.0
	Test Set	0	0.0	0	0.0	0	0.0
A. rubra	Model	147	21.0	279	51.6	0	70.8
	Test Set	100	28.6	116	43.2	226	72.9

strength, grain diameter and arci width (Fig. 3.8). End node probabilities range from P=0.587-0.9810 (Fig. 3.8) and overall model error is 26.4%. However, only 59.6% of the *A. viridis* subsp. *sinuata* pollen grains used to derive this CART model, and 57.1% of test set grains, are classified accurately by the model (Table 3.7).

Similarly, the binary CART model for *A. rubra* pollen identifies annulus width, arci strength and arci width as the most important traits (Fig. 3.9). Annulus width is once again the initial splitting variable; however, grain diameter is not used by the model for separating *A. rubra* from pollen of the other two species. Overall model error is 24.6%, but only 56.8% and 51.9% of *A. rubra* grains used to derive the model and test set grains, respectively, are classified correctly (Table 3.7B). The intermediate pollen morphology of *A. incana* subsp. *tenuifolia* prevents the derivation of a binary CART model for this species. The intermediate pollen morphology of *A. incana* subsp. *tenuifolia* likely accounts for some of the classification error in the binary models for the other two species.

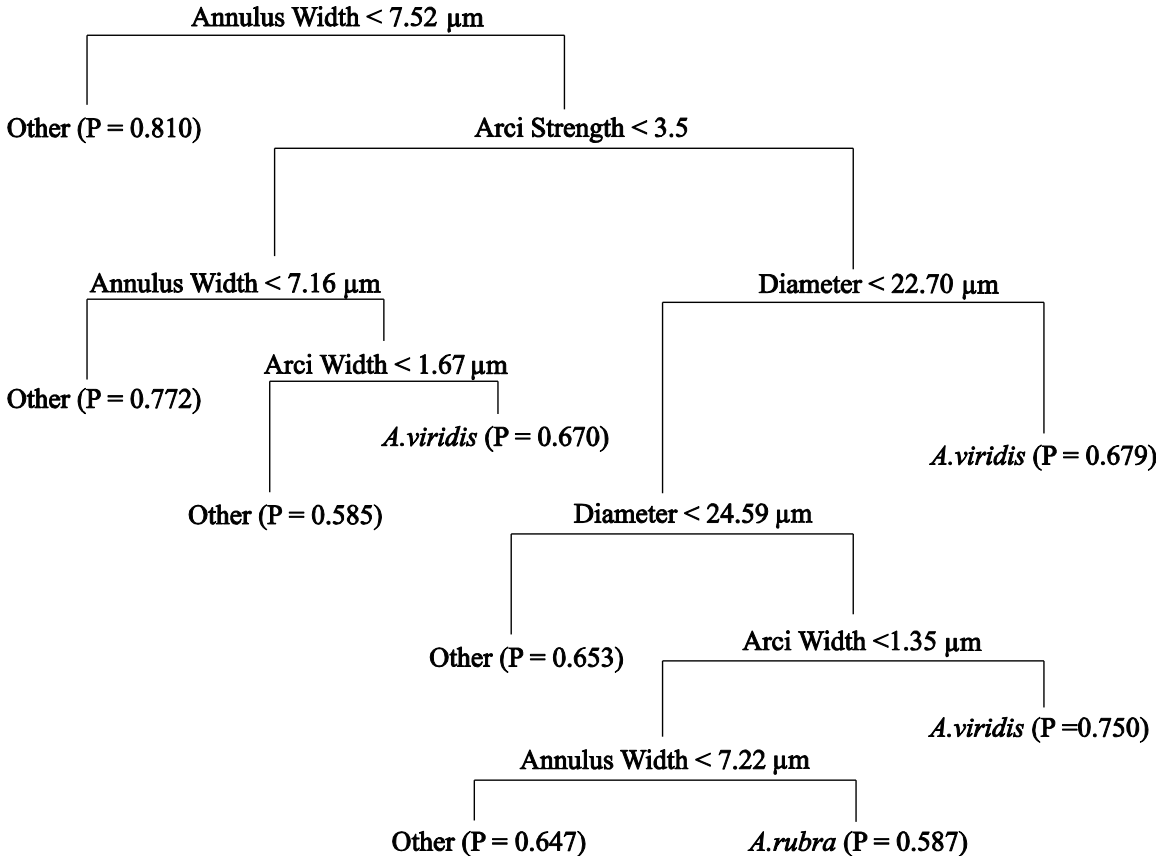


Figure 3.8: CART derived decision tree for the classification of *Alnus viridis* subsp. *sinuata* vs. pooled datasets from *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen (Other). All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate classification and the within model probability of correct classification.

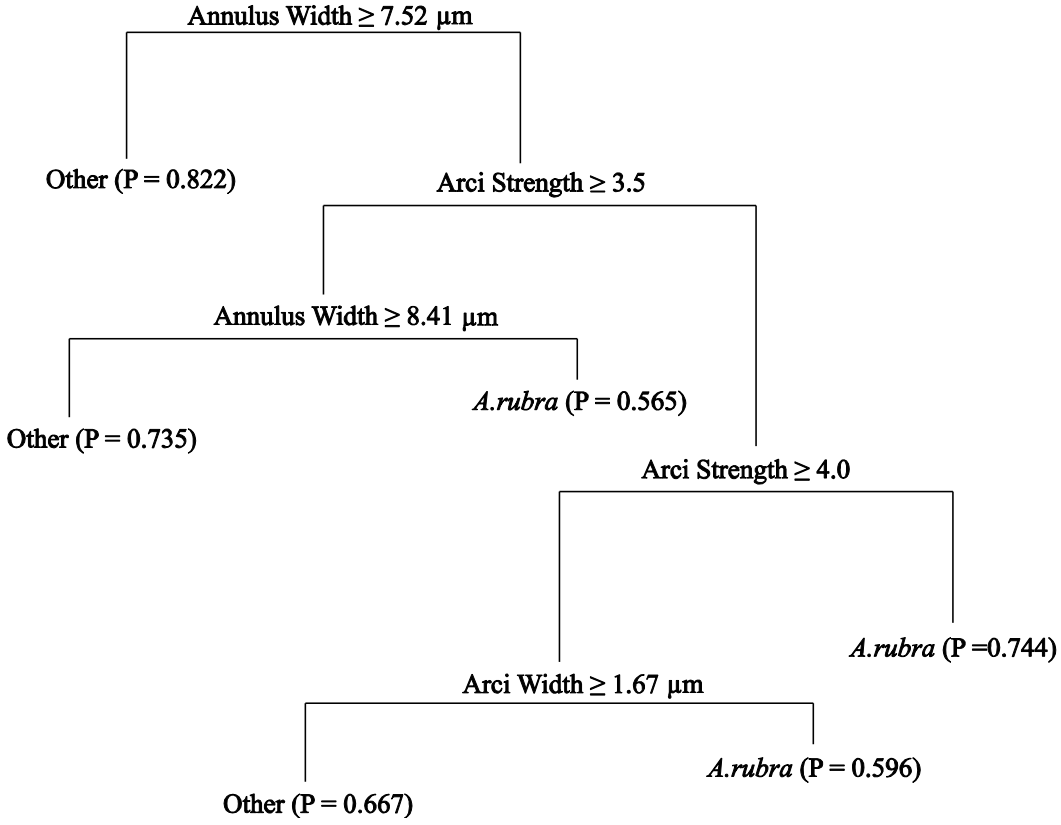


Figure 3.9: CART derived decision tree for the classification of *Alnus rubra* vs. pooled datasets from *Alnus viridis* subsp. *sinuata* and *Alnus incana* subsp. *tenuifolia* pollen (Other). All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate classification and the within model probability of correct classification.

Table 3.7: Species vs. 'Other' classification for *Alnus viridis* subsp. *sinuata* and *Alnus rubra* pollen vs. pooled datasets from the other two alder species. Model classification of reserved 30% test sets is also shown. Total model error for the *Alnus viridis* subsp. *sinuata* binary classification tree is 26.4%. Total model error for the *Alnus rubra* binary classification tree is 24.6%.

	Alnus viridis vs. Other			*Alnus rubra* vs. Other		
		Alder Species			**Alder Species**	
		A. viridis (*n*=700)			***A. rubra*** (*n*=620)	
Data	**Identified As**	**N**	**%**	**Identified As**	**N**	**%**
Model	***A. viridis***	417	59.6	***A. rubra***	352	56.8
Test Set		200	57.1		161	51.9
Model	**Other**	283	40.4	**Other**	268	43.2
Test Set		150	42.9		149	48.1

Reduced Model Classification

As the intermediate morphology of *A. incana* subsp. *tenuifolia* pollen prevents successful classification to the species level, CART analysis was conducted on a reduced dataset that excluded *A. incana* subsp. *tenuifolia* pollen. The reduced species CART model uses only annulus width and arci strength to differentiate *A. rubra* and *A. viridis* subsp. *sinuata* pollen (Fig. 3.10). The resulting decision tree isolates *A. rubra* pollen via annulus width ≥ 7.52 µm and arci strength ≥ 3.5. If arci strength is < 3.5, but annulus width is >8.31 µm, grains are also classified as *A. rubra*. Direct classification of pollen to *A. viridis* subsp. *sinuata* occurs if annulus width <7.52 µm. End node probability estimates are much higher for this reduced species model (P=0.608-0.852) than for CART models that include all species.

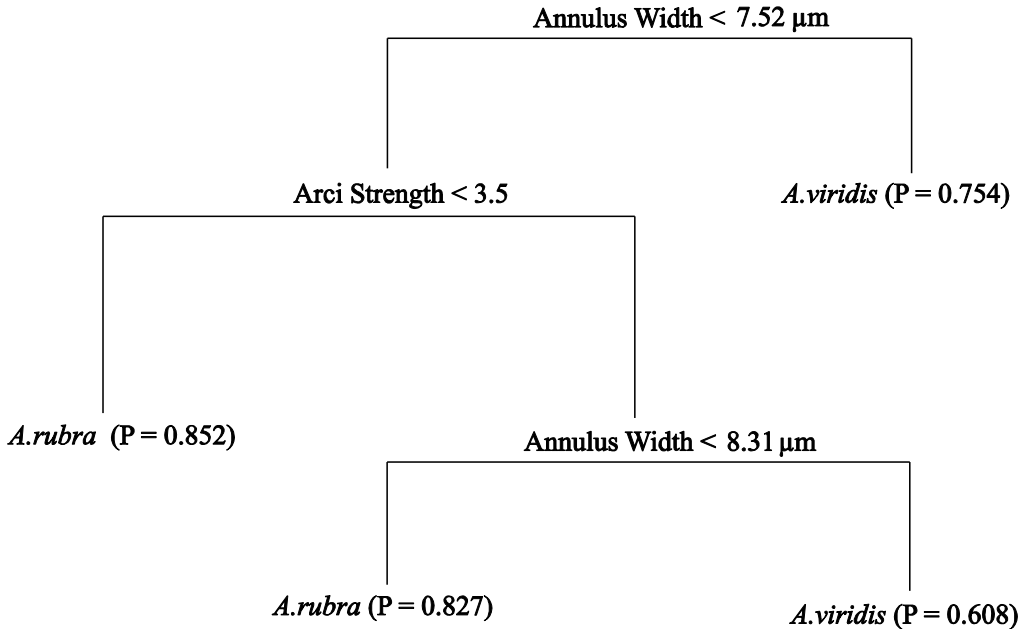


Figure 3.10: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata* and *Alnus rubra* pollen. *Alnus incana* subsp. *tenuifolia* has been removed from the model dataset. All quantitative traits, arci strength and pore protrusion are included as model parameters. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Model and test set accuracy for *A. viridis* subsp. *sinuata* and *A. rubra* classification are also improved when *A. incana* subsp. *tenuifolia* is excluded from the dataset (Table 3.8). The model classifies *A. viridis* subsp. *sinuata* pollen most accurately, with 90.2% of grains used in creating the model correctly identified. *Alnus viridis* subsp. *sinuata* grains reserved as a test set were classified accurately 91.7% of the time. Model classification for *A. rubra* pollen is less accurate at 61.8%, with 58.8% of *Alnus rubra* test set gains classified accurately. Total model error is 23.1%.

Table 3.8: Species classification for reduced species dataset (*Alnus incana* subsp. *tenuifolia* removed) CART model. Model classification of the reserved 30% test set is also shown. Overall model error is 23.1%.

		Alder Species			
		A. viridis* subsp. *sinuata (*n*=700)		***A. rubra*** (*n*=620)	
Identified As	**Data**	***n***	**%**	***n***	**%**
A. viridis	Model	631	90.2	237	38.2
	Test Set	321	91.7	128	41.2
A. rubra	Model	69	9.8	383	61.8
	Test Set	29	8.3	182	58.8

To determine the model that best classifies *A. viridis* subsp. *sinuata* and *A. rubra* pollen grains into species, CART analysis was also performed on the reduced species dataset using only quantitative traits. As with the previous model, the resulting decision tree (Fig. 3.11) starts by classifying grains based on annulus width, but then uses exine thickness and pollen diameter. The model classifies *A. viridis* subsp. *sinuata* and *A. rubra* pollen grains correctly 83.5% and 67.7% of the time, respectively (Table 3.9). Test set grain classification is comparable to model prediction, with *A. viridis* subsp. *sinuata* test grains classified with an accuracy of 82.6% and *A. rubra* test set grains classified 70.6%

correctly. The overall error for this two species model based on quantitative morphological traits only is 23.9%.

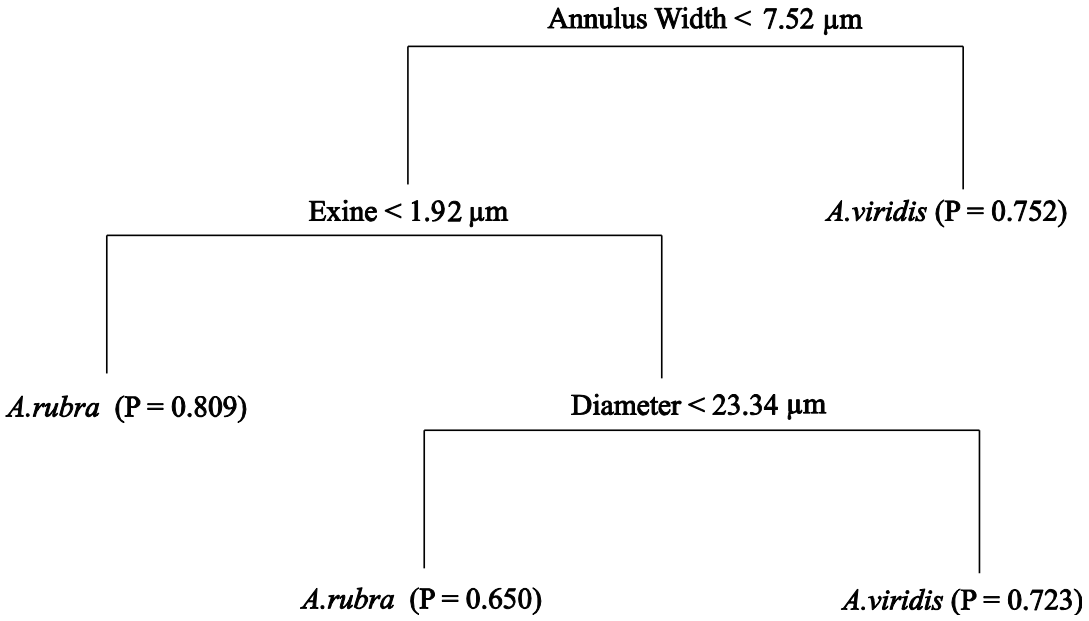


Figure 3.11: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata* and *Alnus rubra* pollen. *Alnus incana* subsp. *tenuifolia* has been removed from the model dataset. Only quantitative traits are included as model parameters. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.9: Species classification accuracy for reduced species dataset (*Alnus incana* subsp. *tenuifolia* removed) CART model. Model derived from quantitative traits only. Model classification of the reserved 30% test set is also shown. Total model error is 23.9%.

		Alder Species			
		A. viridis* subsp. *sinuata (*n*=700)		***A. rubra*** (*n*=620)	
Identified As	**Data**	***n***	**%**	***n***	**%**
A. viridis	Model	584	83.5	200	32.3
	Test Set	289	82.6	91	29.4
A. rubra	Model	116	16.5	420	67.7
	Test Set	61	17.4	219	70.6

Morphological Trait Importance for Classification

Random Forest analysis can be used to compare the importance of the various morphological traits across models (Table 3.10). When trait importance is ranked based on Gini number, annulus width, grain diameter and annulus area are ranked 1-3 for both datasets including categorical traits in the model. However, overall (OOB) model error is reduced by half when *A. incana* subsp. *tenuifolia* is removed from the dataset (44.1% to 21.3%). When Random Forest analysis is performed on the full dataset and two species dataset with categorical variables removed, annulus width, grain diameter and annulus area are again ranked 1-3 for the all species model. Trait importance shifts for the reduced species model; instead, annulus width, diameter and exine thickness are ranked 1-3. For the Random Forest models excluding categorical traits, overall (OOB) model error is again reduced from 48.4% to 24.6% when *A. incana* subsp. *tenuifolia* is excluded (Fig. 3.10).

Sensitivity Analysis: Dataset Reductions in Size

Nested ANOVA Analysis

Nested ANOVA analyses on datasets with reduced sample sizes reveal that sample size impacts quantitative trait comparisons between species to a variable degree depending on the trait and how and to what extent sample size has been reduced. Tables 3.11 through 3.16 outline the results of nested ANOVA models for annulus width between the three alder species as sample size is reduced. Tables outlining the sensitivity of nested ANOVA models to dataset reductions for all other quantitative traits are included in Appendix C.

Table 3.10: (A) Results of Random Forest analysis for full dataset and reduced species dataset models with all quantitative traits as well as arci strength and pore protrusion included as model parameters. (B) Results of Random Forest analysis for full dataset and reduced species dataset models using only quantitative traits. Gini number for each trait is given, as is the rank in importance of each trait (ranks 1-3 are bolded).

	A. Quantitative & Qualitative Traits				**B. Quantitative Traits Only**			
	All Species (OOB Error Rate = 44.1%)		***A. rubra* and *A. viridis*** (OOB Error Rate = 21.3%)		**All Species** (OOB Error Rate = 48.4%)		***A. rubra* and *A. viridis*** (OOB Error Rate = 24.6%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	222.23	**1**	139.73	**1**	362.15	**1**	242.01	**1**
Diameter	213.22	**2**	122.76	**2**	360.73	**2**	207.93	**2**
Annulus Area	199.40	**3**	99.62	**3**	312.34	**3**	138.31	5
Exine Thickness	171.15	4	80.04	4	286.37	4	163.07	**3**
Arci Width	171.36	5	84.26	6	280.71	5	138.58	4
Annulus Height	157.28	6	60.53	7	245.48	6	95.66	6
Arci Strength	103.96	7	82.13	5	-	-	-	-
Pore Protrusion	49.88	8	19.28	8	-	-	-	-

When the number of samples per species is reduced to *n*=15 and *n*=7 (Tables 3.11 and 3.12), the full dataset model indicates significant morphological differences in pollen grains from the same individual for all pair-wise species comparisons (Table 3.4), but this is not the case when sample size is reduced to *n*=15. Significant differences in grains within individual plants are again found in the *n*=7 pair-wise nested model (Table 3.11). The pair-wise species level and individual plant level differences in annulus width remain significant in reduced *n*=15 and *n*=7 models (Table 3.11). Interspecific differences in annulus width remain the greatest source of model variance (Table 3.12), but this decreases from 90.0% to 59.9% of the total model variance, as sample size is reduced (Table 3.12).

If the number of pollen grains per sample is reduced (Tables 3.13 and 3.14), again, species level and individual level differences in annulus width are significant for all pair-wise comparisons (Table 3.13). Both pollen grain per sample reductions (20 and 10 grains per individual plant) show significant differences between pollen grains from individual plants, across all pair-wise comparisons of annulus width. This differs from the full dataset model (Table 3.4), where between grain differences in annulus width was significant for all species. The percent of variance attributable to between species differences in annulus width is more or less maintained as the number of grains measured per individual is reduced (Table 3.14).

When the number of samples per species and grains per sample are reduced, there are significant differences between species and individual plants, but no significant variability in grains within individuals for both the *n*=15 (20 grains per individual) and *n*=15 (10 grains per individual) reductions (Table 3.15). When the dataset is reduced to

Table 3.11: Summary of nested ANOVA analysis of annulus width for pair-wise species comparisons: *Alnus viridis* subsp. *sinuata* (V), *Alnus incana* subsp. *tenuifolia* (I), and *Alnus rubra* (R). ANOVA models for the *n*=15 samples per species dataset and *n*=7 samples per species dataset are shown. Scaled U statistic is the percent overlap in annulus width between species. Bolded numbers highlight significant P-values of less than 0.05.

Dataset		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
***n*=15 Sample / Species**	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	89.26	**< 0.001**	17.09	**< 0.001**	1.66	0.238	72.1
Alnus incana	**V-R**	421.33	**< 0.001**	7.93	**< 0.001**	1.22	1.000	38.3
Alnus rubra	**I-R**	114.2	**< 0.001**	14.6	**< 0.001**	1.4	1.000	66.9
***n*=7 Sample / Species**								
Alnus viridis	**V-I**	25.17	**< 0.001**	16.85	**< 0.001**	2.41	**0.048**	77.3
Alnus incana	**V-R**	91.32	**< 0.001**	13.81	**< 0.001**	3.17	**0.002**	58.5
Alnus rubra	**I-R**	19.24	**< 0.001**	22.42	**< 0.001**	2.97	**0.004**	80.9

Table 3.12: Summary of nested ANOVA analysis and variance component analysis for annulus width. ANOVA models for the *n*=15 and *n*=7 sample per species dataset are shown.

Dataset	Nested ANOVA Model			Variance Component Analysis	
***n*=15 Samples / Species**	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	165.45	82.73	5.1607	90.0
Between Individuals	42	223.63	5.32	0.1637	2.9
Between Grains	1305	532.14	0.41	0.4077	7.1
***n*=7 Samples / Species**					
Between Species	2	33.17	16.59	1.4300	59.9
Between Individuals	18	118.51	6.58	0.5624	23.6
Between Grains	608	239.67	0.39	0.3942	16.5

Table 3.13: Summary of nested ANOVA analysis of annulus width for pair-wise species comparisons: *Alnus viridis* subsp. *sinuata* (V), *Alnus incana* subsp. *tenuifolia* (I), and *Alnus rubra* (R). ANOVA models for dataset reductions to 20 and 10 grains measured per individual plant are shown. Scaled U statistic is the percent overlap in annulus width between species. Bolded numbers highlight significant P-values of less than 0.05.

Dataset		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
All Individuals / Species	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
20 grains / individual								
Alnus viridis	**V-I**	178.00	**< 0.001**	11.36	**< 0.001**	1.54	0.078	66.7
Alnus incana	**V-R**	698.60	**< 0.001**	6.2	**< 0.001**	1.26	1.000	35.4
Alnus rubra	**I-R**	135.04	**< 0.001**	10.86	**< 0.001**	1.28	1.000	68.9
10 grains / individual								
Alnus viridis	**V-I**	91.79	**< 0.001**	5.38	**< 0.001**	0.93	1.000	66.1
Alnus incana	**V-R**	415.15	**< 0.001**	4.4	**< 0.001**	1.17	1.000	33.1
Alnus rubra	**I-R**	83.56	**< 0.001**	6.56	**< 0.001**	1.32	0.993	65.8

Table 3.14: Summary of nested ANOVA analysis and variance component analysis for annulus width. ANOVA models for 20 grains/individual and 10 grains/individual reductions are shown.

Dataset	Nested ANOVA Model			Variance Component Analysis	
20 grains / individual	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	250.82	125.41	3.9361	88.3
Between Individuals	90	305.14	3.39	0.1510	3.4
Between Grains	1767	650.93	0.37	0.3684	8.3
10 grains / individual					
Between Species	2	148.72	74.36	2.3332	81.1
Between Individuals	90	182.43	2.03	0.1650	5.7
Between Grains	837	317.61	0.38	0.3795	13.2

Table 3.15: Summary of nested ANOVA analysis of annulus width for pair-wise species comparisons: *Alnus viridis* subsp. *sinuata* (V), *Alnus incana* subsp. *tenuifolia* (I), and *Alnus rubra* (R). ANOVA models are for combined dataset reductions where sample size is reduced to *n*=15 samples per species and number of grains is reduced to 20 and 10 grains per individual plant, and sample size is reduced to *n*=7 with number of grains per individual reduced to 20 and 10. Scaled U statistic is the percent overlap in annulus width between species. Bolded numbers highlight significant P-values of less than 0.05.

Dataset		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
***n*=15 Samples / Species**	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
20 grains / individual								
Alnus viridis	**V-I**	145.03	**< 0.001**	6.67	**< 0.001**	1.23	1.000	54.1
Alnus incana	**V-R**	319.99	**< 0.001**	6.63	**< 0.001**	1.04	1.000	36.2
Alnus rubra	**I-R**	27.06	**< 0.001**	8.82	**< 0.001**	1.09	1.000	79.4
10 grains / individual								
Alnus viridis	**V-I**	22	**< 0.001**	5.4	**< 0.001**	1.29	1.000	67.6
Alnus incana	**V-R**	126.55	**< 0.001**	2.78	**< 0.001**	1.34	1.000	40.6
Alnus rubra	**I-R**	32.7	**< 0.001**	4.59	**< 0.001**	1.79	0.149	68.0
***n*=7 Samples / Species**								
20 grains / individual								
Alnus viridis	**V-I**	11.79	0.120	11.51	**< 0.001**	0.63	1.000	71.2
Alnus incana	**V-R**	177.21	**< 0.001**	7.65	**< 0.001**	1.34	1.000	32.8
Alnus rubra	**I-R**	87.11	**< 0.001**	6.95	**< 0.001**	0.86	1.000	47.0
10 grains / individual								
Alnus viridis	**V-I**	40.41	**< 0.001**	2.98	**0.020**	1.28	1.000	47.9
Alnus incana	**V-R**	71.18	**< 0.001**	3.08	**0.014**	1.93	0.484	37.2
Alnus rubra	**I-R**	1.23	1.000	2.53	0.098	0.81	1.000	92.3

n=7, 20 grains per individual), species level differences in annulus width are lost for the *A. viridis* subsp. *sinuata* – *A. incana* subsp. *tenuifolia* species pair. The dataset reduction to *n*=7 (10 grains per individual) datasets results in the loss of all significant differences (i.e., between species, between individual plants within species and between grains within individuals) for the *A. incana* subsp. *tenuifolia* - *A. rubra* pair-wise comparison. A comparison of scaled U statistics (Table 3.11-3.16) shows that the least amount of morphological overlap in annulus width occurs between *A. rubra* and *A. viridis* subsp. *sinuata*, a trend that is maintained across sample size reductions. Sample size reduction combined with a reduction in the number of grains measured per plant changes the amount of variance attributable to between grain and between species differences in annulus width (Table 3.16). The two most reduced datasets (*n*=15, 10 grains per individual and *n*=7, 10 grains per individual) have increased between grain variance, 21.5% and 17.9% percent of total variance (Table 3.16), when compared to the full dataset, where 6.1% of total variance is explained by between grain differences (Table 3.3). Between species differences account for 71.9% and 78.9% of total variance in the *n*=15 (10 grains per individual) and *n*=7 (10 grains per individual) nested ANOVA models for annulus width (Table 3.16).

CART and Random Forest Analysis of Datasets Reduced in Sample Size

CART analysis allows a visual comparison of tree models resulting from each of the consecutive dataset reductions in sample size. Because CART outlines a possible method for distinguishing pollen grains to species, comparing CART models also allows an examination of the possible conclusions that could be drawn from using this method with smaller datasets.

Table 3.16: Summary of nested ANOVA analysis and variance component analysis for annulus width. ANOVA models for the *n*=15 (20 and 10 grains/individual) and *n*=7 (20 and 10 grains/individual) are shown.

Dataset		Nested ANOVA Model		Variance Component Analysis	
***n*=15 Samples / Species**	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
20 grains / individual					
Between Species	2	126.70	63.35	4.0273	88.4
Between Individuals	42	123.40	2.94	0.1270	2.8
Between Grains	855	341.57	0.40	0.3995	8.8
10 grains / individual					
Between Species	2	50.38	25.19	1.5527	71.9
Between Individuals	42	79.97	1.90	0.1440	6.7
Between Grains	405	187.94	0.46	0.4640	21.5
***n*=7 Samples / Species**					
20 grains / individual					
Between Species	2	74.91	37.48	4.8286	89.2
Between Individuals	18	66.29	3.68	0.1630	3.0
Between Grains	399	167.13	0.42	0.4189	7.7
10 grains / individual					
Between Species	2	26.78	13.39	1.7557	78.9
Between Individuals	18	19.80	1.10	0.0700	3.1
Between Grains	189	75.82	0.40	0.3989	17.9

a. Reductions in Sample Size

The CART decision tree model for the dataset randomly reduced to *n*=15 samples per species separates pollen grains based on diameter, arci strength, annulus width, arci width and annulus area, with a total model error of 42.2% (Fig. 3.12). This is a different suite of morphological traits compared to the full CART model (Fig. 3.6). Of the *A. viridis* subsp. *sinuata* grains used to create the CART model, 73.7% were classified correctly (Table 3.17). *Alnus incana* subsp. *tenuifolia* and *A. rubra* grains used to create the model were classified 17.7% and 82.3% correctly, respectively. As was the case with full dataset CART models, the 30% reserve test set of pollen grains is classified with accuracy that is consistent with the model accuracy. However, when the grains from those samples excluded from the dataset during the random process of dataset reduction (the OS Test Set) are used to test the CART model, model classification accuracy is substantially reduced (Table 3.17).

When the reduced *n*=7 samples per species dataset is used to produce a CART decision tree (Fig. 3.13), pollen grains are classified to species based on exine thickness, arci strength, diameter and annulus width. The *A. viridis* subsp. *sinuata* end node probability of P=0.471 (Fig. 3.13) is lower than any of the end node probabilities in the full dataset CART models (Fig. 3.6 and Fig. 3.7) and the *n*=15 CART model (Fig. 3.12). As was the case in the *n*=15 CART model, the decision tree does a much poorer job of accurately classifying grains from the OS test set (Table 3.18).

If *A. incana* subsp. *tenuifolia* is removed from the reduced *n*=7 samples per species dataset, the derived CART model (Fig. 3.14), has reduced model error (14.9%) when compared to the CART tree excluding *A. incana* subsp. *tenuifolia*, but including all

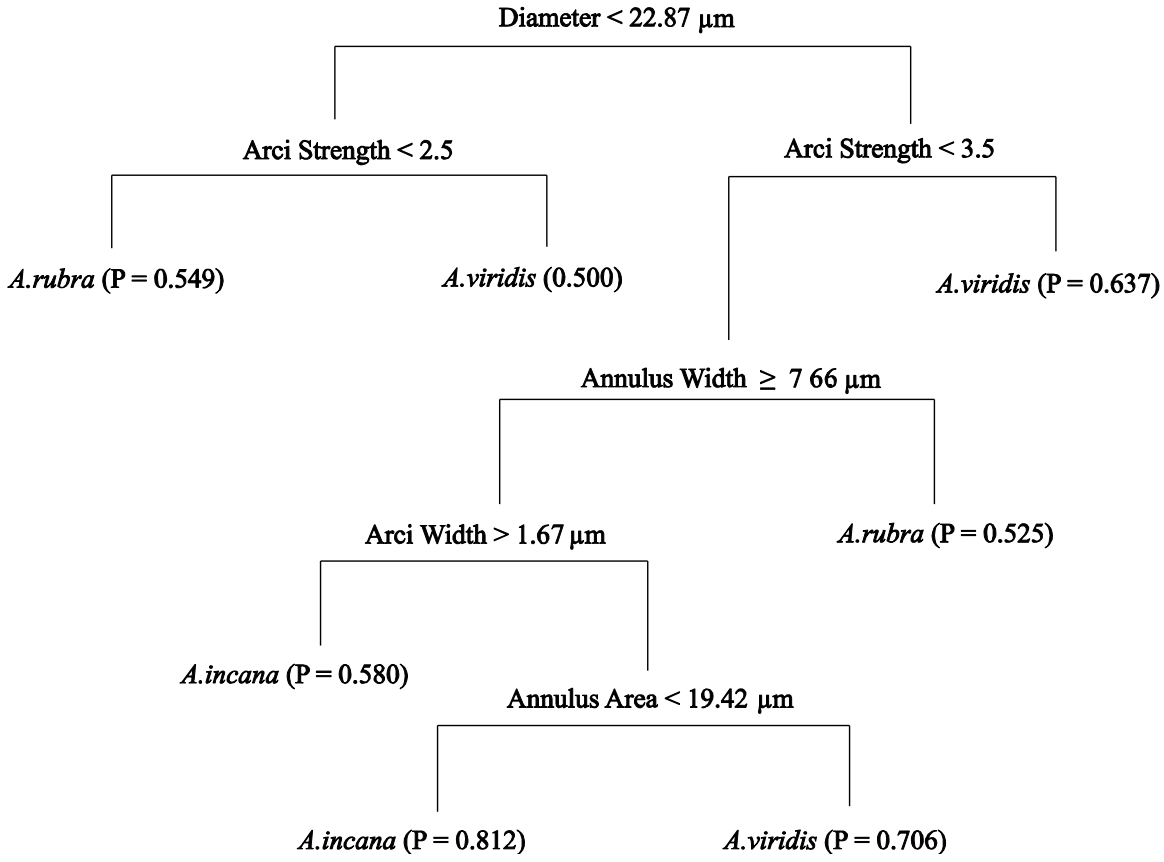


Figure 3.12: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when sample size is reduced to *n*=15 per species. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.17: CART model species classification for reduced *n*=15 samples per grain dataset. Model classification of reserved 30% test set and of the Other Sample (OS) test set is also shown. Total model error is 42.2%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n* = 300)		***A. incana* subsp. *tenuifolia*** (*n* = 300)		***A. rubra*** (*n* = 300)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	221	73.7	102	34.0	40	13.3
	Test Set	102	68.0	61	40.7	21	14.0
	OS Test	396	66.0	175	48.7	103	21.5
Alnus incana	Model	19	6.3	53	17.7	13	4.3
	Test Set	15	10.0	18	12.2	6	40.0
	OS Test	61	10.2	27	7.6	51	10.6
Alnus rubra	Model	60	20.0	145	48.3	247	82.3
	Test Set	33	22.0	71	47.3	123	82.0
	OS Test	143	23.8	157	43.7	326	67.9

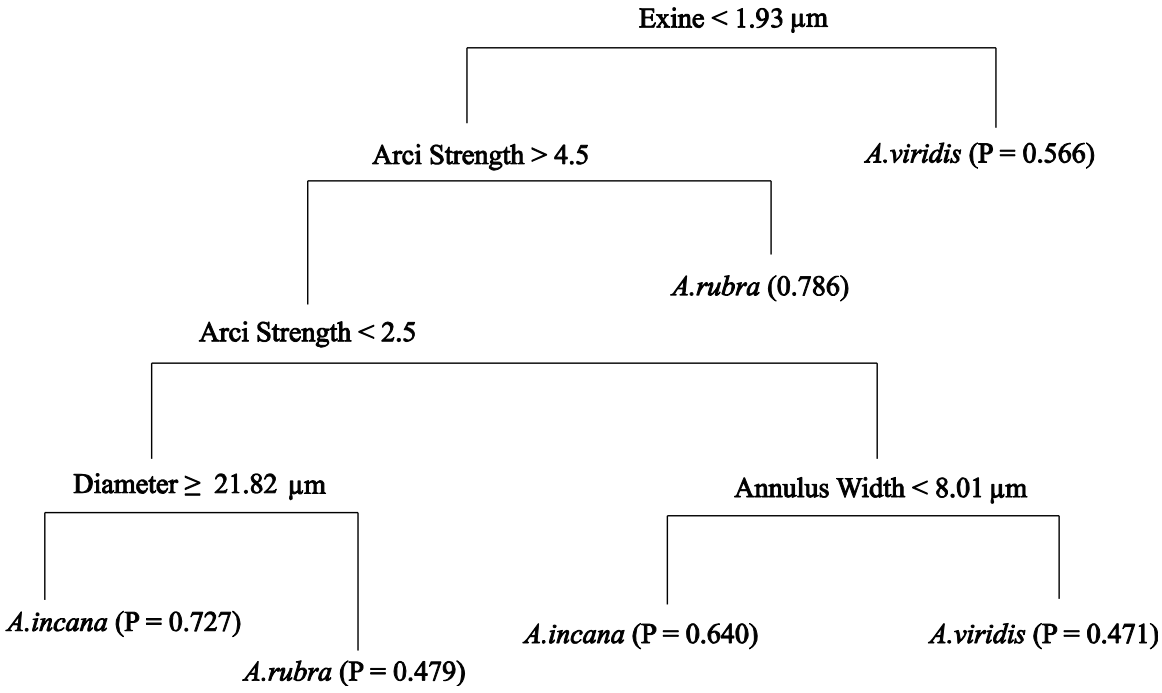


Figure 3.13: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen grains, when sample size is reduced to *n*=7 per species. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.18: CART model species classification for the reduced *n*=7 samples per species dataset. Model classification of reserved 30% test set and of the Other Sample (OS) test set is also shown. Total model error is 44.1%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=140)		***A. incana* subsp. *tenuifolia*** (*n*=140)		***A. rubra*** (*n*=140)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	114	81.4	58	41.7	41	29.3
	Test Set	51	72.9	38	54.3	23	32.9
	OS Test	610	72.6	278	46.3	211	29.3
Alnus incana	Model	5	3.6	32	23.0	10	7.1
	Test Set	4	5.7	18	25.7	4	5.7
	OS Test	80	9.5	49	8.2	35	4.9
Alnus rubra	Model	21	15.0	49	35.3	89	63.6
	Test Set	15	21.4	14	20.0	43	61.4
	OS Test	150	17.9	273	45.5	474	65.8

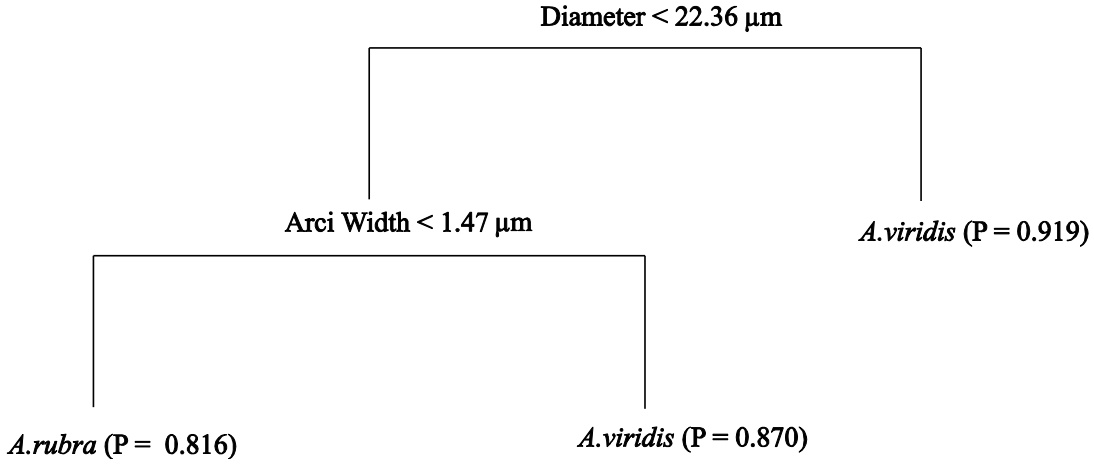


Figure 3.14: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, and *Alnus rubra* pollen (*Alnus incana* subsp. *tenuifolia* data is excluded), when the number of samples per species is reduced to *n*=7. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

samples (23.9%, Fig. 3.10). The model differentiates *A. viridis* subsp. *sinuata* and *A. rubra* pollen grains based on only two traits, diameter and arci strength. Within model classification error for each species is very low with 79.3% and 92.1% of *A. viridis* subsp. *sinuata* and *A. rubra* pollen identified correctly (Table 3.19). However, only 60.4% and 64.4% of OS test set grains are identified correctly for each species (Table 3.19).

Table 3.19: CART model species classification for *n*=7 samples per species, *Alnus viridis* subsp. *sinuata* and *Alnus rubra* (*Alnus incana* subsp. *tenuifolia* removed) dataset. Model classification of OS test set is also shown. Total model error is 14.3%.

		Alder Species			
		A. viridis subsp. sinuata (*n*=140)		***A. rubra*** (*n*=140)	
Identified As	**Data**	***n***	**%**	***n***	**%**
A. viridis	Model	111	79.3	11	7.9
	OS Test	509	60.6	256	35.6
A. rubra	Model	29	20.7	129	92.1
	OS Test	331	39.4	464	64.4

In general, when all species are included, CART model error increases as the number of samples per species is reduced (Tables 3.6, 3.17 and 3.18). However, the opposite is true when *A. incana* subsp. *tenuifolia* is removed from the dataset. The key diagnostic traits for species separation in the CART models also change as sample size is reduced (Figs. 3.6, 3.12, 3.13, 3.14). These trends are echoed by Random Forest models of the *n*=15 and *n*=7 samples per species datasets (Table 3.20). As the full dataset is reduced in size, there are substantial shifts in which morphological traits are identified as most important for species separation (Table 3.20).

Table 3.20: Results of Random Forest analysis comparing the full dataset, reduced *n*=15 sample per species dataset, reduced *n*=7 samples per species dataset and the reduced *n*=7 samples per species (excluding *Alnus incana* subsp. *tenuifolia*) dataset. Gini number for each trait is given, as is a rank of importance for each trait (ranks 1-3 are bolded). Out of Bag (OOB) error rates are provided for each model.

	All Samples (OOB Error Rate = 44.1%)		***n*=15 (30 grains/sample)** (OOB Error Rate = 43.4%)		***n*=7 (30 grains/sample)** (OOB Error Rate = 50.8%)		***n*=7 (*A. incana* removed)** (OOB Error Rate = 13.9%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	222.23	**1**	94.87	**2**	40.98	**3**	21.53	**3**
Diameter	213.22	**2**	110.24	**1**	42.35	**2**	34.78	**1**
Annulus Area	199.40	**3**	83.32	**3**	37.26	5	21.07	4
Exine Thickness	171.15	4	80.44	4	47.08	**1**	16.49	5
Arci Width	171.36	5	77.86	5	37.33	4	25.58	**2**
Annulus Height	157.28	6	70.65	6	36.63	6	9.07	7
Arci Strength	103.96	7	57.64	7	26.19	7	10.95	6
Pore Protrusion	49.88	8	22.50	8	10.27	8	7.10	8

b. Reductions in the Number of Pollen Grains Measured per Sample

When sample size per species is not reduced but the number of pollen grains measured per sample is diminished to 20, the resulting CART model (Fig. 3.15) separates pollen based on annulus width, arci strength, exine thickness and diameter, with 43.2% model classification error (Table 3.21). The *A. viridis* subsp. *sinuata, A. incana* subsp. *tenuifolia* and *A. rubra* pollen grains included in the model were classified with 84.1%, 5.6% and 72.1% accuracy, respectively. When the 'OG' test set, composed of pollen grains in each sample removed during dataset reductions, is used to test the model, *A. viridis* subsp. *sinuata, A. incana* subsp. *tenuifolia* and *A. rubra* are classified with somewhat lower accuracy i.e., 78.3%, 3.7% and 65.2%, respectively.

The CART model resulting from the 10 grain per sample dataset separates grains solely on the basis of annulus width (Fig. 3.16), with 46.4% overall model error. This results in a tree with only two terminal nodes, one that groups pollen as *A. rubra*, the other as *A. viridis* subsp. *sinuata.* Therefore, the 10 grain per sample decision tree results in 0% classification accuracy for *A. incana* subsp. *tenuifolia* pollen (Table 3.22). Again, OG test set classification closely resembles the accuracy for the model set.

There is very little change in the rank importance of morphological traits as the number of pollen grains per sample is reduced (Table 3.23). Annulus height, grain diameter and annulus area are ranked 1-3 in importance across all three models. This is in contrast to dataset reductions by sample size where trait importance shifted across different models. In addition, overall model error remains more or less constant as the number of pollen grains per sample is reduced (Table 3.23). However, in all three models, over 95% of *A. incana* subsp. *tenuifolia* grains are misclassified within the model and across test sets.

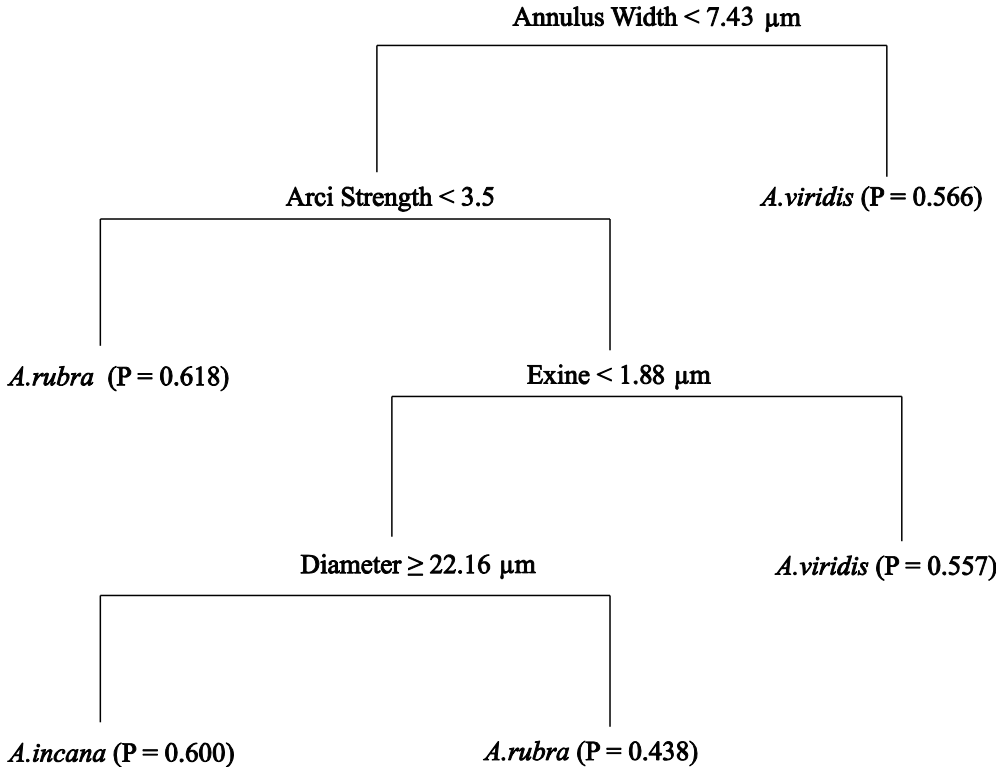


Figure 3.15: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when the number of pollen grains is reduced to 20 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.21: CART model species classification for the 20 grains per sample (all samples included) reduced dataset. Classification of reserved 30% test set and Other Grain (OG) test set is also shown. Total model error is 43.2%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=500)		***A. incana* subsp. *tenuifolia*** (*n*=378)		***A. rubra*** (*n*=434)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	412	84.1	190	50.3	119	27.4
	Test Set	164	78.1	86	53.1	45	24.2
	OG Test	274	78.3	137	50.9	104	33.5
Alnus incana	Model	12	2.4	21	5.6	2	1.0
	Test Set	3	1.4	6	3.7	2	1.0
	OG Test	10	2.9	10	3.7	4	1.3
Alnus rubra	Model	76	15.5	167	44.2	313	72.1
	Test Set	43	20.5	70	43.2	139	74.7
	OG Test	66	18.8	122	45.4	202	65.2

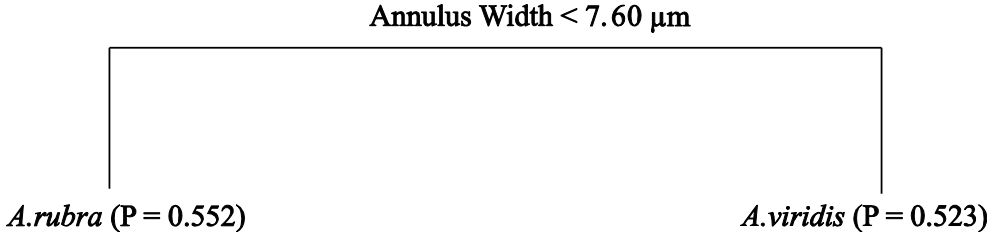


Figure 3.16: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when the number of grains is reduced to 10 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.22: CART model species classification for the 10 grains per sample (all samples included) reduced dataset. Classification of reserved 30% test set and Other grain (OG) test set is also shown. Total model error is 46.4%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=245)		***A. incana* subsp. *tenuifolia*** (*n*=189)		***A. rubra*** (*n*=217)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	191	78.0	115	60.8	59	27.2
	Test Set	94	89.5	42	51.9	25	26.9
	OG Test	551	78.7	286	53.1	227	36.6
Alnus incana	Model	0	0.0	0	0.0	0	0.0
	Test Set	0	0.0	0	0.0	0	0.0
	OG Test	0	0.0	0	0.0	0	0.0
Alnus rubra	Model	54	22.0	74	39.2	158	72.8
	Test Set	11	10.5	39	48.1	68	73.1
	OG Test	149	21.3	253	46.9	393	63.4

Table 3.23: Results of Random Forest analysis for full dataset, reduced 20 grains per sample and 10 grains per sample datasets (all samples included in reductions). Gini number for each trait is given, as is a rank of importance for each trait (ranks 1-3 are bolded). Out of Bag (OOB) error rates are provided for each model.

	All Samples (OOB Error Rate = 44.1%)		***n*=30 (20 grain/species)** (OOB Error Rate = 46.5%)		***n*=30 (10 grain/species)** (OOB Error Rate = 44.2%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	222.23	**1**	148.08	**1**	76.44	**1**
Diameter	213.22	**2**	139.48	**2**	72.33	**2**
Annulus Area	199.40	**3**	131.17	**3**	65.90	**3**
Exine Thickness	171.15	4	122.39	4	53.26	5
Arci Width	171.36	5	110.81	5	54.41	4
Annulus Height	157.28	6	102.03	6	53.18	6
Arci Strength	103.96	7	70.75	7	38.02	7
Pore Protrusion	49.88	8	35.01	8	16.51	8

c. Combined Reductions in Sample Size and Number of Grains per Sample

When the dataset is reduced in sample size and in the number of grains measured per sample to 15 samples per species and 20 grains per sample, CART modelling uses annulus width, arci strength and annulus height to differentiate alder grains to species (Fig. 3.17), with 45.8% model error. Model classification accuracy is 81.4%, 47.1% and 37.1% for *A. viridis* subsp. *sinuata*, *A. incana* subsp. *tenuifolia* and *A. rubra* grains, respectively (Table 3.24). When the OS test set is used to assess model classification, accuracy drops to 77.2%, 25.2% and 38.1% for each species.

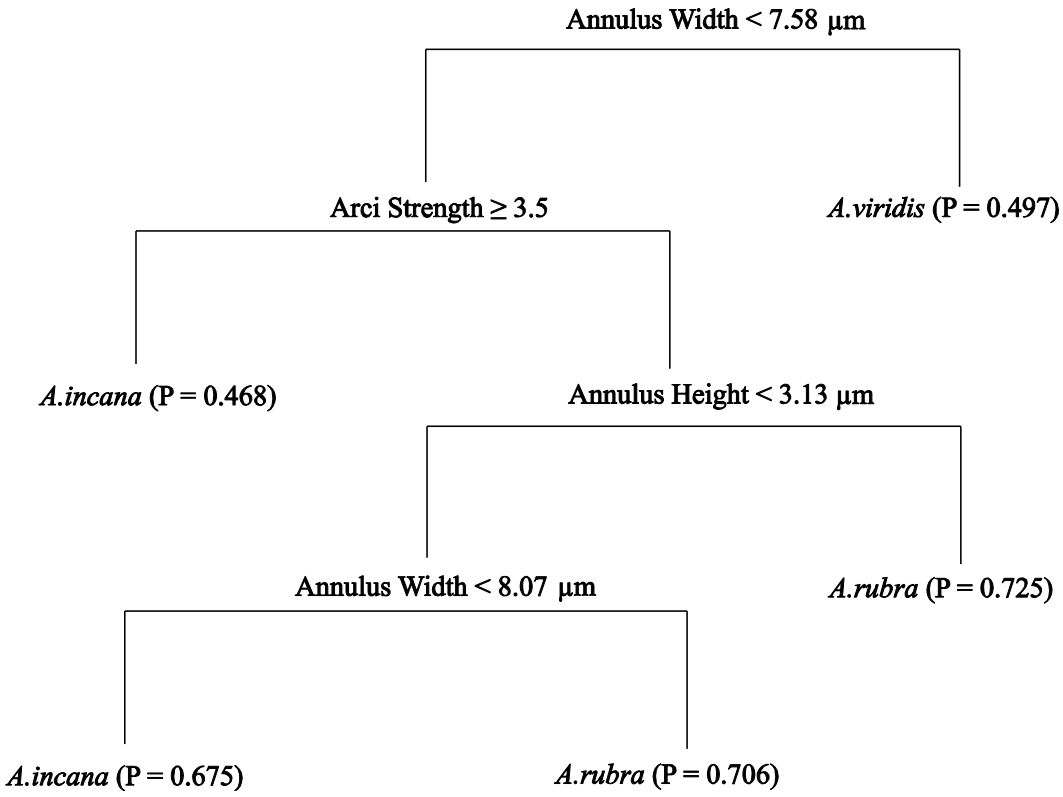


Figure 3.17: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when sample size is reduced to *n*=15 per species and the number of grains is reduced to 20 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.24: CART species classification for the *n*=15 samples per species, 20 grains per sample dataset reduction. Classification of the reserved 30% test set and Other Sample (OS) test set is also shown. Model Error is 45.8%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=210)		***A. incana* subsp. *tenuifolia*** (*n*=210)		***A. rubra*** (*n*=210)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	171	81.4	107	51.0	66	31.4
	Test Set	79	87.8	43	47.8	30	33.3
	OS Test	463	77.2	211	58.6	150	31.3
Alnus incana	Model	26	12.4	78	37.1	45	21.4
	Test Set	10	11.1	32	35.6	27	30.0
	OS Test	98	16.3	91	25.2	147	30.6
Alnus rubra	Model	13	6.2	25	11.9	99	47.1
	Test Set	1	1.1	15	16.7	33	36.7
	OS Test	39	6.5	58	16.1	183	38.1

If the dataset is reduced to 15 samples per species but only 10 grains per sample, CART analysis (Fig. 3.18) separates grains via a different suite of morphological traits (i.e., diameter, annulus height and annulus area), but total model error is reduced 31.8% (Table 3.25). *Alnus viridis* subsp. *sinuata*, *A. incana* subsp. *tenuifolia* and *A. rubra* grains used in the model are classified 69.5%, 60.9% and 74.3% correctly; however, this accuracy is dramatically reduced for all three species when the model is tested using the grains in the OS test set.

Dataset reductions to *n*=7 samples per species with 20 pollen grains measured per sample results in similar shifts in diagnostic splitting variables and decreases in test set accuracy (Fig 3.19; Table 3.26). The *n*=7 samples per species, 10 grains per sample CART model differentiates all three alder pollen types into species groups based only on diameter (Fig. 3.20). While overall model error is 42.9%, it is interesting that 73.5% of the *A. incana* subsp. *tenuifolia* grains are classified correctly (Table 3.27). Model accuracy for differentiating *A. viridis* subsp. *sinuata* (42.9%) and *A. rubra* (55.1%) grains

is lower than for *A. incana* subsp. *tenuifolia*, a result that makes this CART model stand out from other reductions. Correct classification of *A. viridis* subsp. *sinuata* and *A. rubra* grains further drops to 25.8% and 34.5%, respectively, when OS grains are used to test the CART model.

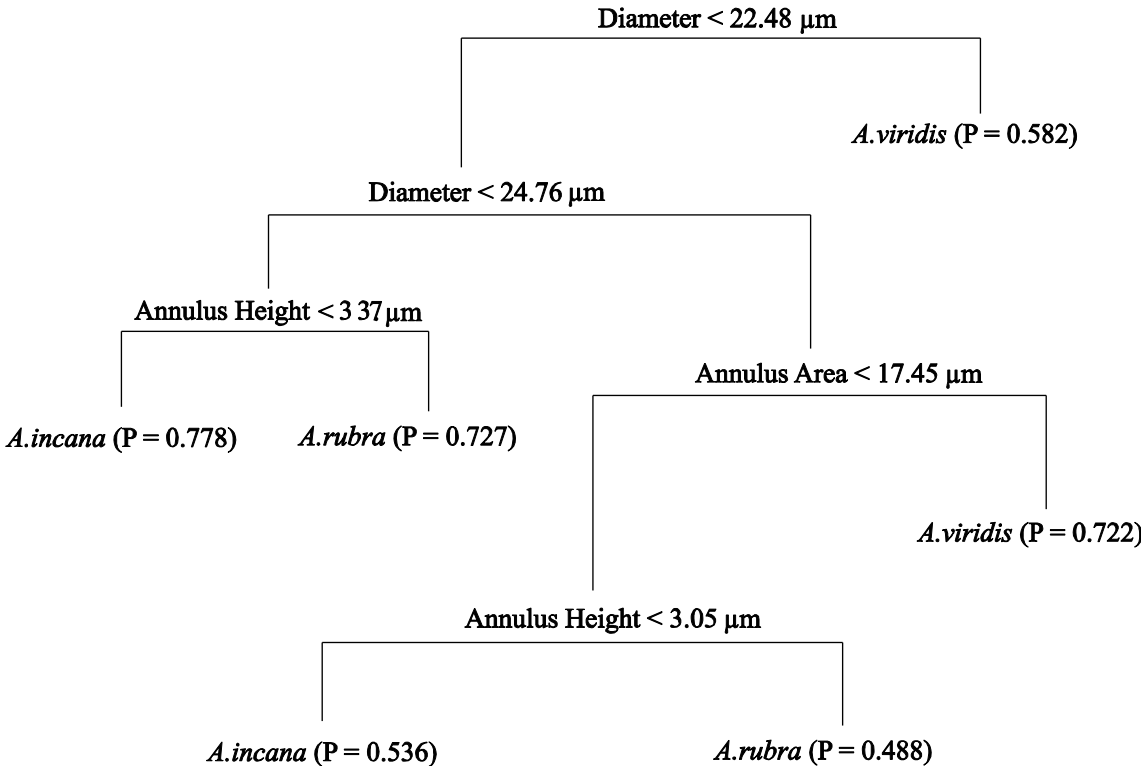


Figure 3.18: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when sample size is reduced to *n*=15 per species and the number of grains is reduced to 10 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.25: CART species classification for the *n*=15 samples per species, 10 grains per sample dataset reduction. Classification of the reserved 30% test set and Other Sample (OS) test set is also shown. Total model error is 31.8%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=105)		***A. incana* subsp. *tenuifolia*** (*n*=105)		***A. rubra*** (*n*=105)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	73	69.5	16	15.2	5	4.8
	Test Set	21	46.7	14	31.1	3	6.7
	OS Test	271	45.2	139	38.7	103	21.4
Alnus incana	Model	24	22.9	64	60.9	22	20.9
	Test Set	15	33.3	15	33.3	11	24.4
	OS Test	210	35.0	111	30.9	163	33.9
Alnus rubra	Model	8	7.6	25	23.8	78	74.3
	Test Set	9	20.0	16	35.6	31	68.9
	OS Test	119	19.8	109	30.3	214	44.6

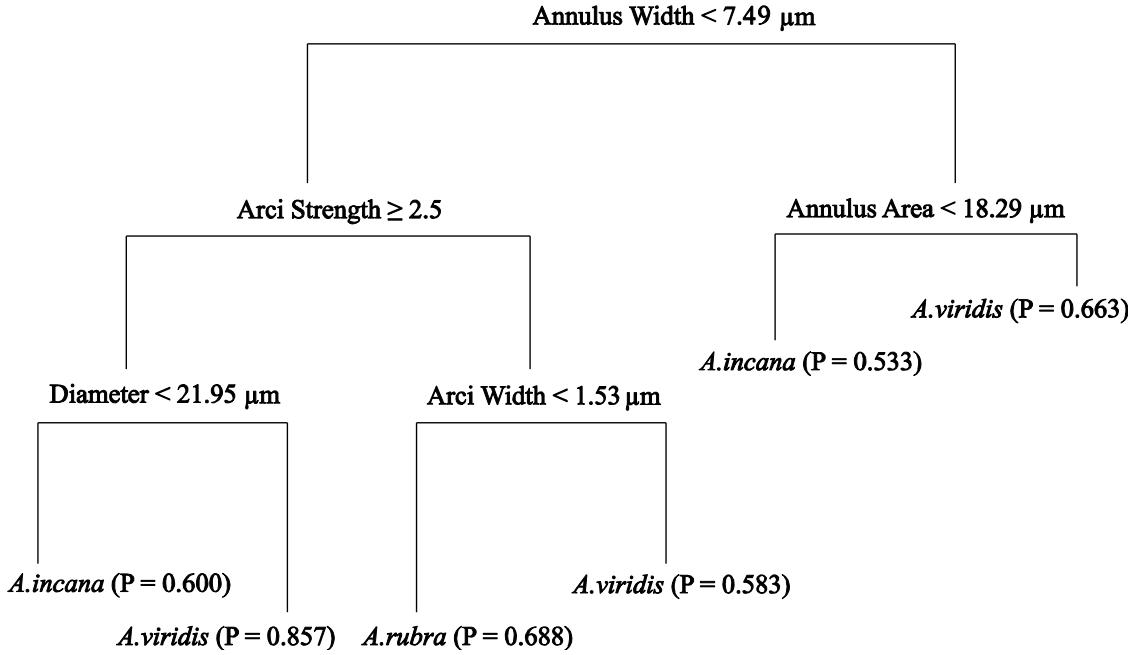


Figure 3.19: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when sample size is reduced to *n*=7 per species and the number of grains is reduced to 20 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.26: CART species classification for the *n*=7 samples per species 20 grains per sample dataset reduction. Classification of the reserved 30% test set and Other Sample (OS) test set is also shown. Total model error is 35.8%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=98)		***A. incana* subsp. *tenuifolia*** (*n*=98)		***A. rubra*** (*n*=98)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	68	69.3	28	28.6	6	6.1
	Test Set	23	54.8	20	47.6	3	7.1
	OS Test	108	12.9	155	25.8	110	15.3
Alnus incana	Model	21	21.4	44	44.9	15	15.3
	Test Set	15	35.7	14	33.3	13	30.9
	OS Test	356	42.3	206	34.4	178	24.7
Alnus rubra	Model	9	9.2	26	26.5	77	78.6
	Test Set	4	9.5	8	19.0	26	61.9
	OS Test	108	12.9	238	39.7	432	60.0

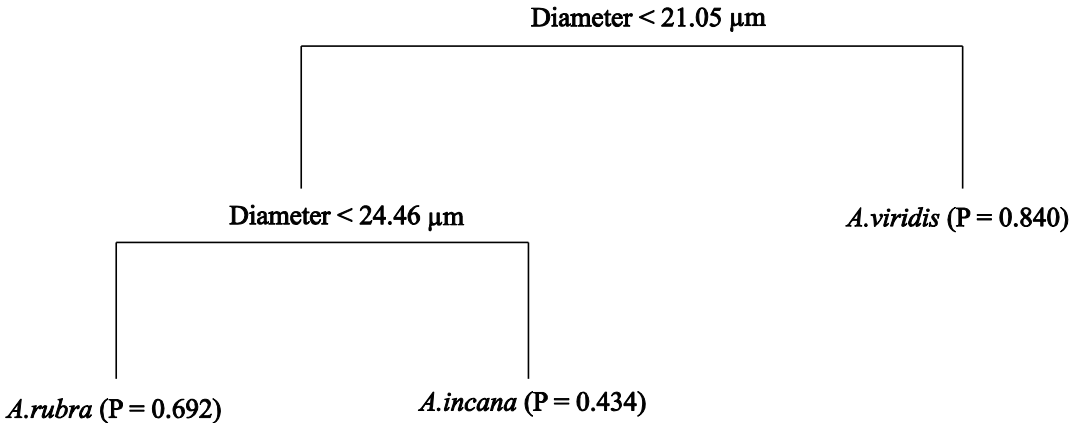


Figure 3.20: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, *Alnus incana* subsp. *tenuifolia* and *Alnus rubra* pollen, when sample size is reduced to *n*=7 and the number of grains is reduced to 10 per sample. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.27: CART species classification for the *n*=7 samples per species 10 grains per sample dataset reduction. Classification of the reserved 30% test set and Other Sample (OS) test set is also shown. Total model error is 42.9%.

		Alder Species					
		A. viridis* subsp. *sinuata (*n*=49)		***A. incana* subsp. *tenuifolia*** (*n*=49)		***A. rubra*** (*n*=49)	
Identified As	**Data**	***n***	**%**	***n***	**%**	***n***	**%**
Alnus viridis	Model	21	42.9	3	6.1	1	2.0
	Test Set	5	23.8	2	9.5	0	0.0
	OS Test	217	25.8	119	19.8	27	3.8
Alnus incana	Model	26	53.1	36	73.5	21	42.9
	Test Set	14	66.7	14	66.6	13	61.9
	OS Test	537	63.9	353	58.9	444	61.6
Alnus rubra	Model	2	4.1	10	20.4	27	55.1
	Test Set	2	9.5	5	23.8	8	38.1
	OS Test	86	10.2	127	21.2	249	34.5

When the dataset is reduced to *n*=7 samples per species and 10 grains per sample, but *A. incana* subsp. *tenuifolia* is removed, the derived CART model (Fig. 3.21) has very low model error (12.2%) when compared to models that include *A. incana* subsp. *tenuifolia,* and when compared to the *A. viridis* subsp. *sinuata* and *A. rubra* reduced species classification model including of all samples and grains (Fig. 3.10). The model differentiates *A. viridis* subsp. *sinuata* and *A. rubra* pollen grains based on diameter, arci strength and exine thickness. Annulus width, which is the most important trait for classification of *A. viridis* subsp. *sinuata* and *A. rubra* when all samples are included, does not appear in the reduced model. Within model classification accuracy for each species is very high with 93.9% and 81.6% of *A. viridis* subsp. *sinuata* and *A. rubra* pollen identified correctly (Table 3.28). However, only 67.7% and 59.1% of *A. viridis* subsp. *sinuata* and *A. rubra* OS test set grains are identified correctly (Table 3.28).

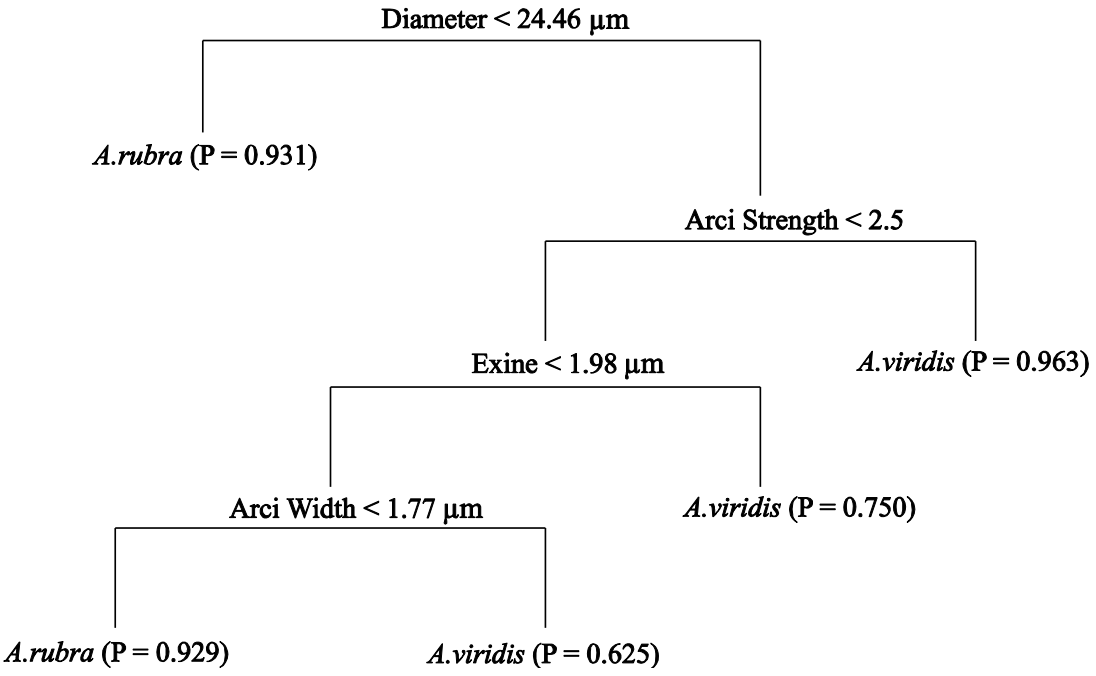


Figure 3.21: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata*, and *Alnus rubra* pollen (*Alnus incana* subsp. *tenuifolia* excluded), when the number of samples per species is reduced to *n*=7 (10 grains measured per sample). All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node. Terminal nodes indicate species classification and the within model probability of correct classification.

Table 3.28: CART model species classification for *n*=7 samples per species and 10 grain per sample, *A. viridis* subsp. *sinuata* and *A. rubra* (*Alnus incana* subsp. *tenuifolia* removed) dataset. Model classification of OS test set is also shown. Total model error is 12.2%.

		Alder Species			
		A. viridis subsp. sinuata (*n*= 49)		***A. rubra*** (*n*= 49)	
Identified As	**Data**	***n***	**%**	***n***	**%**
A. viridis	Model	46	93.9	9	18.4
	OS Test	569	67.7	294	40.9
A. rubra	Model	3	7.1	40	81.6
	OS Test	271	32.3	426	59.1

Random Forest analyses of the datasets that have been reduced both in sample size and in the number of grains per sample indicate high (i.e., 43.9-61.9%) OOB error rates for models based on these combined reductions (Table 3.29). Moreover, the morphological traits ranked 1-3 in importance for classifying pollen grains to species differ across all models; however, diameter and annulus width remain consistently important. When models excluding *A. incana* subsp. *tenuifolia* are compared across reductions using Random Forest analyses (Table 3.30), there is decreased error with dataset reduction. Trait importance again shifts between models; however, annulus width which is ranked most important for differentiating *A. viridis* subsp. *sinuata* and *A. rubra* when all samples are included, is not one of the top three most important traits for the *n*=7 (10 grains per sample) reduction (Table 3.30).

Together, nested ANOVA, CART modelling and Random Forest analysis indicate that reductions in sample size combined with reductions in the number of grains measured per sample result in shifts in morphological trait importance for pollen classification. Further, as datasets are increasingly reduced, rates of misclassification of pollen grains not used in model creation also increase. Interestingy, as datasets including all species are increasingly reduced, model error increases; whereas, the opposite is true when datasets excluding *A. incana* subsp. *tenuifolia* are increasingly reduced.

Sensitivity Analysis: Splitting the Data into Regional Subsets

Before the datasets were split into regional subsets for *A. viridis* subsp. *sinuata* and *A. rubra*, correlation was used to test for associations between latitude and each individual morphology trait for each of the three alder species as well as for all species combined. The vast majority of these correlations are not significant and those that

Table 3.29: Results of Random Forest analysis for the *n*=15, 20 grains per sample and 10 grains per sample datasets and *n*=7, 20 grains per sample and 10 grains per sample datasets. Gini number for each trait is given, as is a rank of importance for each trait (ranks 1-3 are bolded). Out of Bag (OOB) error rates are provided for each model.

	***n*=15 (20 grains/species)** (OOB Error Rate = 49.8%)		***n*=15 (10 grains/species)** (OOB Error Rate = 47.0%)		***n*=7 (20 grains/species)** (OOB Error Rate = 43.9%)		***n*=7 (10 grains/species)** (OOB Error Rate = 61.9%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	72.17	**1**	28.00	**3**	32.00	**2**	14.78	**2**
Diameter	65.63	**2**	42.76	**1**	35.27	**1**	18.60	**1**
Annulus Area	64.68	**3**	29.37	**2**	28.00	4	14.00	**3**
Exine Thickness	62.10	4	27.19	4	24.87	5	12.27	4
Arci Width	54.13	5	25.55	5	29.59	**3**	12.27	5
Annulus Height	53.73	6	25.50	6	23.31	6	11.98	6
Arci Strength	31.11	7	22.72	7	15.77	7	8.65	7
Pore Protrusion	14.96	8	7.82	8	6.16	8	4.63	8

Table 3.30: Results of Random Forest analysis for classification of *Alnus viridis* subsp. *sinuata* and *Alnus rubra*. All datasets (all samples, *n*=7 samples per species, and *n*=7 samples per species 10 grains per sample) exclude *Alnus incana* subsp. *tenuifolia*. Gini number for each trait is given, as is a rank of importance for each trait (ranks 1-3 are bolded). Out of Bag (OOB) error rates are provided for each model.

A. rubra and *A. viridis* only:	**All Samples** (OOB Error Rate = 21.3%)		***n*=7 (all grains per sample)** (OOB Error Rate = 13.9%)		***n*=7 (10 grains per sample)** (OOB Error Rate = 18.4%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	139.73	**1**	21.53	**3**	5.66	5
Diameter	122.76	**2**	34.78	**1**	10.60	**1**
Annulus Area	99.62	**3**	21.07	4	5.72	6
Exine Thickness	80.04	4	16.49	5	8.26	**2**
Arci Width	84.26	6	25.58	**2**	6.28	4
Annulus Height	60.53	7	9.07	7	4.04	7
Arci Strength	82.13	5	10.95	6	6.81	**3**
Pore Protrusion	19.28	8	7.10	8	0.99	8

are significant are generally weakly correlated (Appendix D). The sole exception to this was a significant negative correlation (r = -0.654; $p < 0.001$) between latitude and arci width in *A. viridis* subsp. *sinuata*.

Nested ANOVA Models Comparing Regional Subsets

In general, *A. rubra* pollen from the mainland coast is smaller than *A. rubra* pollen from Vancouver Island, across all morphological traits (Fig. 3.22). However, the extent of morphological overlap varies substantially for each trait; scaled U comparisons between regions range from 63.1% to 98.5% morphological overlap in *A. rubra* pollen (Table 3.31). There are significant differences in exine thickness and diameter in *A. rubra* pollen from Vancouver Island and the mainland coast (Table 3.31). There are also significant differences between individual plants within each region.

The *A. viridis* subsp. *sinuata* dataset was split into three regions (Coast, Inland and North). There is extensive morphological overlap in the pollen of this species across the three different regional subsets (Fig. 3.23), with a general trend of decreasing size in all measured traits between 'Coast' and 'North' samples, with 'Inland' samples intermediate between the two. *Alnus viridis* subsp. *sinuata* pollen is significantly different between regions in annulus width, annulus area and diameter, as indicated by nested ANOVA models (Table 3.32). Coastal and Inland samples differ less in their pollen morphology than samples from the North. This is supported by scaled U statistics of trait distribution overlap between regional datasets. However, as was the case with comparisons between *A. rubra* regional groups, there are also significant differences between individual plants within each regional group across all of the regional comparisons. Full nested ANOVA models comparing regional subsets for all traits show that between region differences account for most of the variance in each model (Table 3.33).

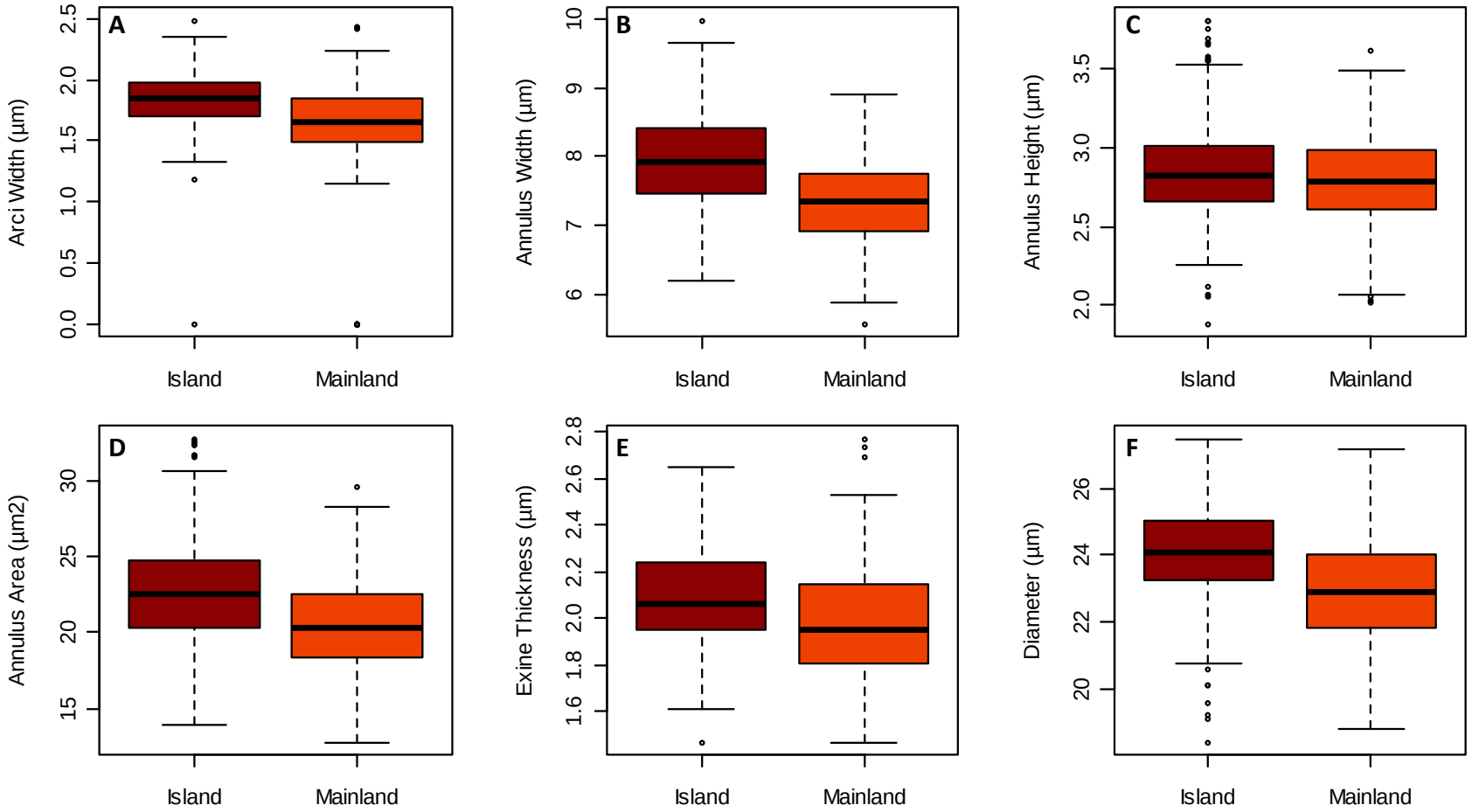


Figure 3.22: Boxplots showing regional variability in *Alnus rubra* pollen for each quantitative morphological trait: arci width (A), annulus width (B), annulus height (C), annulus area (D), exine thickness (E) and grain diameter (F). Solid lines bisecting each boxplot represent the trait median for each region. Box edges mark the first and third quartiles. Whiskers extend to the smallest and largest non-extreme data points.

Table 3.31: Summary of nested ANOVA analysis for pair-wise regional comparisons between the 'Vancouver Island' and 'Mainland Coast' *Alnus rubra* datasets. All quantitative morphological trait comparisons are shown. Scaled U statistic is the percent overlap in each trait between regional datasets. Bolded numbers highlight significant P-values of less than 0.05.

	Nested Comparison (Region)		**Nested Comparison (Region:Sample)**		**Nested Comparison (Region:Sample:Grain)**		**U Statistic: $(U)(2/n_1n_2)$**
Trait	**F**	**P-value**	**F**	**P-value**	**F**	**P-value**	**Overlap (%)**
Arci Width (µm)	0.02	0.882	7.48	**< 0.001**	1.12	0.334	94.2
Annulus Width (µm)	0.07	0.791	13.57	**< 0.001**	1.06	0.390	98.5
Annulus Height (µm)	0.17	0.677	7.79	**< 0.001**	1.21	0.262	95.6
Annulus Area (μm^2)	0.31	0.579	13.29	**< 0.001**	1.19	0.276	97.5
Exine Thickness (µm)	5.75	**0.017**	5.77	**< 0.001**	1.12	0.338	63.1
Grain Diameter (µm)	7.88	**0.005**	12.68	**< 0.001**	0.74	0.733	81.2

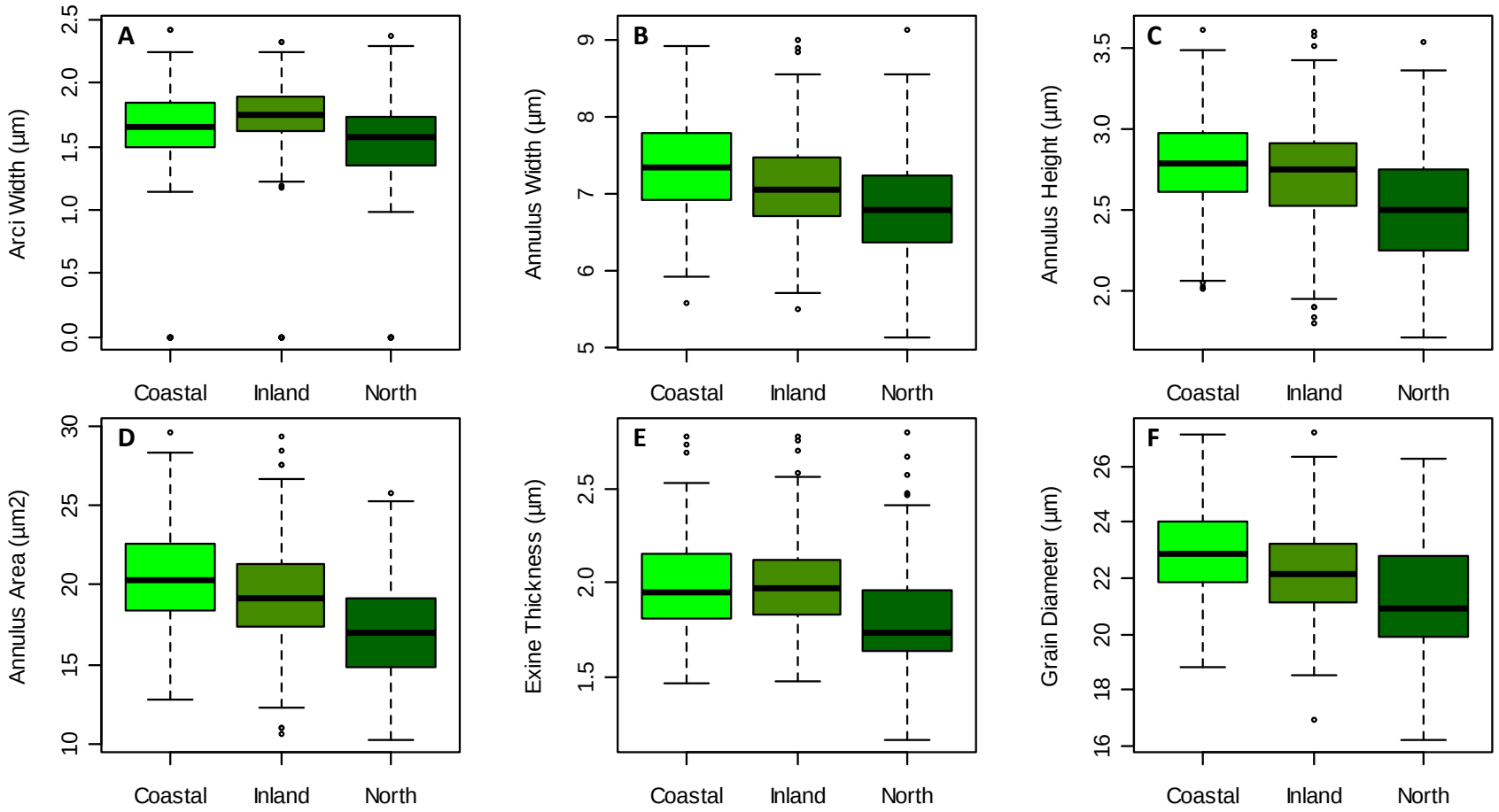


Figure 3.23: Boxplots showing regional variability in *Alnus viridis* subsp. *sinuata* pollen for each quantitative morphological trait: arci width (A), annulus width (B), annulus height (C), annulus area (D), exine thickness (E) and grain diameter (F). Solid lines bisecting each boxplot represent the trait median for each region. Box edges mark the first and third quartiles. Whiskers extend to the smallest and largest non-extreme data points.

Table 3.32: Summary of nested ANOVA analysis for pair-wise regional comparisons between the'Inland (I),' 'Coastal (C)' and 'North (N)' *A. viridis* subsp. *sinuata* datasets. All quantitative morphological trait comparisons are shown. Scaled U statistic is the percent overlap in traits between regional datasets. Bolded numbers highlight significant P-values of less than 0.05.

Trait by Region		Nested Comparison (Region)		Nested Comparison (Region:Sample)		Nested Comparison (Region:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Coast	**I-C**	10.39	**0.023**	3.05	**0.004**	1.35	1.000	79.4
Inland	**I-N**	2.27	1.000	4.93	**< 0.001**	1.12	1.000	79.7
North	**C-N**	29.09	**< 0.001**	3.98	**< 0.001**	0.92	1.000	59.3
Annulus Width								
Coast	**C-I**	16.92	**0.001**	2.51	**0.043**	1.58	1.000	77.7
Inland	**C-N**	85.27	**< 0.001**	4.33	**< 0.001**	2.02	0.207	55.1
North	**I-N**	26.98	**< 0.001**	4.97	**< 0.001**	1.82	0.422	74.9
Annulus Height								
Coast	**C-I**	5.73	0.291	4.02	**< 0.001**	1.49	1.000	87.7
Inland	**C-N**	92.88	**< 0.001**	4.44	**< 0.001**	1.52	1.000	53.6
North	**I-N**	57.12	**< 0.001**	5.99	**< 0.001**	1.89	0.310	64.2
Annulus Area								
Coast	**C-I**	15.40	**0.002**	3.23	**0.002**	1.62	1.000	79.8
Inland	**C-N**	140.26	**< 0.001**	4.66	**< 0.001**	1.97	0.258	45.4
North	**I-N**	67.11	**< 0.001**	6.37	**< 0.001**	2.31	**0.045**	61.6
Exine Thickness								
Coast	**C-I**	0.02	1.000	3.62	**< 0.001**	3.33	0.051	98.1
Inland	**C-N**	71.23	**< 0.001**	6.99	**< 0.001**	1.40	1.000	58.0
North	**I-N**	80.54	**< 0.001**	6.67	**< 0.001**	1.18	1.000	56.4
Grain Diameter								
Coast	**C-I**	25.82	**< 0.001**	9.40	**< 0.001**	0.60	1.000	74.9
Inland	**C-N**	154.85	**< 0.001**	22.52	**< 0.001**	1.21	1.000	51.8
North	**I-N**	66.82	**< 0.001**	27.50	**< 0.001**	1.27	1.000	68.6

Table 3.33: Summary of nested ANOVA analysis and variance component analysis for all quantitative traits by *Alnus viridis* subsp. *sinuata* region.

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.***	**% Variance**
Between Regions	2	5.06	2.53	0.2213	51.1
Between Individuals	20	15.36	0.76	0.0190	4.4
Between Grains	667	128.70	0.19	0.1929	44.5
Annulus Width (µm)					
Between Regions	2	30.96	15.48	1.7550	81.0
Between Individuals	20	28.88	1.44	0.0353	1.6
Between Grains	667	250.50	0.38	0.3756	17.3
Annulus Height (µm)					
Between Regions	2	9.75	4.87	0.5538	84.1
Between Individuals	20	8.88	0.44	0.0117	1.8
Between Grains	667	62.02	0.09	0.0930	14.1
Annulus Area (μm^2)					
Between Regions	2	1304.60	652.28	76.2563	88.2
Between Individuals	20	844.65	42.23	1.1037	1.3
Between Grains	667	6083.30	9.12	9.1203	10.5
Exine Thickness (µm)					
Between Regions	2	5.67	2.83	0.3138	83.1
Between Individuals	20	6.32	0.32	0.0087	2.3
Between Grains	667	36.86	0.06	0.0553	14.6
Grain Diameter (µm)					
Between Regions	2	330.35	165.18	15.7338	83.0
Between Individuals	20	786.17	39.31	1.2447	6.6
Between Grains	667	1315.70	1.97	1.9725	10.4

* Var.Comp. = the variance component (i.e., variance explained by each nested factor)

Multi-trait (CART and Random Forest) Analysis of Regional Subsets

The CART model derived from the 'Vancouver Island' *A. rubra* and 'Coast' *A. viridis* subsp. *sinuata* regional datasets (Fig. 3.24) differentiates the pollen of these two species based on arci width, diameter, annulus width, annulus area and exine thickness. End node probabilities for correct classification are relatively high, ranging from (P=0.611-1.00), and total model error is relatively low at 13.2%. The model classifies *A. viridis* subsp. *sinuata* and *A. rubra* pollen used to build the model with a high degree of accuracy (Table 3.34). However, when the model is used to classify *A. rubra* grains from the mainland coast test set, only 10.4% of grains are identified correctly.

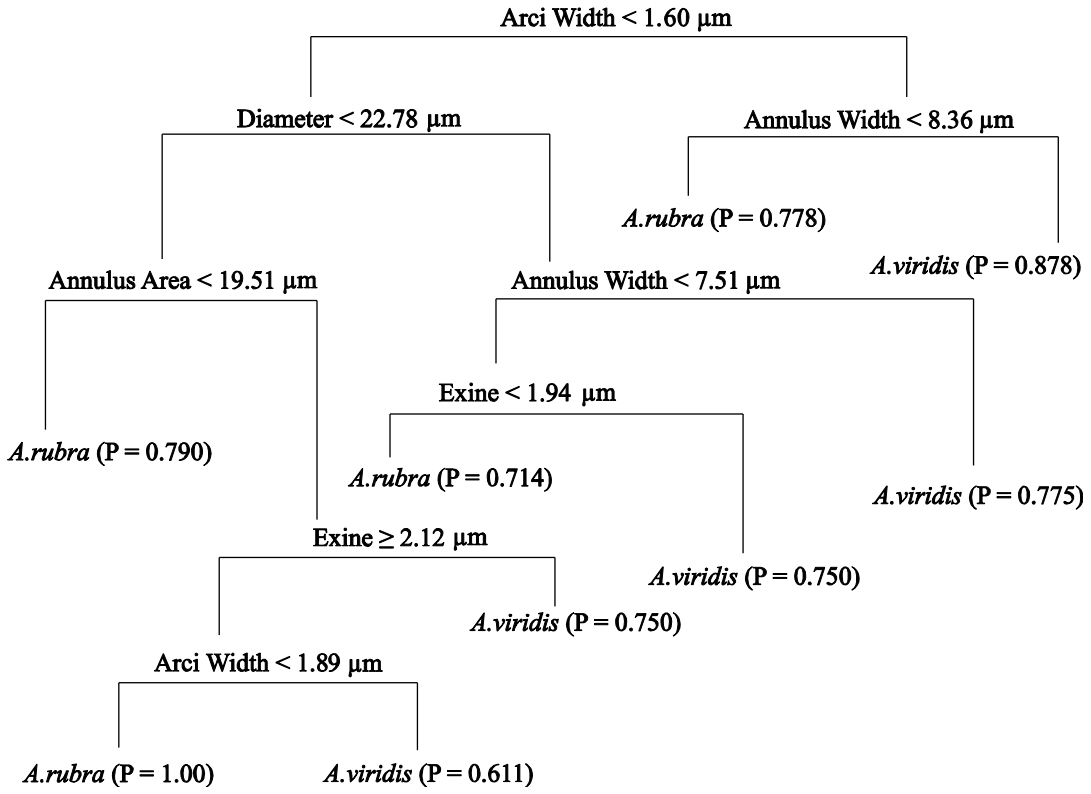


Figure 3.24: CART derived decision tree for the simultaneous classification of *Alnus viridis* subsp. *sinuata* and *Alnus rubra* pollen, based on the 'Coastal' and 'Vancouver Island' *Alnus viridis* subsp. *sinuata* and *Alnus rubra* regional datasets, respectively. All quantitative traits, arci strength and pore protrusion are included as model inputs. Morphological splitting variables and threshold values occur at each internal node.

Table 3.34: CART species classification for 'Coastal' and 'Vancouver Island' *Alnus viridis* subsp. *sinuata* and *Alnus rubra* datasets, respectively. Model test set classification of grains from other regions are also shown where AV = *Alnus viridis* subsp. *sinuata* and AR = *Alnus rubra*. Total model error is 13.2%.

		Alder Species			
		A. viridis* subsp. *sinuata (*n*=210)		***A. rubra*** (*n*=210)	
Identified As	**Data**	***n***	**%**	***n***	**%**
Alnus viridis	Model	185	88.1	39	18.6
	AV Inland	201	83.8	NA	NA
	AV North	216	90.0	NA	NA
	AR Mainland	NA	NA	188	89.5
Alnus rubra	Model	25	11.9	171	81.4
	AV Inland	39	16.3	NA	NA
	AV North	24	10.0	NA	NA
	AR Mainland	NA	NA	22	10.5

As occurred in Random Forest models comparing datasets reduced in size, splitting the dataset into regions also changes the morphological traits that are important for classification of alder pollen to species (Table 3.35). When compared to the original two-species dataset (all samples and all grains per sample, but with *A. incana* subsp. *tenuifolia* data removed), the traits ranked 1-3 in importance shift from annulus width, diameter and annulus area to arci strength, diameter and annulus width. As was the case with sample size reductions and combined sample size and grains measured per sample reductions, splitting the dataset into regional subsets has important implications for identifying the morphological traits that are diagnostic for species identification and for classification accuracy across regions. These results underscore the importance of determining diagnostic pollen identification criteria from pollen samples from across a species' range, rather than from within a limited portion of a species' range.

Table 3.35: Results of Random Forest analysis comparing the reduced species dataset (excluding *Alnus incana* subsp. *tenuifolia*) and the 'Coastal' and 'Vancouver Island' *Alnus viridis* subsp. *sinuata* and *Alnus rubra* regional dataset. Gini number for each trait is given, as is a rank of importance for each trait (ranks 1-3 are bolded). Out of Bag (OOB) error rates are provided for each.

	All Samples (*A. rubra* and *A. viridis*) (OOB Error Rate = 21.3%)		**Regional Model (*A. rubra* and *A. viridis*)** (OOB Error Rate = 19.3%)	
Trait	**Mean Gini**	**Rank**	**Mean Gini**	**Rank**
Annulus Width	139.73	**1**	32.64	**3**
Diameter	122.76	**2**	33.02	**2**
Annulus Area	99.62	**3**	24.00	5
Exine Thickness	80.04	4	23.36	6
Arci Width	84.26	6	31.62	4
Annulus Height	60.53	7	18.17	7
Arci Strength	82.13	5	40.97	**1**
Pore Protrusion	19.28	8	5.38	8

Discussion

Differentiating Alder Pollen

What is most clear from the results of the morphometric analyses performed in this study is that no single morphological trait on its own is sufficient for identifying alder pollen to species. Comparison of morphological trait means and variance component analyses suggest statistically significant differences between all species, but this can be deceptive. Substantial overlap in morphological variability between species indicates that the significant differences between means (Table 3.3) as well as the high percentage of variance attributable to between species effects across traits (Table 3.4) have only limited value for determining which traits are important for species level identification. Instead, the pollen morphologies of three alder species form a morphological continuum, where *A. viridis* subsp. *sinuata* pollen is the smallest across quantitative traits, *A. rubra* is the largest, and *A. incana* subsp. *tenuifolia* is intermediate (Fig 3.4). Moreover, there is little distinction between the three alder species in any of the qualitative traits. A comparison of mean values and/or reliance on limited morphological traits is insufficient when attempting to differentiate these three alder species. As hybridization is known to have played a large role in the evolution of *Alnus* species (Furlow 1979; Chen and Li 2004), ancestral hybridization between clades may play a role in pollen grain similarities between alder species.

CART analysis produces a multi-trait method for species classification by defining important morphological splitting variables and threshold values for these traits. In doing so, CART analysis provides a better tool for classification of alder pollen than is

possible by examining individual morphological traits in isolation. While decision trees derived by CART modeling may seem convoluted in that they can have multiple end nodes per species class, in practice, multiple nodes result in differing pathways to a single classification. This allows the model to encompass morphological variability within one species in a way that is not possible when comparing single trait thresholds.

CART analysis provides a possible solution for alder pollen separation for two of the three alder species that occur naturally in western North America. The classification error for *A. incana* subsp. *tenuifolia* was over 90% in the full CART model produced for the classification of all three alder pollen types (Table 3.5) and 100% in the model that excluded categorical morphological traits (Table 3.6). As *A. incana* subsp. *tenuifolia* pollen morphology is intermediate for all traits when compared to *A. viridis* subsp. *sinuata* and *A. rubra* pollen, morphological overlap is extensive enough to prevent the CART model from defining appropriate trait thresholds by which to classify *A. incana* subsp. *tenuifolia* pollen. The inclusion of *A. incana* subsp. *tenuifolia* pollen in 'species vs. other' binary trees models (Figures 3.8 and 3.9) limited the usefulness of these models for determining splitting traits for the isolation of *A. viridis* subsp. *sinuata* and *A. rubra* pollen. The extreme model error present when attempting to differentiate *A. incana* subsp. *tenuifolia* indicates that the pollen of this species cannot be reliably distinguished from that of *A. viridis* subsp. *sinuata* and *A. rubra.*

The morphology of *A. incana* subsp. *tenuifolia* pollen is an amalgam of the traits that are more characteristic of *A. rubra* and *A. viridis* subsp. *sinuata* pollen. As such, CART modelling (Tables 3.5 and 3.6) classifies the grains of this species as both *A. viridis* subsp. *sinuata* and *A. rubra*. Interestingly, research examining morphological

differences between the pollen of congeneric species in the genus *Betula*, the closest relative genus to *Alnus* (Navarro et al. 2003), found notable differences in morphological traits (i.e., pore depth and pore-diameter ratio) between pollen from tree birches vs. shrub birches (Clegg et al. 2005). Separation into broader growth form groups cannot be made in west coast alder species as 30-52% of pollen from *A. incana* subsp. *tenuifolia* (a shrub) are classified as *A. rubra* (a tree). However, the intermediate pollen morphology of *A. incana* subsp. *tenuifolia* also does not reflect the genetic relatedness of these three species of *Alnus*. The *A. incana* species complex and *A. rubra* both fall within the *Alnus* subgenus in the broader genus *Alnus*, whereas *A. viridis* subsp. *sinuata* is part of the phylogenetically distinct *Alnobetula* subgenus (Chen and Li 2004). These phylogenetic patterns are not reflected in pollen morphology, as 48-60% of *A. incana* subsp. *tenuifolia* are identified as *A. viridis* subsp. *sinuata*. Given that *A. incana* subsp. *tenuifolia* pollen grains are almost evenly split by statistical models between the other two alder species, there is no practical value (either growth form or genetic) in separating all three species into two morphotypes in areas where the three species co-occur.

Alnus incana subsp. *tenuifolia* does not currently occur in areas such as coastal British Columbia. As such, a method for distinguishing alder pollen that excludes this species would be useful in certain areas. Once *A. incana* subsp. *tenuifolia* is removed, CART models provide a far more accurate approximation of species identification for *A. rubra* and *A. viridis* subsp. *sinuata*. However, it must be noted that these models are appropriate for discriminating alder pollen only when it can be safely assumed that *A. incana* subsp. *tenuifolia* does not occur on the landscape, and has not occurred in the pollen source area in the past. Small, light seeds are characteristic of the genus *Alnus*, as

is the ability to disperse rapidly, particularly along waterways, colonizing at speeds of 1000 to 2000 m/yr in continental Europe following the last glacial maximum (Delcourt and Delcourt 1991). These life history characteristics combined with the prevalence of bare mineral soils and substantial changes in climate through the Holocene (Delcourt and Delcourt 1991; Mayle et al. 1993; Wright et al. 1993; Williams et al. 2002; Lacourse 2009) suggest that large shifts in alder species ranges likely occurred along the west coast of North America. It is also possible that alder species that do not currently occur in the Pacific Northwest were actually present in the region in the past. For instance, *Alnus rhombifolia*, currently native to California, may previously have had a more northern distribution (Furlow 1979; Reinink-Smith 2010), in which case its pollen would confound species identification in coastal British Columbia. Furthermore, studies of modern pollen distribution show that alder pollen is blown far outside the boundaries of its current species' ranges (MacDonald and Ritchie 1986), travelling distances in the hundreds of kilometres (Livingstone 1955; Fægri and Iversen 1989; Mayle et al. 1993). For these reasons, the absence of *A. incana* subsp. *tenuifolia* and the discrimination of *A. viridis* subsp. *sinuata* and *A. rubra* pollen must be assumed carefully, especially in areas close to their modern day range limits.

If the assumption of the absence of *A. incana* subsp. *tenuifolia* pollen is met, then classification of *A. viridis* subsp. *sinuata* and *A. rubra* pollen can be achieved via CART modelling. Even when *A. incana* subsp. *tenuifolia* is excluded, model error is too high to conclude definitively that the threshold criteria are effective in distinguishing the pollen of *A. viridis* subsp. *sinuata* and *A. rubra*. However, the error rates do not preclude the models from being used for separating *A. viridis* subsp. *sinuata* and *A. rubra* pollen into

two morphotypes that are analogous to species separation and representative of their shrub vs. tree growth forms.

The quantitative and qualitative CART model isolates *A. viridis* subsp. *sinuata* and *A. rubra* pollen based on annulus width and arci strength (Fig. 3.10), whereas the quantitative trait CART model separates the two species based on annulus width, exine thickness and diameter (Fig. 3.11). Given that the important traits for classification differ between the two models, I propose a method for morphotype separation that modifies the statistical output and uses a combination of the most important categorical and quantitative traits (Table 4.1). Using this method, alder pollen grains are classified as *A. viridis* - type if they have an annulus width < 7.5 µm, arci that are weak to moderate in strength (1-3), grain diameter of < 23 µm, and exine thickness < 1.9 µm. If annulus width is > 7.5 µm, arci are strong (>4), exine thickness is > 1.9 µm and diameter is > 23 µm, then a grain should be classified as *A. rubra* – type. This morphotype classification may not work perfectly for every alder pollen grain encountered in the fossil record, as some grains may conform to opposing group qualifiers. In that instance, I propose grouping an individual grain into the morphotype category for which it conforms to the higher number of traits. If a grain conforms to an even number of traits in both morphotype categories, grain shape may be useful for placing a grain into the *A. rubra* morphotype group, as *A. rubra* pollen have a higher proportion of convex grains than *A. viridis* subsp. *sinuata* pollen which are more variable in shape (Fig. 3.5D).

While all pollen grain measurements performed in this study were made using a software interface at 1000× magnification, most pollen identification is performed at

Table 4.1: Morphological distinctions for *Alnus viridis* - type and *Alnus rubra* - type pollen identification.

	***Alnus* Pollen Morphotype**	
Trait	*A. viridis* - type	*A. rubra* - type
Annulus Width	< 7.5 µm	> 7.5 µm
Grain Diameter	< 23.0 µm	> 23.0 µm
Exine Thickness	< 1.9 µm	> 1.9 µm
Arci Strength	weak/moderate (1-3)	strong (≥4)

lower magnifications (e.g., 400×) with the aid of an ocular micrometer. For this reason, trait threshold values in the proposed alder-type separation method have been modified from the original CART model output to include less decimal precision. As estimations of all qualitative traits were defined and given relative ranks after examining many pollen grains from all three alder species, palynologists should consult alder reference pollen prior to assessing these traits in fossil pollen.

The traits identified as important for the separation of alder pollen grains into a *A. viridis* – type and a *A. rubra* - type (Table 4.1) support the morphological differences pinpointed in visual determinations made previously by palynologists separating alder pollen into 'morphotypes' in the Pacific Northwest (e.g. Lacourse 2005). Analysis also supports a two category system, even if it does rule out a broader tree vs. shrub distinction that includes *A. incana* subsp. *tenuifolia.*

Implications of Alder Pollen Identification for Paleoecological Studies

While a method for species level identification of all three alder species was not achieved, classifying alder pollen into a *A. viridis* – type and *A. rubra* - type is of practical and ecological value because it allows for the distinction between tree alder (*A. rubra* - type) and shrub alder (*A. viridis* – type) pollen in areas where *A. incana* subsp.

tenuifolia is not present such as incoastal British Columbia. This distinction, while primarily associated with growth form, also represents important ecological and functional life history differences that have the potential to enhance insights pertaining to long-term plant community dynamics.

The ability to distinguish the pollen of different species in fossil records is especially important when species have differential responses to environmental conditions and differing species-specific ecologies (Finkelstein et al. 2006), which is the case for *A. viridis* subsp. *sinuata* and *A. rubra* in the Pacific Northwest. While most paleoecological studies from this region have not identified alder pollen past the genus level, studies that separate alder pollen into morphotypes (e.g. Lacourse 2005) show that *A. viridis* and *A. rubra* have different post-glacial histories and ecosystem associations. The differentiation of alders, based on pollen morphotypes, also reveals that individual alder species have strong temporal associations with different conifers. Lacourse (2009) shows clearly that *A. rubra*-type pollen increases in abundance immediately before the arrival of *Picea sitchensis* on northern Vancouver Island following the last glaciation, and that *A. viridis*-type pollen is more closely associated with the arrival and dynamics of *Pinus contorta* and *Tsuga mertensiana*. These patterns suggest that these two alder species played different roles in facilitating the establishment of conifers on northern Vancouver Island. With the increased separation of *A. viridis* –type and *A. rubra* – type pollen, paleoecological studies of facilitation of conifers by these alders on long ecological timescales may be possible (Lacourse 2009).

Increasing the taxonomic resolution of fossil pollen identifications has also provided new paleoecological information for taxa other than *Alnus*. For instance, the

application of CART derived differentiation techniques for *Picea* pollen (Linbladh et al. 2002), when used to examine late-glacial sediments in the eastern United States, permitted a hypothesized shift from *P. glauca* to *P. mariana* in the area to be substantiated (Lindbladh et al. 2007). Finkelstein et al. (2006) note that if not for the differentiation to species of *Acer rubrum* and *Acer saccharum,* the 2000-year trend of decreasing *A. saccharum* and increasing *A. rubrum* populations would not have been discerned. This discovery allowed concern over the possible future disappearance of *A. saccharum* in certain areas to be tempered by the longevity of this trend. Current climate change scenarios mean that within genus taxonomic precision will become increasingly important in the future, as these methods may allow the isolation of exogenous ecosystem drivers such as climate to be isolated from endogenous processes such as succession and competition (Flenley 2003).

Sensitivity Analysis

Using Small Datasets in the Study of Pollen Morphology

Random reductions in sample size (i.e., reductions in the number of samples per species and number of pollen grains per sample) reveal that results of statistical analyses are highly sensitive to changes in sample size. In the three alder species examined here, interspecific differences in pollen morphology are sufficiently strong to remain consistently significant across some sample size reductions. However, this trend is not sustained when the dataset is reduced to n=15 and n=7 samples per species with only 10 grains per sample included i.e., dataset sizes that are typical of other palynological studies focused on pollen identification. For these dataset sizes, the statistically significant difference between species is lost across most morphological traits. Moreover, as datasets

become smaller between grain differences account for a higher percentage of total model variance. Therefore, with progressive dataset size reductions, results begin to less accurately reflect interspecific morphological differences and fail to represent the full spectra of pollen grain morphology within species and within individuals.

CART and Random Forest models derived from the reduced sample size datasets delineate different morphological traits for classifying pollen grains into species groups, and error rates are substantially increased for those models that include all three species. These trends are further amplified when sample size is reduced in combination with a reduction in the number of grains per sample. These reductions result in a complete shift, and/or re-ranking of importance, in the morphological splitting variables identified by CART and Random Forest models as necessary for differentiating alder pollen types. Furthermore, classification accuracy is diminished from 70% to less than 20%, when a CART model derived from a reduced dataset is used to classify test set grains (Table 3.26). Perhaps most interesting is the fact that CART model classification success actually increases to 73.5% for *A. incana* subsp. *tenuifolia* pollen compared to the models derived from the full dataset, where accuracies were between 0% and 5%. Also interesting is the reduction in model error in successively smaller datasets when *A. incana* subsp. *tenuifolia* is removed from models. It is therefore worrisome that a study design that includes fewer samples per species and fewer pollen grains per sample can result in the false conclusions that: a) *A. incana* subsp. *tenuifolia* pollen grains can be discriminated from *A. viridis* subsp. *sinuata* and *A. rubra* pollen and that b) a different suite of traits for separating *A. viridis* subsp. *sinuata* and *A. rubra* pollen can be used to differentiate *A. viridis* subsp. *sinuata* and *A. rubra* pollen with classification accuracies of

over 90% (Table 3.19 and 3.28). These results demonstrate that sample size can have a substantial impact on conclusions about interspecific differences in pollen morphology. As small sample sizes fail to represent the natural variability in pollen morphology, studies using small sample sizes may therefore draw false conclusions regarding the morphological differentiation of pollen from different species.

Statistical results are less sensitive to random reductions in the number of pollen grains measured per sample than to reductions in the number of samples per species. In these reduced datasets, the traits used for classification are consistent between CART models and Random Forest models and model error rates increase only slightly. This implies that, for future studies, effort would be best placed in collecting pollen from many individuals of a species, as opposed to increasing the number of grains per sample. At least in alder, the morphology of pollen from individual plants is relatively similar.

While some palynological studies caution against using too small a representation of any given population when formulating a method for morphological differentiation of pollen (e.g. Mäkelä 1996; Lindbladh et al. 2002), this recommendation is not generally followed. The research presented here specifically outlines the potential risk of drawing conclusions from analyses performed with small sample sizes (< 15 samples per species) and draws into question the effectiveness of species level methods for distinguishing pollen types based on small sample sizes. For instance, in their study delineating a method for the separation of *Betula* pollen to species, Clegg et al. (2005) examined five species, but collected samples from only 5 – 13 plants per species. The results of this study are used to classify pollen into tree and shrub types based on threshold values for single morphological traits (pore depth and pore-diameter ratio). Over 80% of shrub

birches were identified correctly as such and 74.9% of tree birches. However, given the results of sensitivity analyses performed here, it is unlikely that intraspecific differences in pollen morphology were adequately represented in their morphometric measurements and interspecific trait overlap may have been minimized. Had Clegg et al. (2005) included a larger number of samples per species, it is highly possible that they would not have been able to separate *Betula* tree pollen from shrub pollen. Moreover, percentages for correct classification presented by Clegg et al. (2005) represent the classification of pollen grains from the same individuals used to formulate their identification method. 'Other Sample (OS)' test sets used in this study show that classification success can be drastically reduced when a given classification scheme is used to separate pollen from individual plants not used in method development. Presenting classification success based only on the identification of pollen used to create the classification method is misleading.

Many palynological studies use sample sizes comparable to, or even smaller than those used by Clegg et al. (2005). In their study using CART analysis to differentiate the pollen of three species of *Picea*, Lindbladh et al. (2002) use on average 12 samples per species in model derivation, and like Clegg et al. (2005), base the success of their proposed model on the classification of pollen grains from the same *Picea* individuals that were used to create their model. The greater sensitivity of results to reductions in sample size shown here, when compared to reductions in the number of grains measured per sample, indicate that both of these studies would have been better served by examining more individual plants from each of their respective species and perhaps fewer grains per individual plant. If time is a limitation when formulating study designs for

pollen identification, efforts would be best placed in increasing the number of samples (i.e., sampling the pollen of more individual plants), as opposed to sampling more pollen grains from one individual. Nonetheless, both large sample size (> n=25 or more preferably $n \geq 30$) and numerous grain replications are recommended, especially when attempting to differentiate congeneric pollen types.

Using Spatially Biased Pollen Collections in the Study of Pollen Morphology

Nested ANOVA, CART modelling, and Random Forest analyses demonstrate that pollen morphology varies significantly by region. While there is extensive trait distribution overlap between species, there are also consistent trends in pollen morphology between regions. For example, *A. viridis* subsp. *sinuata* pollen from northern British Columbia and the Yukon are significantly smaller across all measured traits than *A. viridis* subsp. *sinuata* pollen grains from Vancouver Island and the mainland coast. *Alnus rubra* pollen grains from trees on Vancouver Island are, on average, larger than pollen from *A. rubra* on the mainland.

As was the case when the dataset was randomly reduced in size, when data from each species is split into regional subsets, the resulting CART model identifies different splitting variables than for CART models derived from the full dataset. The regional CART model (Fig. 3.24) derived from Vancouver Island *A. rubra* pollen and coastal *A. viridis* subsp. *sinuata* pollen has surprisingly low error (13.2%) when classifying pollen grains used to derive the model; however, when the model is used to classify *A. rubra* pollen from the mainland coast, accuracy is drastically reduced from 81.4% to 10.5% (Table 3.34). This research shows that basing pollen identification on samples from only a limited region within a species' range can result in conclusions about pollen

morphology and/or pollen classification that are not representative of pollen grains from the larger population as a whole.

The regional differences in *A. rubra* pollen morphology may be a result of genetic variation in whole-plant morphological and ecophysiological traits between *A. rubra* 'families' (Dang et al. 1994). Individual *A. rubra* trees have varied survival success across different regions, and survive much better when planted near their place of origin (Xie et al. 2008). Common garden experiments have grouped *A. rubra* into three 'families' within British Columbia based on genotype × environment interactions (Hamann et al. 2000). This genetic structure is a result of the fact that west coast *A. rubra* populations are remnants of the mesophytic forests of the Miocene (Furlow 1979) and have had time to genetically diversify within specific regions (Chen and Li 2004; Xie et al. 2008). *Alnus viridis* subsp. *sinuata* is a more recent colonizer of the Pacific Northwest, entering this area in the Quaternary (Furlow 1979). It is interesting therefore that the morphology of *A. viridis* subsp. *sinuata* pollen also varies by region. One reason may be that *A. viridis* and *A. incana* species complexes display within species genetic differentiation between habitats as a result of adaptations to particular environmental conditions, and possibly as a result of their need to facilitate symbiont genotype selection (Anderson et al. 2009). Given that there is genetic plasticity in both *A. rubra* and *A. viridis* subsp. *sinuata* across regions, should it then to be expected that these genetic differences will impact pollen morphology?

Regional differences in pollen morphology will likely vary depending on the species and/or genus in question. This research only identifies regional differences in the pollen morphology of alder species in western North America. Nonetheless, this research

underscores the importance of sampling from throughout a species' range if undertaking research concerned with pollen identification. It is troublesome therefore that Lindbladh et al. (2002) recommend that pollen collections for morphological studies should be based on samples from limited regions within a species range, effectively biasing the morphological variation represented in these collections. In devising a method to distinguish the pollen of two *Pinus* species in New England, Barton et al. (2011) cite this recommendation by Lindbaldh et al. (2005) as the reason for collecting pollen samples from a very limited region. CART modelling of this *Pinus* dataset resulted in 71% to 90% classification accuracy, leading Barton et al. (2011) to conclude that the low error in their classification model further supports the argument for species level differentiation of pollen based on models derived from samples collected from small regions within a species' range. Here, the CART model produced from regional subsets of alder pollen also results in very low model error (i.e. 13.2%), but this regional model is not effective for identifying alder pollen from outside of the model region. This research is the first to demonstrate the biased CART modelling that can result when samples are collected from limited regions.

Broader Implications for the Science of Palynology

Perhaps the most interesting implication of the regional differences identified here in alder pollen morphology is in how this finding relates to the broader palynological assumption of temporal and spatial stability in pollen morphology. The foundation of paleoecological reconstructions of past vegetation is accurate pollen identification, which is based on correlation in pollen morphology between pollen from modern plants and fossil pollen grains found in sediment records (Fægri and Iversen 1989). The correlation

between past and present pollen assumes that pollen morphology has not changed through time. However, if pollen morphology of individual species differs between regions, it may not be safe, in all instances, to assume that pollen morphology is conserved over long periods of time.

Davis (2000) argues that making assumptions, such as the temporal and spatial stability of pollen morphology, may be based on flawed underlying theory and may therefore be detrimental to the science of paleoecology. The temporal stability of pollen through time is widely assumed in palynology because no new species have been identified as arising in the Quaternary (Huntley 2001) and species level adaptation is considered unimportant when set against a backdrop of rapidly changing Quaternary climate regimes (Walker 1990; Huntley 1991; Whitlock and Bartlein 1997; Huntley 2001). However, recent experiments in plant ecology reveal plant adaptation over much shorter time frames than expected (Davis et al. 2005). These transplant experiments, which indicate that discrete populations of a species grow better in their original environment than when they are transplanted to a different site occupied by the same species (Davis et al. 2005), produce similar results to transplantation experiments that have been performed using *A. rubra* (Xie et al. 2008).

While the regional differences in alder pollen morphology presented here do not directly lead to a discussion of short-term plant adaptation and / or how it may impact pollen morphology, they nonetheless suggest that the assumption of temporal stability of pollen morphology requires thorough examination. To argue that species level identification of pollen should be based on regional datasets (e.g. Linbladh et al. 2002; Barton et al. 2011) does a disservice to the science of palynology. If methods for pollen

identification are not transferrable across a current species range, then they cannot be assumed to be effective for identifying pollen through time either, particularly given the dramatic shifts in species ranges that have occurred during the Quaternary (e.g. Williams et al. 2004). If, in some species, pollen morphology differs significantly over distances of a few hundred kilometres, then pollen morphology is also likely to shift over tens of thousands of years. What is needed are pollen identification methods determined using large collections that include many individual plants from across an entire species range. Methods for pollen discrimination based on this type of study design have the best chance of representing the morphological variability present in an entire population and pinpointing a combination of morphological traits that are useful for pollen identification, regardless of slight morphological changes over space or time.

Conclusion

This research demonstrates clearly that no single morphological trait can be used to distinguish the pollen of the three alder species that occur in the Pacific Northwest. The pollen morphology of each species is highly variable across traits, with a large degree of trait overlap between species. While a single morphological trait cannot be used for alder pollen separation, multi-trait CART modelling provides trait thresholds for differentiating *A. viridis* subsp. *sinuata* pollen and *A. rubra* pollen into ecologically relevant morphotypes (i.e., *A. rubra* - type and *A. viridis* - type) analogous to species separation. The main morphological traits that distinguish these two pollen morphotypes are annulus width, diameter, arci strength, exine thickness and grain shape. The number of pores per grain cannot be used to identify alder pollen to species (cf. Reinink-Smith 2010). *Alnus incana* subsp. *tenuifolia* pollen cannot be distinguished from that of *A. viridis* subsp. *sinuata* and *A. rubra.* Given the intermediate morphology of *A. incana* subsp. *tenuifolia* pollen, the proposed morphotype method should only be used in areas where it can be safely assumed that *A. incana* subsp. *tenuifolia* is absent and has been absent from an area for the entire period of record. While this sort of *a priori* assumption about the absence of a particular species through time can be dangerous, in certain areas such as coastal British Columbia, this assumption is likely safe as no *A. incana* subsp. *tenuifolia* macrofossils have been found in late Quaternary sediments.

The validity of using certain pollen traits, namely diameter, exine thickness and pore number, to identify pollen types has been questioned due to potential correlations between grain size, exine thickness and the number of apertures per grain and correlations between these traits and environmental conditions (Lee 1978; Kurtz et al.

1958). While pore number is not being posited here as a useful identification trait, it has been suggested that the number of pores per grain may vary depending where on an individual plant a pollen grain is formed (Flenley 2003). Moreover, grain size and shape are the likeliest pollen grain attributes to be impacted by laboratory treatments (Mäkelä 1996). However, much of the risk involved in the use of these traits for pollen identification is mitigated in the proposed morphotype method for differentiating the pollen of *A. viridis* subsp. *sinuata* and *A. rubra* because no trait is used in isolation to identify either pollen type. In addition, the large number of samples collected for each alder species allows a realistic representation of within individual and between individual trait variations.

Sensitivity analysis shows that in order to capture the variability in pollen morphology within a single species, low sample sizes must be avoided when attempting to identify diagnostic morphological traits. In addition, since individual plants produce morphologically similar pollen, it is better to assess fewer grains from as many individual plants as possible, rather than to measure numerous grains from only a small number of plants. Regional differences in alder pollen morphology indicate that effort should be made to include samples from across a species' range, so that identification criteria are not biased by regional differences in pollen morphology. Caution is advised when using methods for pollen identification that have been derived from samples collected from limited regions (cf. Barton et al. 2011). Lastly, the regional differences in morphology identified in both *A. viridis* subsp. *sinuata* and *A. rubra* pollen highlight the potential need for further investigation into the assumption of spatial and temporal stability of

pollen morphology as well as the amount of variation in pollen morphology attributable to shifting regional genotypes.

References

Allen, G.B., K.J. Brown, and R.J. Hebda. 1999. Surface pollen spectra from southern Vancouver Island, British Columbia, Canada. Canadian Journal of Botany **77**: 786-799.

Andersen, T. 1960. Silicone oil as a mounting medium for pollen grains. Danmarks Geologisk Undersögelse **4**: 1–24.

Anderson, M.D., R.W. Ruess, D.D. Myrold, and D.L. Taylor. 2009. Host species and habitat affect nodulation by specific *Frankia* genotypes in two species of *Alnus* in interior Alaska. Oecologia **160**: 619–630.

Arsenault, A., J.J. Clague, and R.W. Mathewes. 2007. Late Holocene vegetation and climate change at Moraine Bog, Tiedemann Glacier, southern Coast Mountains, British Columbia. Canadian Journal of Earth Sciences **44**: 707-719.

Banner, A., J. Pojar and G.E. Rouse. 1983. Postglacial paleoecology and successional relationships of a bog woodland near Prince Rupert, British Columbia. Canadian Journal of Forest Restoration. **13**: 938-947.

Barton, A.M., A.M. Nurse, K. Michaud, and S.W. Hardy. 2011. Use of CART analysis to differentiate pollen of red pine (*Pinus resinosa*) and jack pine (*P. banksiana*) in New England. Quaternary Research **75**: 18-23.

Bennett, K.D., and K.J. Willis. 2001. Pollen. In: *Tracking Environmental Change Using Lake Sediments. Volume 3: Terrestrial, Algal, and Silicaceous Indicators*. J.P. Smol, H.J.B. Birks and W.M. Last (eds). Kluwer Academic Publishers, Dordrecht, The Netherlands.

Birks, H.J.B. 1993. Quaternary paleoecology and vegetation science – current contributions and possible future developments. Review of Palaeobotany and Palynology **79**: 153-177.

Blackmore, S., J.A.J. Steinmann, P.P. Hoen, and W. Punt. 2003. The northwest European pollen flora, 65, Betulaceae and Corylaceae. Review of Paleobotany and Palynology **123**: 71-98.

Bormann, B.T., and R.C. Sidle. 1990. Changes in productivity and distribution of nutrients in a chronosequence at Glacier Bay National Park, Alaska. Journal of Ecology **78**: 561-578.

Brasier, M.D. 1980. *Microfossils*. Chapman and Hall, London, United Kingdom.

Breiman, L. 2001. Random Forests. Machine Learning **45**: 5-32.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Wadsworth, Belmont, CA..

Brown, K.J., and R.J. Hebda. 2003. Coastal rainforest connections disclosed through a Late Quaternary vegetation, climate, and fire history investigation from the Mountain Hemlock Zone on southern Vancouver Island, British Columbia, Canada. Review of Palaeobotany and Palynology **123**: 247–269.

Carl Zeiss MicroImaging. 2008. Axiovision Release 4.7.1. (http://www.zeiss.de/axiovision). Carl Zeiss. Jena, Germany.

Chapin, F.S., L.R. Walker, C.L. Fastie, and L.C. Sharman. 1994. Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. Ecological Monographs **64**: 149-175.

Chen, Z., and J. Li. 2004. Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. International Journal of Plant Sciences **165**: 325-335.

Connell, J.H., and R.O. Slatyer. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. The American Naturalist **111**: 1119-1144.

Clague, J.J., B. Wohlfarth, J. Ayotte, M. Eriksson, I. Hutchinson, R.W. Mathewes, I.R. Walker, and L. Walker. 2004. Late Holocene environmental change at treeline in the northern Coast Mountains, British Columbia, Canada. Quaternary Science Reviews **23**: 2413-2431.

Clegg, B.F., W. Tinner, D.G. Gavin, and F.S. Hu. 2005. Morphological differentiation of *Betula* (birch) pollen in northwest North America and its paleoecological application. The Holocene **15**: 229-237.

Dang, Q.L., C.Y. Xie, C.C. Ying, and R.D. Guy. 1994. Genetic variation of ecophysiological traits in red alder (*Alnus rubra*). Canadian Journal of Forest Restoration **24**: 2140-2156.

Davis, M.B. 2000. Palynology after Y2K – understanding the source area of pollen in sediments. Annual Review of Earth and Planetary Sciences **28**: 1-18.

Davis, M.B., R.G. Shaw, and J.R. Etterson. 2005. Evolutionary responses to changing climate. Ecology **86**: 1704-1714.

Delcourt, H.R., and P.A. Delcourt. 1991. *Quaternary Ecology: A Paleoecological Perspective*. Chapman and Hall, London.

Douglas, G.W., G.B. Straley, D.V. Meidinger, and J. Pojar (eds). 1998. *Illustrated Flora of British Columbia. Volume 2: Dicotyledons (Balsaminaceae Through Cucurbitaceae).* B.C. Ministry of Environment, Lands & Parks and B.C. Ministry of Forests. Victoria.

Fægri, K., and J. Iversen. 1989. *Textbook of Pollen Analysis*. Fourth ed. Blackwell, Oxford.

Finkelstein, S.A., K. Gajewski, and A.E. Viau. 2006. Improved resolution of pollen taxonomy allows better biogeographical interpretation of post-glacial forest development: Analyses from North American Pollen Database. Journal of Ecology **94**: 415-430.

Flenley, J. 2003. Some prospects for lake sediment analysis in the 21st century. Quaternary International **105**: 77-80.

Flora of North America North of Mexico. 1993+. Flora of North America Editorial Committee, eds. 16 vols. New York, NY.

Furlow, J.J. 1979. The systematics of the American species of *Alnus* (Betulaceae). Rhodora **81**: 1-121.

Gavin, D.G., J.S. Mclachlan, L.B. Brubaker, and K.A. Young. 2001. Postglacial history of subalpine forests, Olympic Peninsula, Washington, USA. The Holocene **11**: 177-188.

Glover, T., and K. Mitchell. 2002. *An Introduction to Biostatistics*. Waveland Press Inc. Long Grove, IL.

Gray J. 1991. Tetrahedraletes, Nodospora and the 'cross' tetrad: An accretion of myths. In: *Pollen and spores: Patterns of diversification, Systematics Association Special Volume Number 44*. S. Blackmore, and S. Barnes (eds.). Oxford University Press, New York, NY: 49–87.

Grienenberger, E., S.S. Kim, B. Lallemand, P. Geoffroy, D. Heintz, C. Souza, T.Heitz, C.J. Douglas, and M. Legrand. 2010. Analysis of tetraketide a-pyrone reductase function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. The Plant Cell. **22**: 4067-4083.

Hamann, A., M.P. Koshy, G. Namkoong, and C.C. Ying. 2000. Genotype x environmental interactions in *Alnus rubra*: developing seed zones and seed transfer guidelines with spatial statistics and GIS. Ecological Management **136**: 107-119.

Hansen, B.C.S., and D.R. Engstrom. 1996. Vegetation history of Pleasant Island, southeastern Alaska, since 13,000 yr B.P. Quaternary Research **46**: 161-175.

Hebda, R.J. 1995. British Columbia vegetation and climate history with focus on 6 ka BP. Géographie physique et Quaternaire **49**: 55–79.

Heusser, C.J. 1973. Modern pollen spectra from Mount Rainer,Washington. Northwest Science **47**: 1–8.

Hu, F.S., B.P. Finney, and L.B. Brubaker. 2001. Effects of Holocene *Alnus* expansion on aquatic productivity, nitrogen cycling, and soil development in southwestern Alaska. Ecosystems **4**: 358-368.

Huntley, B. 1991. How plants respond to climate change: Migration rates, individualism and the consequences for plant communities. Annals of Botany **67**: 15-22.

Huntley, B. 2001. Reconstructing past environments from the Quaternary paleovegetation record. Proceedings of the Royal Irish Academy **101**: 1-18.

Huntley, B., R.A. Spicer, W.G. Chaloner, and E.A. Jarzembowski. 1993. The use of climate response surfaces to reconstruct palaeoclimate from Quaternary pollen and plant macrofossil data. Philosophical Transactions of the Royal Society of London **341**: 215-224.

Kapp, R.O., O.K. Davis, and J.E. King. 2000. *Ronald O. Kapp's Pollen and Spores*. 2nd Edition. American Association of Stratigraphic Palynologists Foundation. College Station, TX.

Kurtz E.B., and J.L. Liverman. 1958. Some effects of temperature on pollen characters. Bulletin of the Torrey Botanical Club **85**: 136–138.

Lacourse, T. 2005. Late Quaternary dynamics of forest vegetation on northern Vancouver Island, British Columbia, Canada. Quaternary Science Reviews **24**: 105-121.

Lacourse, T. 2009. Environmental change controls postglacial forest dynamics through interspecific differences in life-history traits. Ecology **90**: 2149-2160.

Lacourse, T., R.W. Mathewes, and D.W. Fedje. 2005. Late-glacial vegetation dynamics of the Queen Charlotte Islands and adjacent continental shelf, British Columbia, Canada. Palaeogeography, Palaeoclimatology, Palaeoecology **226**: 36-57.

Lantz, T.C., S.E. Gergel, and G.H.R. Henry. 2010. Response of green alder (*Alnus viridis* subsp. *fruticosa*) patch dynamics and plant community composition to fire and regional temperature in north-western Canada. Journal of Biogeography **37**: 1597-1610.

Lee, S. 1978. A factor fnalysis study of the functional significance of angiosperm pollen. Systematic Botany **3**: 1-19.

Liaw A., and M. Wiener. 2002. Classification and Regression by randomForest. R News **2**(3): 18-22.

Lindbladh, M., R. O'Connor, and G.L. Jacobson. 2002. Morphological analysis of pollen grains for paleoecological studies: Classification of *Picea* from eastern North America. American Journal of Botany **89**: 1459-1467.

Lindbladh, M., W.W. Oswald, D.R. Foster, E.K. Faison, J. Hou, and Y. Huang. 2007. A late-glacial transition from *Picea glauca* to *Picea mariana* in southern New England. Quaternary Research **67**: 502-508.

Livingstone, D.A. 1955. Some pollen profiles from Arctic Alaska. Ecology **36**: 587-600.

Logan, M. 2010. *Biostatistical Design and Analysis Using R: A Practical Guide*. Wiley-Blackwell. West Sussex, United Kingdom.

MacDonald, G.M., and J.C. Ritchie. 1986. Modern pollen spectra from the western interior of Canda and the interpretation of late Quaternary vegetation development. New Phytologist **103**: 245-268.

Mäkelä, E.M. 1996. Size distinctions between *Betula* pollen types – A review. Grana **35**: 248-256.

Mathewes, R.W. 1973. A palynological study of postglacial vegetation changes in the University Research Forest, southeastern British Columbia. Canadian Journal of Botany **51**: 2085-2103.

Mayle, F.E., A.J. Levesque, and L.C. Cwynar. 1993. *Alnus* as an indicator taxon of the Younger Dryas cooling in eastern North America. Quaternary Science Reviews **12**: 295-305.

Minckley, T.A., P.J. Bartlein, C. Whitlock, B.N. Shuman, J.W. Williams, and O.K. Davis. 2008. Associations among modern pollen, vegetation, and climate in western North America. Quaternary Science Reviews **27**: 1962-1991.

Navarro, E., J. Bousquet, A. Moiroud, A. Munive, D. Piou, and P. Normand. 2003. Molecular phylogeny of *Alnus* (Betulaceae), inferred from nuclear ribosomal DNA ITS sequences. Plant and Soil **254**: 207-217.

Niinemets, Ü., and F. Valladares. 2006. Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecological Monographs **76**: 521-547.

Payne, R.J., M. Lamentowicz, and E.A.D. Mitchell. 2011. The perils of taxonomic inconsistency in quantitative palaeocology: experiments with testate amoeba data. Boreas **40**: 15-27.

Prentice, I.C. 1988. Records of vegetation in time and space: the principles of pollen analysis. *Vegetation History*. B. Huntley and T. Webb (eds.). Kluwer Academic Publishers. Norwell, MA: 18-42.

Raven, P.H., R.F. Evert and S.E. Eichorn. 2005. *Biology of Plants*. 7th Edition. W.H. Freeman and Company Publishers. New York, NY.

Reinink-Smith, L.M. 2010. Variations in alder pore number – a possible new correlation tool for the Neogene Kenai lowland, Alaska. Palynology **34**: 180-194.

Richard, P. 1970. Atlas pollinique des arbres et de quelques arbustes indigénes du Quèbec. III. Angiospermes (Salicacées, Myricacées, Juglandacées, Corylacées, Fagacées, Ulmacées). Le Naturaliste Canadien **97**: 97-161.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Seppä, H., and K.D. Bennett. 2003. Quaternary pollen analysis: recent progress in paleoecology and paleoclimatology. Progress in Physical Geography **4**: 548-579.

SPSS for Windows. 2002. Release 11.5.0, Standard Version. LEAD Technologies Inc. Microsoft Corporation.

Sutton, C.D. 2004. Classification and regression trees, bagging and boosting. *Handbook of Statistics*. Volume 24. C.R. Rao, E.J. Wegman, and J.L. Solka (eds.). Elsevier. Amsterdam, The Netherlands: 303-329.

Thompson, R.S., K.H. Anderson, and P. J. Bartlein. 1999. U.S. Geological Survey Professional Paper 1650 A&B. In: *Atlas of Relations Between Climatic Parameters and Distributions of Important Trees and Shrubs in North America*. URL=http://greenwood.cr.usgs.gov/pub/ ppapers/p1650-a/pages/conifers.html.

Therneau, T.M., B. Atkinson, and B. Ripley. 2009. Rpart: Recursive Partitioning. R package. Version 3.1-45. (http://CRAN.R-project.org/package=rpart).

Titus, J.H. 2009. Nitrogen-fixers *Alnus* and *Lupinus* influence soil characteristics but not colonization by later successional species in primary succession on Mount St. Helens. Plant Ecology **203**: 289–301.

USDA, NRCS. 2011. The PLANTS Database (http://plants.usda.gov, 7 April 2011). National Plant Data Center, Baton Rouge, LA 70874-4490 USA.

Venables, W.N., and B.D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, NY.

Walker, D. 1990. Purpose and method in Quaternary palynology. Review of Palaeobotany and Palynology **64**: 13-27.

Whitehead, D.R. 1961. A note on silicone oil as a mounting medium for fossil and modern pollen. Ecology **42**: 591.

Whitlock, C., and P.J. Bartlein. 1997. Vegetation and climate change in northwest America during the past 125 kyr. Nature **388**: 57-60.

Whitlock, M., and D. Schluter. 2009. *The Analysis of Biological Data*. Roberts & Co, Greenwood Village, CO..

Williams, J.W., D.M. Post, L.C. Cwynar, A.F. Lotter and A.J. Levesque. 2002. Rapid and widespread vegetation responses to past climate change in the North Atlantic region. Geology **30**: 971-974.

Williams, J.W., B.N. Shuman, T. Webb, P.J. Bartlein, and P.L. Leduc. 2004. Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecological Monographs **74**: 309-334.

Wittborn, J., K.V. Rao, G. El-Ghazaly, and J.R. Rowley. 1996. Substructure of spore and pollen grain exines in *Lycopodium*, *Alnus*, *Betula*, *Fagus* and *Rhododendron*. Grana **35**: 185-198.

Wright, H.E., J.E. Kutzbach, T. Webb III, W.F. Ruddiman, F.A. Street-Perrott, and P.J. Bartlein (eds.). 1993. *Global Climates Since the Last Glacial Maximum*. University of Minnesota Press, Minneapolis, MN.

Xie, C.Y. 2008. Ten-year results from red alder (*Alnus rubra*) provenance-progeny testing and their implications for genetic improvement. New Forests **36**: 273–284.

Appendix A – *Alnus* Pollen Reference Samples

Table A.1: Sample List and Collection Locations for *Alnus* Herbarium Specimens

Species	Sample ID	Sample Location*
Alnus incana	43248[1]	Sumass Mt., 1 km west of Quading farm, BC
subsp. *tenuifolia*	V105976[2]	Juab near Nephi, UT
(*n*=27)	V107227[2]	Similkameen River Valley, BC
	V109975[2]	Jasper National Park, Athabasca River banks, near Pocahontas, AB
	V111096[2]	6.4 km north of Trail, Columbia River Valley, BC
	V165973[2]	Bulkley-Nechako Regional District, Maclure Lake near Telkwa, BC
	V170499[2]	10 km north of Clearwater Village (De Kelver's place), BC
	V177016[2]	Stikine River basin, north end of Eddontenajon Lake, BC
	V205824[2]	Fraser-Fort George Regional District, Miworth, Prince George, BC
	V221486[2]	Peace River Regional District, Carbon Creek, Ten Mile Creek, BC
	V45684[2]	Big Lake, Caribou, BC
	V61455[2]	above Nelson, BC
	V6417[2]	Dawson, YT
	V6421[2]	Armstrong, Okanagan, BC
	V65180[2]	Sheep Creek at Turner Valley, AB
	V72230[2]	Jaffray, BC
	V73200[2]	Cedar Mt, Latah County, ID
	V134341[3]	Trail, BC
	V29468[3]	Nelson, BC
	V061521[3]	Trail, Topping Creek, BC
	V082070[3]	near Lower Post, BC
	V160213[3]	Anarchist Mt, BC
	V149481[3]	White Lake, Salmon Arm Forest District, BC
	V151891[3]	McLure Ferry, Kamloops, BC
	V151907[3]	Osoyoos Oxbows, north of Lake, BC
	RM25[4]	Buck Hills Ecological Reserve, near Lumby, BC
	RM26[4]	near Radium, BC
Alnus rubra	003837[1]	Allouette Lake area, BC
(*n*=31)	003841[1]	Cowichan Station, SW Duncan (Koksilah River Drainage), BC
	003843[1]	Beaver Lake, BC
	30974[1]	University of Victoria campus, BC
	30995[1]	University of Victoria campus, Victoria, BC
	24722[1]	Cordova Bay, Victoria, BC
	V112076[2]	Lighthouse Park, west Vancouver, trail 1, grid-3d, BC
	V125583[2]	west edge of Santa Cruz Mts., 2 miles south of Davenport, CA
	V137600[2]	Powell River, BC
	V138413[2]	Spanish Banks, University of British Columbia area, Vancouver, BC
	V196177[2]	1.6 km west of Falls City, Black Rock rd, sec. 17, Polk County, OR
	V213587[2]	Vancouver Island, Port Alberni, Somass River delta, BC
	V28165[2]	Spruce Cove, Trinidad, Humboldt County, CA
	V6337[2]	Vancouver, West Point Grey, BC

	V6338[2]	Masset, BC
	V6345[2]	Greater Vancouver Regional District, Burquitlam, BC
	V6681[2]	Dryas Island in Fraser River at Hope, BC
	V71258[2]	Oak Ranch Creek, 1.6 km E of route 47, Columbia County, OR
	V78952[2]	Vancouver, Point Grey, BC
	V88028[2]	Cheam Lake, BC
	V78631[2]	Oak Ranch Creek, 1.6 km E of route 47, Columbia County, OR
	V1391[3]	Esquimalt district, Goldstream, BC
	V34428[3]	Fernwood, Saltspring Island, BC
	V101645[3]	Killarney Rd, Victoria, BC
	V47631[3]	Kitimat, BC
	V148972[3]	Ganges, Salt Spring Island, BC
	V115524[3]	Port Neville Inlet, BC
	V98039[3]	Mayne Island, BC
	V126473[3]	Somass Delta, Port Alberni, BC
	V187605[3]	Kal Lake Provincial Park, Vernon, Cosens Creek, BC
	RM28[4]	UBC Campus, Vancouver, BC
Alnus viridis	23960[1]	Lower Bertha Falls, BC
subsp. *sinuata*	31308[1]	4.8 km S of Nadina River crossing, W end of Francois Lake, BC
(*n*=35)	014952[1]	Timothy Mtn, Timothy Lake, Lac la Hache, BC
	003865[1]	Matheson Lake, Vancouver Island, BC
	029190[1]	Haley Lake, Marmot District, Vancouver Island, BC
	003864[1]	Mt. Arrowsmith, Vancouver Island, BC
	V125584[2]	King County, Deception Creek, 12.8 km W of Stevens Pass, WA
	V137595[2]	Powell Lake, Rainbow Lodge, BC
	V3409[2]	Bennett, BC
	V60917[2]	Gray Creek, BC
	V45951[2]	Snohomish County, Cascade Mountains, Perry Creek trail, WA
	V87864[3]	Alsek River, ca. 2.5 km N of the British Columbia Boundary, AK
	V1389[3]	Mt. Mclean, BC
	V6363[3]	Skidegate, Queen Charlotte Islands (Haida Gwaii), BC
	V29472[3]	bank of Kicking Horse Gorge, Yoho National Park, AB
	V22839[3]	Langara Island, Queen Charlotte Islands (Haida Gwaii), BC
	V45469[3]	Allouette Lake area, BC
	V47627[3]	Kitimat, BC
	V47628[3]	Kitselas Canyon Rd., NE of Terrace, BC
	V49593[3]	S shore of Charlotte Lake, W. Chilcotin, BC
	V54399[3]	9.6 km north of Prince George, BC
	V062037[3]	Copper Island, Atlin, BC
	V060843[3]	Silver Creek, Hope, BC
	V114447[3]	Rossland, Record Ridge, BC
	V069231[3]	Murray Ridge, NE of Ft St. James, BC
	V075162[3]	Red Pass, Park Headquarters, BC
	V115522[3]	Port Neville Inlet, BC
	V144564[3]	Stewart-Cassiar Hwy, BC
	V87436[3]	Alsek River, 5km S of the Yukon border, BC
	V112152[3]	4 km SE of Mackenzie Ranger Station, BC
	V182763[3]	McBride Peak, 7 km from highway up Mount View Rd, BC

V165960[3]	Sewell Inlet, S of Sewell Inlet, Moresby Island, BC
V38329[3]	Lake Whatcom, WA
V89293[3]	Apex Mt. Rd, 1.6 km from intersection with Old Penticton Rd., BC
V139552[3]	Chase Shuswap Territory, near Pillar Lake, BC

[1] University of Victoria Herbarium, Department of Biology, Victoria, BC

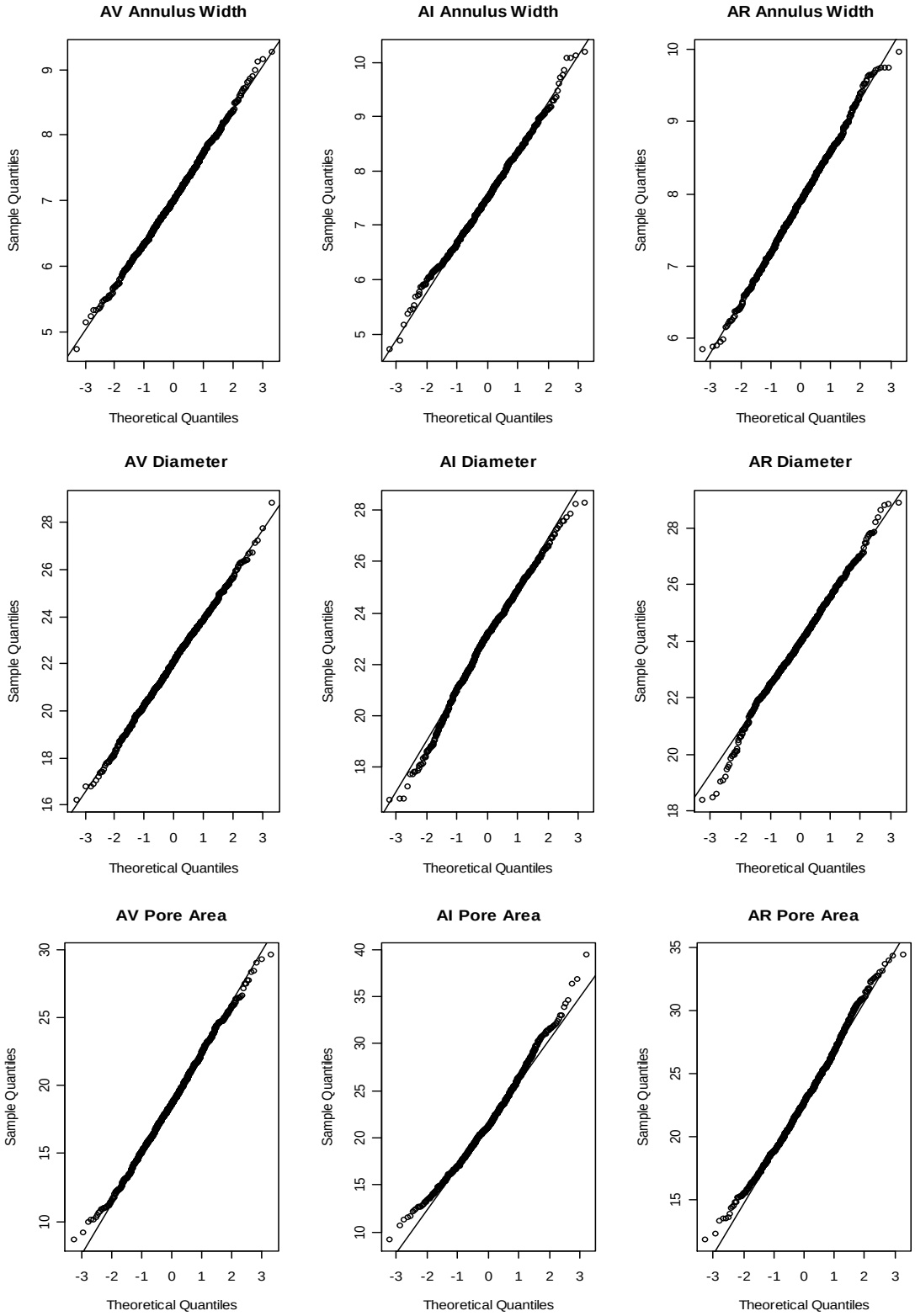
[2] University of British Columbia Herbarium, Beaty Biodiversity Museum, Vancouver, BC

[3] Royal British Columbia Museum Herbarium, Victoria, BC

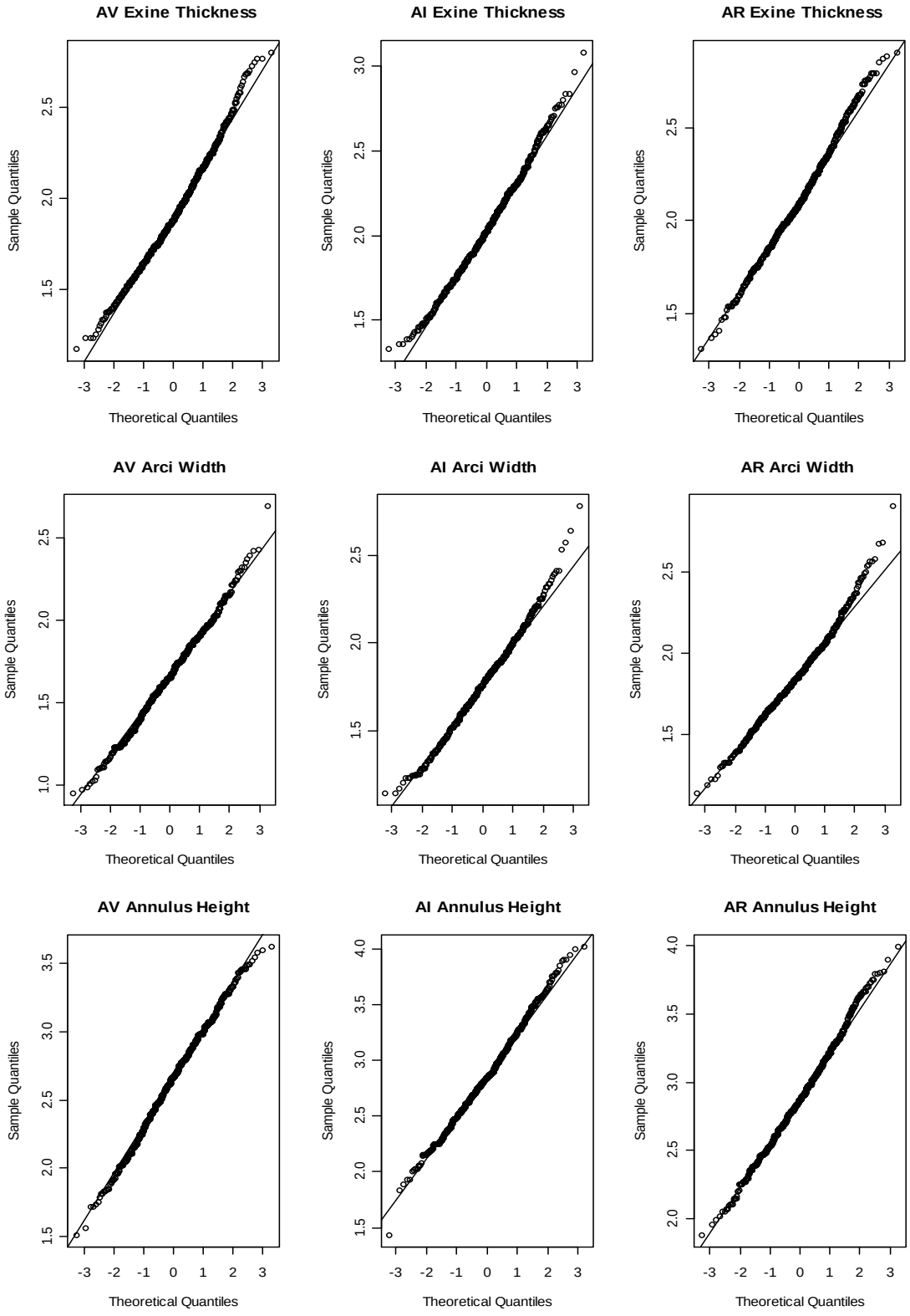
[4] The pollen reference collection of Dr. R. Mathewes, Simon Fraser University, Burnaby, BC

* Sample locations as per herbarium reference sheets

Appendix B – QQ Plots *



* AV=*Alnus viridis* subsp. *sinuata*, AI=*Alnus incana* subsp. *tenuifolia* and AR=*Alnus rubra*


AV Exine Thickness
AI Exine Thickness
AR Exine Thickness
AV Arci Width
AI Arci Width
AR Arci Width
AV Annulus Height
AI Annulus Height
AR Annulus Height
Sample Quantiles
Theoretical Quantiles

Appendix C – Results of Additional Sensitivity Analyses

Table C.1: Nested ANOVA (*n*=15 samples/species)

Quantitative Trait by Source of Variance		Nested ANOVA Model		Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.***	**% Variance**
Between Species	2	14.89	7.44	0.4587	76.9
Between Individuals	42	23.60	0.56	0.0147	2.5
Between Grains	1305	160.80	0.12	0.1232	20.7
Annulus Width (µm)					
Between Species	2	165.45	82.73	5.1607	90.0
Between Individuals	42	223.63	5.32	0.1637	2.9
Between Grains	1305	532.14	0.41	0.4077	7.1
Annulus Height (µm)					
Between Species	2	11.58	5.79	0.3313	72.8
Between Individuals	42	34.28	0.82	0.0240	5.3
Between Grains	1305	129.80	0.10	0.0995	21.9
Annulus Area (μm^2)					
Between Species	2	3504.40	1752.20	106.2180	86.6
Between Individuals	42	6675.10	158.93	4.9127	4.0
Between Grains	1305	15067.00	11.55	11.5460	9.4
Exine Thickness (µm)					
Between Species	2	8.48	4.24	0.2380	77.7
Between Individuals	42	28.26	0.67	0.0207	6.7
Between Grains	1305	62.44	0.05	0.0478	15.6
Grain Diameter (µm)					
Between Species	2	1076.90	538.44	33.6707	91.7
Between Individuals	42	1401.80	33.38	1.0460	2.8
Between Grains	1305	2613.90	2.00	2.0030	5.5

* Var.Comp. = the variance component (i.e., variance explained by each nested factor)

Table C.2: Pair-wise nested ANOVA (*n*=15 samples/species).

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: (U)(2/n_1n_2)
Arci Width	**Pair-wise***	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	36.70	**< 0.001**	4.18	**< 0.001**	0.75	1.000	76.2
Alnus incana	**V-R**	115.67	**< 0.001**	5.43	**< 0.001**	0.94	1.000	60.8
Alnus rubra	**I-R**	22.21	**< 0.001**	3.88	**< 0.001**	0.79	1.000	60.8
Annulus Width								
Alnus viridis	**V-I**	89.26	**< 0.001**	17.09	**< 0.001**	1.66	0.238	72.1
Alnus incana	**V-R**	421.33	**< 0.001**	7.93	**< 0.001**	1.22	1.000	38.3
Alnus rubra	**I-R**	114.2	**< 0.001**	14.6	**< 0.001**	1.4	1.000	66.9
Annulus Height								
Alnus viridis	**V-I**	82.22	**< 0.001**	9.19	**< 0.001**	1.02	1.000	70.5
Alnus incana	**V-R**	92.61	**< 0.001**	7.01	**< 0.001**	0.95	1.000	68.5
Alnus rubra	**I-R**	0.17	1.000	8.36	**< 0.001**	0.94	1.000	98.7
Annulus Area								
Alnus viridis	**V-I**	135.95	**< 0.001**	17.35	**< 0.001**	1.56	0.469	66.3
Alnus incana	**V-R**	311.52	**< 0.001**	9.84	**< 0.001**	1.27	1.000	46.3
Alnus rubra	**I-R**	31.39	**< 0.001**	14.63	**< 0.001**	1.47	0.836	81.6
Exine Thickness								
Alnus viridis	**V-I**	109.73	**< 0.001**	13.1	**< 0.001**	1.97	**0.026**	67.5
Alnus incana	**V-R**	155.89	**< 0.001**	13.25	**< 0.001**	2.04	**0.015**	60.3
Alnus rubra	**I-R**	3.65	**< 0.001**	17.1	**< 0.001**	1.27	1.000	92.9
Grain Diameter								
Alnus viridis	**V-I**	48.24	**< 0.001**	21.1	**< 0.001**	0.85	1.000	76.9
Alnus incana	**V-R**	512.39	**< 0.001**	12.36	**< 0.001**	0.80	1.000	34.2
Alnus rubra	**I-R**	248.18	**< 0.001**	16.22	**< 0.001**	1.04	1.000	54.4

* Pair-wise abbreviations: *A. viridis* subsp. *sinuata* (V), *A. incana* subsp *tenuifolia* (I) and *A. rubra* (R)

Table C.3: Nested ANOVA (*n*=7 samples/species)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	2.84	1.42	0.0857	20.8
Between Individuals	18	14.69	0.82	0.1519	36.8
Between Grains	608	106.73	0.18	0.1755	42.5
Annulus Width (µm)					
Between Species	2	33.17	16.59	1.4300	59.9
Between Individuals	18	118.51	6.58	0.5624	23.6
Between Grains	608	239.67	0.39	0.3942	16.5
Annulus Height (µm)					
Between Species	2	2.29	1.14	0.0057	1.3
Between Individuals	18	19.87	1.10	0.3333	74.3
Between Grains	608	66.44	0.11	0.1093	24.4
Annulus Area (μm^2)					
Between Species	2	734.54	367.27	21.6643	63.0
Between Individuals	18	3881.20	215.62	0.5911	1.7
Between Grains	608	7390.00	12.16	12.1550	35.3
Exine Thickness (µm)					
Between Species	2	5.96	2.98	0.3329	40.7
Between Individuals	18	11.67	0.65	0.4333	53.0
Between Grains	608	31.50	0.05	0.0517	6.3
Grain Diameter (µm)					
Between Species	2	188.62	94.31	3.0929	49.5
Between Individuals	18	1307.80	72.66	1.3308	21.3
Between Grains	608	1107.50	1.82	1.8215	29.2

Table C.4: Pair-wise nested ANOVA (*n*=7 samples/species)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	5.74	0.289	3.42	**0.002**	1.17	1.000	79.5
Alnus incana	**V-R**	16.32	**0.001**	6.39	**< 0.001**	1.47	1.000	65.0
Alnus rubra	**I-R**	2.59	1.000	4.28	**< 0.001**	1.15	1.000	83.1
Annulus Width								
Alnus viridis	**V-I**	25.17	**< 0.001**	16.85	**< 0.001**	2.41	**0.048**	77.3
Alnus incana	**V-R**	91.32	**< 0.001**	13.81	**< 0.001**	3.17	**0.002**	58.5
Alnus rubra	**I-R**	19.24	**< 0.001**	22.42	**< 0.001**	2.97	**0.004**	80.9
Annulus Height								
Alnus viridis	**V-I**	18.13	**< 0.001**	6.63	**< 0.001**	0.81	1.000	80.4
Alnus incana	**V-R**	13.65	**0.004**	11.22	**< 0.001**	1.38	1.000	85.0
Alnus rubra	**I-R**	0.52	1.000	12.5	**< 0.001**	1.13	1.000	96.2
Annulus Area								
Alnus viridis	**V-I**	34.27	**< 0.001**	14.09	**< 0.001**	1.59	0.080	74.8
Alnus incana	**V-R**	62.13	**< 0.001**	18.29	**< 0.001**	2.08	0.193	68.2
Alnus rubra	**I-R**	2.66	1.000	21.71	**< 0.001**	1.60	1.000	92.5
Exine Thickness								
Alnus viridis	**V-I**	54.21	**< 0.001**	12.22	**< 0.001**	1.07	1.000	64.9
Alnus incana	**V-R**	113.09	**< 0.001**	11.98	**< 0.001**	1.57	1.000	50.7
Alnus rubra	**I-R**	10.25	**0.025**	13.84	**< 0.001**	1.59	1.000	82.4
Grain Diameter								
Alnus viridis	**V-I**	1.00	1.000	45.45	**< 0.001**	0.76	1.000	98.2
Alnus incana	**V-R**	92.48	**< 0.001**	38.72	**< 0.001**	0.76	0.716	66.4
Alnus rubra	**I-R**	64.97	**< 0.001**	34.28	**< 0.001**	0.69	1.000	69.8

Table C.5: Nested ANOVA (All Samples, 20 grains/sample)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	22.30	11.15	0.3387	66.8
Between Individuals	90	58.49	0.65	0.0255	5.0
Between Grains	1767	252.97	0.14	0.1432	28.2
Annulus Width (µm)					
Between Species	2	250.82	125.41	3.9361	88.3
Between Individuals	90	305.14	3.39	0.1510	3.4
Between Grains	1767	650.93	0.37	0.3684	8.3
Annulus Height (µm)					
Between Species	2	19.37	9.68	0.2894	68.8
Between Individuals	90	64.11	0.71	0.0305	7.2
Between Grains	1767	178.39	0.10	0.1010	24.0
Annulus Area (μm^2)					
Between Species	2	5739.20	2869.60	88.8313	84.5
Between Individuals	90	10425.00	115.83	5.2390	5.0
Between Grains	1767	19529.00	11.05	11.0520	10.5
Exine Thickness (µm)					
Between Species	2	14.22	7.12	0.2129	74.2
Between Individuals	90	46.67	0.52	0.0235	8.2
Between Grains	1767	89.43	0.05	0.0506	17.6
Grain Diameter (µm)					
Between Species	2	1156.10	578.06	17.6913	83.8
Between Individuals	90	2667.00	29.63	1.3795	6.5
Between Grains	1767	3599.40	2.04	2.0370	9.7

Table C.6: Pair-wise nested ANOVA (All Samples, 20 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	26.59	**< 0.001**	4.36	**< 0.001**	1.11	1.000	78.6
Alnus incana	**V-R**	166.13	**< 0.001**	4.65	**< 0.001**	1.16	1.000	58.5
Alnus rubra	**I-R**	46.81	**< 0.001**	4.82	**< 0.001**	1.19	1.000	79.1
Annulus Width								
Alnus viridis	**V-I**	178.00	**< 0.001**	11.36	**< 0.001**	1.54	0.078	66.7
Alnus incana	**V-R**	698.60	**< 0.001**	6.2	**< 0.001**	1.26	1.000	35.4
Alnus rubra	**I-R**	135.04	**< 0.001**	10.86	**< 0.001**	1.28	1.000	68.9
Annulus Height								
Alnus viridis	**V-I**	107.51	**< 0.001**	7.19	**< 0.001**	0.95	1.000	72.4
Alnus incana	**V-R**	169.51	**< 0.001**	6.83	**< 0.001**	0.97	1.000	66.1
Alnus rubra	**I-R**	2.69	1.000	7.13	**< 0.001**	1.01	1.000	94.5
Annulus Area								
Alnus viridis	**V-I**	210.21	**< 0.001**	11.64	**< 0.001**	1.36	0.569	65.1
Alnus incana	**V-R**	540.85	**< 0.001**	8.83	**< 0.001**	1.21	1.000	44.1
Alnus rubra	**I-R**	45.42	**< 0.001**	11.39	**< 0.001**	1.21	1.000	80.1
Exine Thickness								
Alnus viridis	**V-I**	86.23	**< 0.001**	11.39	**< 0.001**	1.61	**0.038**	75.3
Alnus incana	**V-R**	281.78	**< 0.001**	9.03	**< 0.001**	1.67	**0.012**	56.4
Alnus rubra	**I-R**	47.54	**< 0.001**	11.48	**< 0.001**	1.48	0.203	80.8
Grain Diameter								
Alnus viridis	**V-I**	96.60	**< 0.001**	16.84	**< 0.001**	1.11	1.000	74.8
Alnus incana	**V-R**	585.74	**< 0.001**	12.6	**< 0.001**	1.00	1.000	43.8
Alnus rubra	**I-R**	163.74	**< 0.001**	14.57	**< 0.001**	1.45	0.276	69.4

Table C.7: Nested ANOVA (All Samples, 10 grains/sample)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	11.46	5.73	0.1723	49.6
Between Individuals	90	34.89	0.39	0.0240	6.9
Between Grains	837	126.60	0.15	0.1512	43.5
Annulus Width (µm)					
Between Species	2	148.72	74.36	2.3332	81.1
Between Individuals	90	182.43	2.03	0.1650	5.7
Between Grains	837	317.61	0.38	0.3795	13.2
Annulus Height (µm)					
Between Species	2	14.26	7.12	0.2165	62.6
Between Individuals	90	36.70	0.41	0.0320	9.3
Between Grains	837	81.42	0.09	0.0973	28.1
Annulus Area (μm^2)					
Between Species	2	3711.30	1855.60	57.7142	77.7
Between Individuals	90	5981.10	66.46	5.5470	7.5
Between Grains	837	9194.70	10.99	10.9850	14.8
Exine Thickness (µm)					
Between Species	2	5.05	2.52	0.0713	47.3
Between Individuals	90	27.75	0.31	0.0260	17.3
Between Grains	837	44.70	0.05	0.0534	35.4
Grain Diameter (µm)					
Between Species	2	631.95	315.98	9.7681	75.6
Between Individuals	90	1185.60	13.17	1.1130	8.6
Between Grains	837	1709.90	2.04	2.0429	15.8

Table C.8: Pair-wise nested ANOVA (All samples, 10 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	10.63	**0.020**	2.38	**< 0.001**	0.82	1.000	79.1
Alnus incana	**V-R**	77.58	**< 0.001**	2.3	**< 0.001**	0.78	1.000	58.1
Alnus rubra	**I-R**	27.97	**< 0.001**	3.22	**< 0.001**	1.36	0.706	79.4
Annulus Width								
Alnus viridis	**V-I**	91.79	**< 0.001**	5.38	**< 0.001**	0.93	1.000	66.1
Alnus incana	**V-R**	415.15	**< 0.001**	4.4	**< 0.001**	1.17	1.000	33.1
Alnus rubra	**I-R**	83.56	**< 0.001**	6.56	**< 0.001**	1.32	0.993	65.8
Annulus Height								
Alnus viridis	**V-I**	76.57	**< 0.001**	3.57	**< 0.001**	0.64	1.000	66.5
Alnus incana	**V-R**	127.75	**< 0.001**	4.05	**< 0.001**	0.81	1.000	59.4
Alnus rubra	**I-R**	1.57	1.000	4.7	**< 0.001**	0.75	1.000	94.5
Annulus Area								
Alnus viridis	**V-I**	121.64	**< 0.001**	5.18	**< 0.001**	0.64	1.000	61.6
Alnus incana	**V-R**	362.32	**< 0.001**	5.79	**< 0.001**	1.10	1.000	39.1
Alnus rubra	**I-R**	30.23	**< 0.001**	7.09	**< 0.001**	1.06	1.000	78.3
Exine Thickness								
Alnus viridis	**V-I**	33.35	**< 0.001**	5.52	**< 0.001**	1.13	1.000	76.1
Alnus incana	**V-R**	88.69	**< 0.001**	5.37	**< 0.001**	1.27	1.000	64.6
Alnus rubra	**I-R**	12.37	**0.008**	7.07	**< 0.001**	1.17	1.000	87.4
Grain Diameter								
Alnus viridis	**V-I**	31.13	**< 0.001**	7.39	**< 0.001**	1.01	1.000	80.1
Alnus incana	**V-R**	311.89	**< 0.001**	5.94	**< 0.001**	1.01	1.000	42.3
Alnus rubra	**I-R**	112.96	**< 0.001**	5.91	**< 0.001**	0.83	1.000	60.4

Table C.9: Nested ANOVA (*n*=15, 20 grains/sample)

Quantitative Trait by Source of Variance		Nested ANOVA Model		Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	9.85	4.93	0.2747	60.7
Between Individuals	42	34.20	0.81	0.0335	7.4
Between Grains	855	123.14	0.14	0.1440	31.8
Annulus Width (µm)					
Between Species	2	126.70	63.35	4.0273	88.4
Between Individuals	42	123.40	2.94	0.1270	2.8
Between Grains	855	341.57	0.40	0.3995	8.8
Annulus Height (µm)					
Between Species	2	6.97	3.48	0.1907	60.4
Between Individuals	42	26.25	0.62	0.0260	8.2
Between Grains	855	84.59	0.10	0.0989	31.3
Annulus Area (μm^2)					
Between Species	2	2513.40	1256.70	77.3967	83.4
Between Individuals	42	4021.30	95.75	4.2270	4.6
Between Grains	855	9582.60	11.21	11.2080	12.1
Exine Thickness (µm)					
Between Species	2	4.71	2.35	0.1187	61.5
Between Individuals	42	24.08	0.57	0.0260	13.5
Between Grains	855	41.33	0.05	0.0483	25.0
Grain Diameter (µm)					
Between Species	2	507.98	253.99	14.7760	80.8
Between Individuals	42	1358.70	32.35	1.5175	8.3
Between Grains	855	1709.90	2.00	1.9999	10.9

Table C.10: Pair-wise nested ANOVA (*n*=15, 20 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	6.54	0.188	5.12	**< 0.001**	0.96	1.000	84.0
Alnus incana	**V-R**	70.46	**< 0.001**	5.19	**< 0.001**	1.08	1.000	56.3
Alnus rubra	**I-R**	32.65	**< 0.001**	6.95	**< 0.001**	1.35	1.000	73.9
Annulus Width								
Alnus viridis	**V-I**	145.03	**< 0.001**	6.67	**< 0.001**	1.23	1.000	54.1
Alnus incana	**V-R**	319.99	**< 0.001**	6.63	**< 0.001**	1.04	1.000	36.2
Alnus rubra	**I-R**	27.06	**< 0.001**	8.82	**< 0.001**	1.09	1.000	79.4
Annulus Height								
Alnus viridis	**V-I**	55.25	**< 0.001**	6.2	**< 0.001**	0.87	1.000	71.4
Alnus incana	**V-R**	51.29	**< 0.001**	6.58	**< 0.001**	1.34	1.000	72.4
Alnus rubra	**I-R**	1.08	1.000	6.24	**< 0.001**	0.92	1.000	96.3
Annulus Area								
Alnus viridis	**V-I**	133.66	**< 0.001**	7.97	**< 0.001**	0.92	1.000	58.6
Alnus incana	**V-R**	230.82	**< 0.001**	8.59	**< 0.001**	1.20	1.000	46.3
Alnus rubra	**I-R**	3.22	1.000	9.04	**< 0.001**	0.89	1.000	90.2
Exine Thickness								
Alnus viridis	**V-I**	22.63	**< 0.001**	15.1	**< 0.001**	1.54	0.557	82.8
Alnus incana	**V-R**	99.95	**< 0.001**	8.44	**< 0.001**	1.62	0.341	61.3
Alnus rubra	**I-R**	27.04	**< 0.001**	12.73	**< 0.001**	1.06	1.000	81.0
Grain Diameter								
Alnus viridis	**V-I**	94.62	**< 0.001**	17.82	**< 0.001**	0.88	1.000	66.3
Alnus incana	**V-R**	256.45	**< 0.001**	15.9	**< 0.001**	0.56	1.000	47.9
Alnus rubra	**I-R**	32.28	**< 0.001**	14.17	**< 0.001**	0.96	1.000	81.3

Table C.11: Nested ANOVA (*n*=15, 10 grains/sample)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	3.48	1.74	0.0993	51.4
Between Individuals	42	10.41	0.25	0.0170	8.8
Between Grains	405	31.12	0.08	0.0768	39.8
Annulus Width (µm)					
Between Species	2	50.38	25.19	1.5527	71.9
Between Individuals	42	79.97	1.90	0.1440	6.7
Between Grains	405	187.94	0.46	0.4640	21.5
Annulus Height (µm)					
Between Species	2	3.93	1.97	0.1080	46.4
Between Individuals	42	14.55	0.35	0.0250	10.7
Between Grains	405	40.33	0.10	0.0996	42.8
Annulus Area (μm^2)					
Between Species	2	1126.30	563.13	33.7233	66.9
Between Individuals	42	2405.90	57.28	4.5080	8.9
Between Grains	405	4941.60	12.20	12.2020	24.2
Exine Thickness (µm)					
Between Species	2	2.37	1.19	0.0620	48.4
Between Individuals	42	10.97	0.26	0.0220	17.2
Between Grains	405	17.82	0.04	0.0440	34.4
Grain Diameter (µm)					
Between Species	2	445.12	222.56	14.0340	81.8
Between Individuals	42	505.90	12.05	0.9910	5.8
Between Grains	405	865.97	2.14	2.1382	12.5

Table C.12: Pair-wise nested ANOVA (*n*=15, 10 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	16.45	**0.001**	3.04	**< 0.001**	0.52	1.000	75.9
Alnus incana	**V-R**	39.62	**< 0.001**	3.78	**< 0.001**	0.67	1.000	63.8
Alnus rubra	**I-R**	5.29	0.379	2.19	**0.015**	0.56	1.000	84.2
Annulus Width								
Alnus viridis	**V-I**	22	**< 0.001**	5.4	**< 0.001**	1.29	1.000	67.6
Alnus incana	**V-R**	126.55	**< 0.001**	2.78	**< 0.001**	1.34	1.000	40.6
Alnus rubra	**I-R**	32.7	**< 0.001**	4.59	**< 0.001**	1.79	0.149	68.0
Annulus Height								
Alnus viridis	**V-I**	27.45	**< 0.001**	3.71	**< 0.001**	1.18	1.000	68.4
Alnus incana	**V-R**	32.73	**< 0.001**	2.99	**< 0.001**	1.15	1.000	67.9
Alnus rubra	**I-R**	0.05	1.000	3.77	**< 0.001**	0.88	1.000	97.7
Annulus Area								
Alnus viridis	**V-I**	42.66	**< 0.001**	6.02	**< 0.001**	1.63	0.385	71.8
Alnus incana	**V-R**	108.46	**< 0.001**	3.57	**< 0.001**	1.51	0.792	46.1
Alnus rubra	**I-R**	9.15	**0.046**	5.19	**< 0.001**	1.58	0.515	81.6
Exine Thickness								
Alnus viridis	**V-I**	32.66	**< 0.001**	4.5	**< 0.001**	1.48	0.839	76.1
Alnus incana	**V-R**	46.79	**< 0.001**	5.85	**< 0.001**	1.76	0.175	62.5
Alnus rubra	**I-R**	0.81	1.000	9.29	**< 0.001**	1.40	1.000	94.1
Grain Diameter								
Alnus viridis	**V-I**	15.24	**0.002**	6.14	**< 0.001**	0.68	1.000	76.3
Alnus incana	**V-R**	212.3	**< 0.001**	4.72	**< 0.001**	0.77	1.000	28.1
Alnus rubra	**I-R**	96.05	**< 0.001**	5.45	**< 0.001**	0.86	1.000	49.9

Table C.13: Nested ANOVA (*n*=7, 20 grains/sample)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	8.29	4.14	0.5100	73.6
Between Individuals	18	10.29	0.57	0.0205	3.0
Between Grains	399	64.83	0.16	0.1625	23.4
Annulus Width (µm)					
Between Species	2	74.91	37.48	4.8286	89.2
Between Individuals	18	66.29	3.68	0.1630	3.0
Between Grains	399	167.13	0.42	0.4189	7.7
Annulus Height (µm)					
Between Species	2	5.72	2.86	0.3057	71.5
Between Individuals	18	12.98	0.72	0.0315	7.4
Between Grains	399	36.04	0.09	0.0903	21.1
Annulus Area (µm^2)					
Between Species	2	1614.00	806.98	98.8914	85.8
Between Individuals	18	2065.40	114.74	5.1765	4.5
Between Grains	399	4471.50	11.21	11.2070	9.7
Exine Thickness (µm)					
Between Species	2	5.01	2.51	0.2814	80.1
Between Individuals	18	9.78	0.54	0.0250	7.1
Between Grains	399	17.96	0.04	0.0449	12.8
Grain Diameter (µm)					
Between Species	2	404.54	202.27	23.9714	86.5
Between Individuals	18	620.54	34.47	1.6170	5.8
Between Grains	399	851.58	2.13	2.1343	7.7

Table C.14: Pair-wise nested ANOVA (*n*=7, 20 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	1.26	1.000	3.6	**0.00**	0.63	1.000	81.2
Alnus incana	**V-R**	51.98	**< 0.001**	3.23	**0.00**	0.70	1.000	47.4
Alnus rubra	**I-R**	34.99	**< 0.001**	3.52	**0.00**	0.97	1.000	58.7
Annulus Width								
Alnus viridis	**V-I**	11.79	**0.120**	11.51	**< 0.001**	0.63	1.000	71.2
Alnus incana	**V-R**	177.21	**< 0.001**	7.65	**< 0.001**	1.34	1.000	32.8
Alnus rubra	**I-R**	87.11	**< 0.001**	6.95	**< 0.001**	0.86	1.000	47.0
Annulus Height								
Alnus viridis	**V-I**	28.6	**< 0.001**	11.88	**< 0.001**	1.18	1.000	73.2
Alnus incana	**V-R**	64.19	**< 0.001**	8.34	**< 0.001**	0.85	1.000	55.1
Alnus rubra	**I-R**	6.52	**0.191**	4.51	**< 0.001**	0.94	1.000	83.7
Annulus Area								
Alnus viridis	**V-I**	27.81	**< 0.001**	14.57	**< 0.001**	0.78	1.000	71.9
Alnus incana	**V-R**	154.83	**< 0.001**	10.31	**< 0.001**	0.98	1.000	37.7
Alnus rubra	**I-R**	40.29	**< 0.001**	6.41	**< 0.001**	0.73	1.000	59.9
Exine Thickness								
Alnus viridis	**V-I**	24.99	**< 0.001**	12.79	**< 0.001**	0.99	1.000	81.1
Alnus incana	**V-R**	110.51	**< 0.001**	11.44	**< 0.001**	1.66	0.954	45.8
Alnus rubra	**I-R**	32.77	**< 0.001**	12.7	**< 0.001**	1.46	1.000	70.1
Grain Diameter								
Alnus viridis	**V-I**	12.05	**0.010**	22.6	**< 0.001**	0.82	1.000	82.8
Alnus incana	**V-R**	185.12	**< 0.001**	9.19	**< 0.001**	0.91	1.000	32.9
Alnus rubra	**I-R**	94.21	**< 0.001**	15.79	**< 0.001**	1.01	1.000	55.1

Table C.15: Nested ANOVA (*n*=7, 10 grains/sample)

Quantitative Trait by Source of Variance	Nested ANOVA Model			Variance Component Analysis	
Arci Width (µm)	**df**	**SS**	**MS**	**Var.Comp.**	**% Variance**
Between Species	2	1.71	0.86	0.0843	35.7
Between Individuals	18	4.93	0.27	0.0130	5.5
Between Grains	189	26.28	0.14	0.1391	58.8
Annulus Width (µm)					
Between Species	2	26.78	13.39	1.7557	78.9
Between Individuals	18	19.80	1.10	0.0700	3.1
Between Grains	189	75.82	0.40	0.3989	17.9
Annulus Height (µm)					
Between Species	2	2.39	1.69	0.1943	59.9
Between Individuals	18	5.93	0.33	0.0220	6.8
Between Grains	189	20.43	0.11	0.1081	33.3
Annulus Area (μm^2)					
Between Species	2	736.91	368.46	46.2771	74.8
Between Individuals	18	801.42	44.52	3.2170	5.2
Between Grains	189	2334.40	12.35	12.3510	20.0
Exine Thickness (µm)					
Between Species	2	1.36	0.68	0.0486	36.7
Between Individuals	18	6.12	0.34	0.0290	21.9
Between Grains	189	10.35	0.05	0.0547	41.4
Grain Diameter (µm)					
Between Species	2	221.72	110.86	14.1314	81.8
Between Individuals	18	214.95	11.94	0.9770	5.7
Between Grains	189	410.10	2.17	2.1698	12.6

Table C.16: Pair-wise nested ANOVA (*n*=7, 10 grains/sample)

Trait by Species		Nested Comparison (Species)		Nested Comparison (Species:Sample)		Nested Comparison (Species:Sample:Grain)		U Statistic: $(U)(2/n_1n_2)$
Arci Width	**Pair-wise**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**F-Ratio**	**P-value**	**Overlap (%)**
Alnus viridis	**V-I**	3.44	1.000	1.05	1.000	0.82	1.000	78.1
Alnus incana	**V-R**	11.87	**0.014**	2.12	**0.351**	0.59	1.000	65.3
Alnus rubra	**I-R**	2.57	1.000	2.98	**0.021**	1.14	1.000	85.6
Annulus Width								
Alnus viridis	**V-I**	40.41	**< 0.001**	2.98	**0.020**	1.28	1.000	47.9
Alnus incana	**V-R**	71.18	**< 0.001**	3.08	**0.014**	1.93	0.484	37.2
Alnus rubra	**I-R**	1.23	1.000	2.53	**0.098**	0.81	1.000	92.3
Annulus Height								
Alnus viridis	**V-I**	35.68	**< 0.001**	3.47	**0.004**	1.57	1.000	53.4
Alnus incana	**V-R**	12.16	**0.012**	3.65	**0.002**	1.21	1.000	70.4
Alnus rubra	**I-R**	4.57	0.554	2.65	**0.066**	1.97	0.445	85.2
Annulus Area								
Alnus viridis	**V-I**	54.03	**< 0.001**	3.81	**0.001**	1.46	1.000	46.4
Alnus incana	**V-R**	48.42	**< 0.001**	4.72	**< 0.001**	1.63	1.000	50.4
Alnus rubra	**I-R**	0.95	1.000	3.17	**0.011**	1.44	1.000	94.5
Exine Thickness								
Alnus viridis	**V-I**	13.65	**0.005**	3.83	**0.001**	0.37	1.000	70.1
Alnus incana	**V-R**	23.74	**< 0.001**	6.28	**< 0.001**	1.20	1.000	62.7
Alnus rubra	**I-R**	0.68	1.000	9.08	**< 0.001**	1.58	1.000	93.1
Grain Diameter								
Alnus viridis	**V-I**	28..59	**< 0.001**	5.42	**< 0.001**	0.98	1.000	60.1
Alnus incana	**V-R**	111.70	**< 0.001**	7.63	**< 0.001**	1.02	1.000	32.4
Alnus rubra	**I-R**	20.66	**< 0.001**	3.76	**0.001**	0.94	1.000	63.4

Appendix D – Results of Correlation Analyses

Table D.1: Pearson's Product Moment Correlations between Quantitative Variables and Latitude.

Variable by Species Group	t-statistic	*r*	P-value
A. viridis* subsp. *sinuata			
Arci Width	-4.32	-0.654	**< 0.001**
Annulus Width	-1.10	-0.215	0.281
Annulus Height	-2.17	-0.397	**0.039**
Annulus Area	-1.87	-0.350	0.073
Exine Thickness	-1.24	-0.241	0.226
Diameter	0.38	0.075	0.710
A. incana* subsp. *tenuifolia			
Arci Width	-0.41	-0.082	0.684
Annulus Width	-0.17	-0.036	0.860
Annulus Height	-0.32	-0.065	0.748
Annulus Area	-0.17	-0.036	0.860
Exine Thickness	-0.03	-0.007	0.973
Diameter	-1.25	-0.243	0.222
A. rubra			
Arci Width	0.68	0.134	0.504
Annulus Width	-0.28	-0.056	0.780
Annulus Height	-0.61	-0.122	0.545
Annulus Area	-0.49	-0.098	0.626
Exine Thickness	1.53	0.293	0.138
Diameter	-0.41	-0.081	0.687
All Species			
Arci Width	-2.91	-0.292	**0.005**
Annulus Width	-3.48	-0.343	**0.001**
Annulus Height	-3.07	-0.306	**0.003**
Annulus Area	-3.41	-0.337	**< 0.001**
Exine Thickness	-2.38	-0.242	**0.019**
Diameter	-3.94	-0.382	**< 0.001**

Table D.2: Spearman's Rank Correlations between Qualitative Variables and Latitude.

Variable by Species Group	S-statistic	*r*	P-value
A. viridis* subsp. *sinuata			
Arci Strength	3234.77	0.013	0.950
Pore Protrusion	3424.73	-0.045	0.822
Grain Shape	4194.49	-0.280	0.157
Pore Number	3770.99	-0.151	0.452
A. incana* subsp. *tenuifolia			
Arci Strength	3268.07	0.002	0.990
Pore Protrusion	4238.61	-0.294	0.137
Grain Shape	4043.04	-0.234	0.240
Pore Number	3292.13	-0.005	0.981
A. rubra			
Arci Strength	3958.53	-0.208	0.297
Pore Protrusion	4525.31	-0.381	**0.050**
Grain Shape	3715.94	-0.134	0.504
Pore Number	2978.55	0.091	0.652
All Species			
Arci Strength	168064.90	-0.254	**0.014**
Pore Protrusion	152751.20	-0.140	0.182
Grain Shape	109784.70	0.181	0.083
Pore Number	133791.3	0.002	0.986

Appendix E – R Statistical Code *

QQ Plot Analysis of Normality:

```
>qqnorm(trait, main="QQ Plot Title")
```

F-Test for Equal Variance:

```
> var.test(trait1, trait2)
```

Standard Boxplot:

```
>par(mfrow=c(1,1))
>boxplot(AVtrait, AItrait, ARtrait), names=c("A.viridis","A.incana","A.rubra"),ylab="Trait Name" , col=c("green","blue","red"))
```

Pearson's Product Moment Correlation Test:

```
>wilcox.test(trait1, trait2, paired=FALSE)
```

Spearman's Rank Correlation Test:

```
>wilcox.test(trait1, trait2, paired=FALSE)
```

Single Sample ANOVA with Tukey's Correction:

```
>ModelName<-aov(trait~species)
>summary(ModelName)
>TukeyHSD(ModelName)
```

Nested ANOVA Model:

```
>library(nlme)
>ModelName<-aov(trait~Species+Error(Species/Sample/Grain), dataset))
>summary(ModelName)
```

Pair-wise Nested ANOVA Comparison with Bonferroni Correction:

```
>ModelName<-lm(trait~Species/Sample/Grain)
>anova(ModelName)
>p.adjust(anova(ModelName))
```

CART Modeling using Rpart:

```
>library(rpart)
# Fit the Model
>ModelName<-factor(dataset$Species,levels=0:1,labels=c("A.incana","A.rubra","A.viridis"))
>fit1<-rpart(ModelName~trait1+trait2…+trait8, data=dataset,method='class')
# Results Summary
```

* dataset specific code shown in red

CART Modeling using Rpart (continued):

```
>print(fit1)
>printcp(fit1)
>plotcp(fit1)
>summary(fit1)
# Plot Rpart Decision Tree
>plot(fit1,uniform=TRUE, main="Decision Tree Title")
>text(fit1, use.n=TRUE, all=TRUE, cex=.8)
# Prune the Decision Tree Model
>pfit1<- prune(fit1, cp=fit1$cptable[which.min(fit1$cptable[,"xerror"]),"CP"])
# Export Decision Tree
>post(pfit1, file = "C:\\Users\\Laura\\Desktop\\Tree.ps", title = "FileName")
# Test Model Predictions
>predict(pfit1,TestSetName ,type="class")
```

Random Forest Analysis:

```
>library(randomForest)
# Fit the Model
>ModelName<-factor(dataset$Species,levels=0:1,labels=c("A.rubra","A.viridis"))
>fit1<-randomForest(ModelName~ trait1+trait2…+trait8,data=dataset)
# Results Summary
>print(fit1)
>importance(fit1)
```

Linear Discriminant Function Analysis:

```
>library(MASS)
# Fit the Model
>fit1<-lda (Species~ trait1+trait2…+trait8, data=dataset, method="moment", CV=FALSE)
# Test Model Predictions
>test<-predict(fit1, TestSetName)
>table(test$class)
# Linear Discriminant Function Plot
> plot(fit1, main="PlotTitle", legend(legend=c("LegendText")))
```