THE COST OF LONGEVITY: LOSS OF SEXUAL FUNCTION IN NATURAL CLONES OF
Populus tremuloides

by

Dilara Ally

B.Sc., The University of Toronto, April 1998
M.Sc., The University of British Columbia, December 2000

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

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

January 2008

© Dilara Ally, 2008
ABSTRACT

Most clonal plants exhibit a modular structure at multiple levels. At the level of the organs, they are characterized by functional modules, such as, internodes, leaves, branches. At the level of the genetic individual (clone or genet), they possess independent evolutionary and physiological units (ramets). These evolutionary units arise through the widespread phenomenon of clonal reproduction, achieved in a variety of ways including rhizomes, stolons, bulbils, or lateral roots. The focus of this study was *Populus tremuloides*, trembling aspen, a dioecious tree that reproduces sexually by seed and asexually through lateral roots. Local forest patches in western populations of *Populus tremuloides* consisted largely of multiple genotypes. Multi-clonal patches were dominated by a single genotype, and in one population (Riske Creek) we found several patches (five out of 17) consisting of a single genotype. A second consequence of modularity is that during the repeated cycle of ramet birth, development and death, somatic mutations have the opportunity to occur. Eventually, the clone becomes a mosaic of mutant and non-mutant cell lineages. We found that neutral somatic mutations accumulated across 14 microsatellite loci at a rate of between $10^{-6}$ and $10^{-5}$ per locus per year. We suggest that neutral genetic divergence, under a star phylogeny model of clonal growth, is an alternative way to estimate clone age. Previous estimates of clone age couple the mean growth rate per year of shoots with the area covered by the clone. This assumes a positive linear relationship between clone age and clone size. We found, however, no repeatable pattern across our populations in terms of the relationship of either shape or size to the number of somatic changes. A final consequence of modularity is that during clonal growth, natural selection is relaxed for traits involving sexual function. This means that mutations deleterious to sexual function can accumulate, reducing the overall sexual fitness of a clone. We coupled neutral genetic divergence within clones with pollen fitness data to infer the rate and effect of mildly deleterious mutations. Mutations reduced relative sexual fitness in clonal aspen populations by about $0.12 \times 10^{-3}$ to $1.01 \times 10^{-3}$ per year. Furthermore, the decline in sexual function with clone age is evidence that clonal organisms are vulnerable to the effects of senescence.
# TABLE OF CONTENTS

**ABSTRACT** ........................................................................................................................... ii  
**TABLE OF CONTENTS** ........................................................................................................ iii  
**LIST OF TABLES** ............................................................................................................. iv  
**LIST OF FIGURES** ............................................................................................................... v  
**ACKNOWLEDGEMENTS** ...................................................................................................... vi  
**DEDICATION** ...................................................................................................................... viii  
**CO-AUTHORSHIP STATEMENT** .......................................................................................... ix  
**Chapter 1 : INTRODUCTORY REMARKS** ......................................................................... 1  
  REFERENCES .......................................................................................................................... 10  
**Chapter 2 : FINE-SCALE GENETIC STRUCTURING OF *Populus tremuloides* PATCHES**  
  IN THREE CLONAL POPULATIONS. ....................................................................................... 13  
  **INTRODUCTION** .............................................................................................................. 14  
  **MATERIALS AND METHODS** .......................................................................................... 18  
  **RESULTS** .......................................................................................................................... 29  
  **DISCUSSION** ...................................................................................................................... 35  
  REFERENCES .......................................................................................................................... 45  
**Chapter 3 : CAN CLONE SIZE SERVE AS A PROXY FOR CLONE AGE? AN**  
  **EXPLORATION USING COALESCENCE THEORY** ............................................................. 59  
  **INTRODUCTION** .............................................................................................................. 60  
  **MATERIAL AND METHODS** ............................................................................................. 65  
  **RESULTS** ............................................................................................................................ 74  
  **DISCUSSION** ...................................................................................................................... 82  
  REFERENCES .......................................................................................................................... 100  
**Chapter 4 : ON THE ESTIMATION OF CLONE AGE IN THE PRESENCE OF AGE-STRUCTURE**  
  **AND DENSITY-DEPENDENT GROWTH.** ............................................................................. 105  
  **INTRODUCTION** .............................................................................................................. 106  
  **METHODS** ........................................................................................................................ 111  
  **RESULTS AND DISCUSSION** ........................................................................................... 114  
  REFERENCES .......................................................................................................................... 127  
**Chapter 5 : A GENOMIC DELETERIOUS MUTATION RATE FOR THE LOSS OF SEX IN A**  
  **NATURAL POPULATION OF *Populus tremuloides*** ......................................................... 129  
  **MAIN TEXT** ...................................................................................................................... 130  
  **METHODS AND MATERIALS** .......................................................................................... 139  
  REFERENCES .......................................................................................................................... 177  
**Chapter 6 : CONCLUDING REMARKS** ............................................................................. 181  
  REFERENCES .......................................................................................................................... 185  
**APPENDIX A-2** ................................................................................................................ 199  
**APPENDIX B-2** ................................................................................................................ 206
LIST OF TABLES

Table 2-1. Estimates of clonal and genetic diversity across five microsatellite loci in three populations of Populus tremuloides............................................................... 43
Table 2-2. The log-likelihoods for the different relationship types in all three populations.................................................................................................................. 44
Table 3-1. A table describing the 14 microsatellite loci used in the study. ................. 96
Table 3-2. Somatic mutations detected across 14 microsatellite loci in two populations of P. tremuloides. .................................................................................................. 97
Table 3-3. Different linear models for the relationship between \( \pi_k \) and size ............... 98
Table 5-1. Estimates of regression coefficients for the full model on the raw data ...... 171
Table 5-2. The deleterious genomic mutation rate for sexual function obtained from Populus is comparable to estimates obtained from MA experiments on different organisms for traits related to sexual function...................................................... 172
STable 5-3. The final model is compared with the models rejected by the stepwise analysis...................................................................................................................... 173
STable 5-4. Estimates of regression coefficients for the square-root transformed data . 174
STable 5-5. A sensitivity analysis for the likelihood on \( M \) and \( V_E \) ......................... 175
STable 5-6. Maximum likelihood value of the shape parameter, \( \alpha \) ...................... 176
LIST OF FIGURES

Figure 2-1. (Colour plate) A map showing the location of the three *P. tremuloides* populations........................................................................................................................................... 39
Figure 2-2. Examples of patches from three populations of *P.tremuloides* illustrate the presence of multiple genets.................................................................................................................. 40
Figure 2-3. A single genotype predominates in a patch................................................................................................................. 41
Figure 2-4. The number of multi-locus genotypes per ramet sampled was significantly higher on the perimeter of a patch than in the centre of the patch.................................................. 42
Figure 3-1. A hydraulic excavation of part of a *Populus tremuloides* clone shows the pattern of interconnection among ramets.................................................................................................................. 90
Figure 3-2. Clonal growth through time............................................................... 91
Figure 3-3. A comparison of two different growth models...................................... 93
Figure 3-4. Clone shape does not follow a perfect circle........................................ 94
Figure 3-5. In Riske Creek, the presence of herbivory changes the area and morphology of a clone.................................................................................................................................................. 95
Figure 4-1. The life cycle used in the simulations.................................................. 120
Figure 4-2. A stable age distribution is reached ..................................................... 122
Figure 4-3. A plot of $\hat{\pi}_k$, the genome-wide pairwise estimate of age against true genet age............................................................. 123
Figure 4-4. Frequency distributions of average pairwise diversity at 20 microsatellite loci............................................................................................................................................... 124
Figure 4-5. $\hat{\pi}_k$ and genet age exhibit no relationship........................................... 125
Figure 4-6. In age-structured clones with density dependence (600 generations), sampling only from one age class has little impact on the estimate of true genet age ... 126
Figure 5-1. A genet/clone is initiated by a single genotype or seed............................ 161
Figure 5-2. In Riske Creek, for a given the environment, relative clone sexual fitness declines significantly with increasing clone age........................................................................... 162
Figure 5-3. A histogram showing the variation in mean sexual fitness in the Riske Creek population .................................................................................................................................................. 164
Figure 5-4. An ANOVA showed that state of catkin emergence at collection time did not explain sexual fitness variation in Riske Creek.......................................................... 165
Figure 5-5. A scatter plot shows that ramet age as estimated from tree ring data does not explain variation in mean sexual fitness.............................................................. 166
Figure 5-6. Results from an outlier analysis on the untransformed data with only clone age in the model........................................................................................................................................... 167
Figure 5-7. A linear regression of untransformed sexual fitness against moisture gradient was significant ($F_{1,18} = 4.759$, p value = 0.043, $R^2 = 0.1514$). ............................................................................. 168
Figure 5-8. The proportion of homozygous loci in a clone does not change with time, as measured by divergence .......................................................................................................................... 169
Figure 5-9. Sexual and asexual fitness measures are only weakly related.............. 170
ACKNOWLEDGEMENTS

“We do not receive wisdom, we must discover it for ourselves, after a journey through the wilderness, which no one else can make for us, which no one can spare us, for our wisdom is the point of view from which we come at last to regard the world.”

-Marcel Proust

I am indebted to a number of people who helped me during the course of my PhD. I am most grateful to my co-supervisors Sally Otto and Kermit Ritland. Thank you Sally, for challenging and insightful questions (always), enlightening discussions, frank comments, and for the small (and sometimes big) nudges of encouragement to move outside of my comfort zone. My scientific thinking has become more sophisticated through our interactions.

Thank you Kermit for providing me the resources and intellectual space to explore my creative scientific side regardless of the outcome.

My committee members: Michael Doebeli, Rosie Redfield, Joerg Bohlmann for your useful comments on my thesis.

I am grateful to Carol Ritland for help in the lab, as well for her support outside of school.

The end of a PhD is a written thesis, evidence of work done, not what and how we, as students struggled with and overcame. I believe a doctorate is partly earned in the struggle. The PhD is littered with small and big acomodadors. An acomodador is, as described by Paulo Coelho in his book The Zahir, “a giving up point…an event in our lives that is responsible for us failing to progress...can make cowards of us and prevent us from moving on.” I would like to thank my dear friends who, through much love and care, gave me courage to move forward. Thank you my ‘isa’ friends: Lisa O’Connell, Marissa LeBlanc, and Risa Sargent. Jessica Hill, for your wonderful wisdom and wit. Carrie Holt for your small packages of kindnesses. To Kim Goodliffe and Leslie Palleson for reminding me to play.

Many thanks also to OLMers: Chrissy Spencer, Aleeza Gerstein, Crispin Jordan (for our walks), Leithen M’Gonigle (for representing dorks ‘r us), Fred Guilllame, and Michelle Tseng. Also thanks to friends: Charles Chen, Jackie Ngai, Sam Yeaman, Alistair Blachford, and Allyson Longmuir

I would like to acknowledge and thank the Zoology Computing Unit: Sanja LeBlanc, Alistair Blachford, and Richard Sullivan. Without the ZCU, I would not have been able to run my likelihood programme.
Field Assistants: Katie Aitken, Jamie Skidmore, Mario-Pineda Krch. I owe a debt of gratitude to the most patient field assistant, Jodie Krakowski for putting up with my spatial and navigational dyslexia and my somewhat demanding personality.

I am grateful to Gillian Leung, Johnny Hsu, two tireless and committed volunteers who counted over 1 million pollen grains and Laura Glaubach, who helped me perform over 25,000 PCRs.

Field Sites and Equipment: Kathy Martin for use of the Riske Creek field site, Sally Aitken for the use of the microscope and pole pruner, Cyndi Smith (Waterton Lakes National Park), Claire Trethewey and the UBC Research Forest.

Blake Matthews, Diane Srivastava, Rasmus Astrup, Valerie Lemay, Michael Whitlock, and Carrie Holt for discussions on statistics and likelihood estimation.

NSERC, Nature Trust of BC: Brink/McLean Grassland Conservation Fund Award, UBC, PEO Education Fund for financial support during my PhD.

Finally and most importantly, I would like to thank Jabus Tyerman, my best friend. When the PhD seemed most daunting, you inspired with support and love, always quietly reminding me that I am more than just a scientist.
DEDICATION

TO MY FATHER
CO-AUTHORSHIP STATEMENT

Chapter 2 resulted from a collaboration among Sarah (Sally) Otto (research supervisor), Kermit Ritland (research supervisor), and me. Kermit wrote the likelihood program used to analyze relationships within and between aspen patches. Sally provided insightful and detailed comments on the manuscript. I designed the experiment, collected and analyzed the field and genetic data, and wrote the manuscript.

Chapter 3 was collaboration among Sally Otto, Kermit Ritland, and me. Sally assisted both in brainstorming about data analysis and in applying coalescence theory to clone age estimation. In addition, she provided detailed and insightful revisions to the manuscript. Kermit provided the resources and financial aid to carry out the field and molecular data collection. I designed the experiment, collected and analyzed the genetic data, and wrote most of the manuscript.

Chapter 4 resulted from a collaboration between Fred Guillaume and I. He wrote the individual-based stochastic simulation program, Nemo ver 2.0.2. Additionally, he helped design the experimental set up for the simulations and provided comments on the manuscript. I analyzed the data and wrote the manuscript.

Chapter 5 was the result of a collaboration among Sally Otto, Kermit Ritland and me. Sally assisted with experimental design, data analysis, and worked on the integral for the likelihood. In addition, she provided many extremely helpful and detailed comments on the manuscript. Kermit provided the resources and financial aid to carry out the field and molecular data collection. I designed the experiment, collected the field and genetic data with the help of several undergraduate assistants, analyzed the data, wrote the likelihood program in R, and was responsible for writing the manuscript.
Chapter 1 : INTRODUCTORY REMARKS

“All these inferences and formulae remain dry as dust until applied to the actual situation. What evidence is there concerning the total frequency of mutation, the relative frequencies of mutations with different grades of detriment, the amount to which they act as dominants or recessives, and finally, the degree and manner in which the population is encumbered by them?” H.J. Muller (1950)

In his 1950 work entitled “Our Load of Mutations,” H.J. Muller began to speculate on the frequency of mutations and their effect. Comparing the ratio of spontaneous recessive lethals to “visible detrimentals” to the same ratio induced by X-rays in *Drosophila melanogaster*, he surmised that small effect mutations were two or three times more frequent than mutations with larger effects. Estimating the number of genes in the *Drosophila* genome as 5,000-10,000 genes, he went on to suggest that 1 in every 20 gametes has a new mutation, while the per locus mutation rate was 1 in 100,000.

Since 1950, a vast amount of research has measured the genomic deleterious mutation rate in many different model and laboratory organisms, in part due to experimental methods established by Muller (1928) and developed further by Mukai (1964). In Mukai’s seminal experiments (1964), natural selection on a single *Drosophila melanogaster* chromosome was relaxed, thus allowing mutations to accumulate (mutation accumulation (MA) experiments). These mutations were found to have a measurable effect on fitness related traits such as viability. Mukai (1964) found that mutational effects caused a 0.4% decline in mean viability per generation. When extrapolated to the haploid genome, his minimum estimate of the genomic deleterious mutation rate, U, was 0.14 new mutations per generation. Although subsequent to Mukai’s work (1964) questions have been raised about the validity of his estimates (Fry, 1999; Keightley,
1996; Keightley and Eyre-Walker, 1999), it was still important as one of the first attempts to empirically assess mutational parameters.

Although the majority of mutations are thought to be harmful (Keightley and Eyre-Walker, 1999; Keightley and Lynch, 2003), recent evidence suggests that some proportion of newly arisen mutations have fitness enhancing effects (Joseph and Hall, 2004; MacKenzie et al., 2005; Shaw et al., 2000). Newly arisen mutations are likely to have a deleterious effect because they can occur in either protein coding sequences (Hoekstra and Coyne, 2007) or regulatory regions (Carroll, 2000) where they disrupt biochemical function, physiological networks, and/or alter gene expression. Perturbations in highly adapted developmental, physiological and biochemical networks lead to higher rates of mortality and reduced fecundity in some individuals within a population. Mutations that reduce fitness can have selective effects that range from almost neutral to lethal. The presence of these mutations in a population imposes a genetic load or a loss of fitness on the population (Haldane, 1937). Independent of a mutation’s selective effect, the number of “genetic deaths” in a population at a mutation-selection equilibrium is driven by the rate at which deleterious mutations occur over the whole genome, U (Haldane, 1937). Framing the discussion in terms of individual mutations and their dominance coefficients, Muller (1950) postulated that mutations with slight phenotypic effects could reach higher equilibrium frequencies. Thus, the total loss of fitness or the number of genetic deaths for slightly deleterious mutations can be equivalent to that of a lethal mutation, which causes the death of a single individual. He stated that

“…each detrimental mutant gene, no matter how slight its phenotypic effect, produces, on average, one eventual half-death of a zygote, or what may be termed one
genome-death, when it acts as a recessive in causing elimination, and one complete zygote death, i.e. two genome deaths when acting as a dominant.” MULLER (1947)

Knowledge about the frequency of mutation is fundamental to a wide range of evolutionary theory. For example knowledge about U may help to support or refute the mutational deterministic hypothesis (MDH) for why sex would be beneficial (Kondrashov, 1994). Under this hypothesis, if the genomic mutation rate per generation is greater than one then a substantial deleterious genetic load is imposed on the population. This is especially true in the presence of negative epistasis, where deleterious mutations reduce fitness more than expected by the individual fitness effects. Thus, sex is beneficial because recombination effectively purges the population of these deleterious mutations by bringing them together into a low fitness genotype.

Evidence for the assertion that most mutations reduce fitness come from MA experiments on a variety of organisms such as Caenorhabiditis elegans (Keightley and Caballero, 1997b), Drosophila (Fry, 1999), Escherichia coli (Kibota and Lynch, 1996), Amsinckia (Schoen, 2005), and Arabidopsis thaliana (Schultz, 1999). Collectively these experiments demonstrated significant declines in fitness of the MA line relative to the control. One criticism that has been raised about the analyses of these experiments is estimation of the mutation rate and effect size did not allow for the possibility of fitness-enhancing mutations (Shaw, 2002). Furthermore, recent work has demonstrated that approximately 6-50% of new mutations increase reproductive fitness (Joseph and Hall, 2004; MacKenzie et al., 2005; Shaw et al., 2000).

Our understanding of the deleterious mutation rate, U, is limited by a taxonomic bias. At present, we have direct estimates of the slightly deleterious mutation rate for
only two plants, *Amsinckia* (Johnston and Schoen, 1995; Schoen, 2005) and *Arabidopsis* (Schultz, 1999; Shaw *et al.*, 2000), both of which are herbaceous annuals. These studies suggest a range from 0.0024 to 2.89 per diploid genome per sexual generation, depending upon the trait under consideration. Lethal mutation rates, however, have been calculated on a broader taxonomic range using seedlings that fail to perform photosynthetic carbon fixation, i.e., chlorophyll deficient seedlings. These studies use the relationship between the per locus mutation rate, $\mu$, and the frequency of the mutant allele, $q$, at a mutation-selection equilibrium. When a mutant allele is completely recessive ($h=0$) and lethal ($s=1$), this relationship simplifies to $\hat{q} = \sqrt{\mu}$. A comparison of the lethal mutation rates in 10 annuals to three long-lived perennials suggests that long-lived perennials have rates that are ~10-20 times higher (Klekowski, 1998). This method cannot be broadly applied to outcrossed woody tree species because it requires populations to be at a mutation-selection equilibrium. It is unlikely that populations of outcrossed woody tree species are near this equilibrium because the number of generations in years required establishing a mutation-selection balance is much greater than the length of time that tree populations have occupied their current habitats. One review on tree range size shifts suggests that most temperate habitats were under the glaciers 10-12,000 years BP (Davis and Shaw, 2001). Furthermore, as Muller (1950) pointed out, lethals are unlikely to segregate at the same equilibrium frequencies as moderate to weak mutations in plant populations.

Although direct estimates of the genomic mutation rate are limited in plants, indirect estimates of the genetic load imposed by recessive deleterious mutations are available by way of estimates of the number of lethal equivalents. A lethal equivalent is a set of mutant alleles that if found in a single diploid individual, homozygous for all
mutant alleles, the probability of death would be 100% (Morton and Crow, 1956). Thus, unlike a recessive lethal allele, a lethal equivalent is the composite effect of many slightly deleterious alleles at many different loci (Klekowski, 1988). Estimated by examining relative offspring survivorship from selfed and outcrossed matings, those mutations removed during soft selection, i.e. pollen or ovule competition, are missed (Klekowski Jr, 1988). Estimates in woody tree species (*Pinus, Pseudotsuga*, and *Eucalyptus*) suggest the number of lethal equivalents is between 3-27 per zygote or 1.5-13.5 per gamete (Franklin, 1972; Griffin *et al.*, 1987; Sorensen, 1969). In contrast, the number of lethal equivalents is much smaller in some reported herbaceous plants: 0.79 per zygote in *Phlox drummondii* (Levin, 1984), 3.67 per zygote in *Allium schoenoprasum* (Stevens and Bougourd, 1988), 1.96 per gamete in *Fagopyrum esculentum* (Komaki, 1982), 2.7 per zygote in *Secale cereale* (Landes, 1939), and 1.2-4.5 per zygote in *Medicago sativa* (Cooper and Brink, 1940). Conversion of lethal equivalents into a genomic deleterious mutation rate requires species specific information on the dominance coefficient, h. Although there is some evidence for the slight recessivity of newly arisen mutations (in *Amsinckia*: h~0.28-0.35: Johnston and Schoen, 1995), there is little empirical work on the degree of dominance in a wide range of species (*Drosophila*: Garcia-Dorado and Caballero, 2000; *C.elegans*: Peters *et al.*, 2003; *Arabidopsis*: Shaw and Chang, 2006).

Despite almost two decades of research, we still have limited information on how often new mutations of small effect might arise. Therefore, in response to Muller’s call for, “evidence…concerning the total frequency of mutation [and] the relative frequencies of mutations with different grades of detriment,” one goal of my thesis was to obtain an estimate of the genomic deleterious mutation rate, U, in a natural population of trembling
aspen, *Populus tremuloides* (Michaux.). Unlike the lethal mutation rate obtained on red mangrove (Klekowski and Godfrey, 1989), this estimate of $U$ is for slightly deleterious mutations specific to sexual function. Secondly, in contrast to Klekowski and Godfrey (1989) who restrict their attention to lethal mutations, we estimate the distribution of mutational effects for *P. tremuloides*. Finally, we compare our estimate of the mutation rate to those obtained in annuals and ask how a given mutation rate and distribution of selective effects may influence “the degree and manner in which the population is encumbered by them.” (Muller, 1950)

The origin and accumulation of mutations is a time-dependent process. This fact means differences in development and growth patterns may lead to differences in mutation rates. One developmental feature differentiating animal and plants is the distinction between the germ line and the soma. Although cell divisions occur in the germ line creating the opportunity for mutations to arise and be passed on to offspring, this is not true of mutations that arise in the soma of animals. Somatic mutations may be expressed, but because of the separation between the two cell lines, these are not inherited by the next generation (Antolin and Strobeck, 1985; Buss, 1983). In plants, and particularly in clonal plants, there is no separation between germ and soma cells, and growth occurs largely through the iteration of structural units, e.g., leaves, branches, meristematic apices. Thus meristematic totipotency results in a direct incorporation of somatic variants into germ cells. This is particularly true for clonal plants where the production of mitotic offspring or ramets could confer “immortality” to the plant clone (Slatkin, 1985; Sutherland and Watkinson, 1986). Thus, mutations affecting survivorship or reproductive fitness can arise among ramets within a clone, and as a consequence,
genetically-based variation in fitness can arise within a single individual. This increased opportunity for mutation, however, is balanced by the increased potential for intra-organismal selection (Otto and Orive, 1995).

Plants may have higher mutation rates or perhaps a higher proportion of beneficial mutations to compensate for the local environmental conditions because of their inability to escape stressors (Whittle and Johnston, 2006). Recent evidence suggests that surface and soil-based agents such as irradiation, humidity, and UV and biotic agents such as pathogens may affect the occurrence of mutation and recombination in plants (Albrecht et al., 2004; Kovalchuk et al., 1998; Lebel et al., 1993; Lucht et al., 2002). Oxidative stress increases the number of mutations because free radicals and other chemicals modify ribonucleotides, and deoxyribonucleotides. Cells under high oxidative stress often have inefficient or altered repair mechanisms (Britt, 1999; Cromie et al., 2001).

Trembling aspen, a dioecious member of the willow family (*Salicaceae*), has a mixed reproductive system generating two kinds of progeny: ramets derived from asexual reproduction (mitotic growth) and genets derived from seeds. In the thesis, genet and clone are used interchangeably to refer to a collection of ramets. Although *P. tremuloides* reproduces sexually in all parts of its range, it is thought that the harsh semi-arid environmental conditions of the West have severely reduced the probability of seed germination and seedling survival. With the failure of sexual seedlings, *P. tremuloides* has persisted in the western part of its range through extensive clonal propagation via the production of root suckers (Barnes, 1966; Barnes, 1975; Mitton and Grant, 1996). Large aspen clones have been reported in central Alberta and in the Rocky Mountains. These clones occupy areas ranging from a tenth of an acre to over 100 acres and consist of a few
dozen trees to thousands of ramets (Kemperman and Barnes, 1976). The enormous size of some of these clones has led researchers to postulate that they are quite old. For example, in central Alberta, it is hypothesized that seedling populations of aspen were last established shortly after the retreat of the Laurentide and Cordilleran glaciers, 8000B.P (Alley, 1976).

Clonal growth has features similar to Mukai’s (1964) mutation accumulation (MA) experiment. First, similar to a mutation accumulation experiment, it is seeded by a single genotype, i.e., seed. As the clone grows and reproduces asexually, each mitotic offspring (ramet) represents essentially a replicate line. With every round of asexual reproduction, somatic mutations can and do accrue resulting in among ramet or line divergence creating mosaicism within a single individual (Di Rienzo et al., 1998; O’Connell and Ritland, 2004; for aspen clones see; Tuskan et al., 1996). Secondly, in plant clones although the relaxation of selection is not achieved by bottlenecking ramet populations, natural selection is relaxed for traits involved in sexual function during clonal growth. This is because sexual offspring do not contribute to clonal fitness, thus mutations that affect sexual function are effectively neutral in the face of selection at the clone level (Eckert, 2002). At the end of a typical MA experiment, because moderate to weak mutations accumulate and fix due to the action of drift, overall mean fitness declines and lines diverge phenotypically. In a population of clones, older clones are expected to have lower mean sexual fitness but higher variance than younger clones.

In Chapter 2, I examined three populations to determine if the ecology was appropriate to consider clonal aspen as a natural mutation accumulation experiment. Thus, in the second chapter I describe patch composition, in an attempt to assess the
relative magnitude of sexual to asexual recruitment. Unlike laboratory-based MA experiments where the duration of the experiment is known, clone age is not readily observed. Thus, in Chapter 3, I explore the use of clonal divergence assayed at neutral microsatellite loci and clone size as measures of clone age. In chapter 4, I assess how robust estimates of clonal age are to changes in demographic models of clonal growth. Lastly, in Chapter 5, I estimate the genomic deleterious mutation rate and the distribution of effect size from clones in a natural population of aspen using both the Bateman-Mukai and maximum likelihood methods. As a widespread phenomenon, clonality offers the potential to estimate mildly deleterious mutation rates in the wild in a wide range of organisms.
REFERENCES


Johnston, M. O., and D. J. Schoen, 1995 Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. Science 267: 226-228.


Chapter 2: Fine-scale genetic structuring of *Populus tremuloides* patches in three clonal populations.

*A version of this chapter will be submitted for publication.
Ally D, Ritland K, Otto SP. Fine-scale genetic structuring of *Populus tremuloides* patches in three clonal populations.*
INTRODUCTION

Most clonal plants exhibit a modular structure at multiple levels. At the level of the organ, they are characterized by functional module - discrete physiological units, such as leaves, flowers, and branches. At the level of the individual or genet, clonal plants have evolutionary modules capable of responding to selection because they possess heritable trait variation and provide the opportunity for intra-organismal selection to act (Otto and Orive, 1995). These evolutionary units arise through the widespread phenomenon of clonal reproduction. Asexual reproduction is generally accomplished in one of two ways: by vegetative propagation (e.g., lateral root suckers, rhizomes, stolons, bulbils, and vegetative spread) or through the formation of unfertilized seed (apomixis) (Silander, 1985). Organisms with a mixed reproductive system have the capacity to generate two kinds of progeny: mitotically-derived (ramets) and meiotically-derived (genets).

Rapid clonal spread, which increases the number of effective clonal propagules relative to sexual offspring, can result in altered patterns of population genetic structure. In the absence of mutation, extensive clonality produces offspring genetically identical to the parent plant, yielding lowered population (sensu panmictic population) level genotypic diversity, increased divergence among subpopulations (demes), and decreased effective population sizes (Orive, 1993). Limitations in resources and space coupled with the high levels of clonal offspring restrict the numbers of sexual offspring and lead to fewer genotypes in any one population. Among taxa with the capacity for both sexual and asexual reproduction, there is considerable variability in levels of genetic and genotypic diversity (Ellstrand and Roose, 1987a; Hamrick and Godt, 1989). This wide
variability detected in different taxa may imply multiple mechanisms operating to affect
genet turnover and may underlie differences in the rate of evolutionary change among
species (Poore and Fagerstrom, 2001).

Within populations of multiple genets of a single species, the spatial arrangement
of successful sexual and asexual offspring arises from biotic influences on establishment
and survival as well as effective dispersal distances of the respective reproductive forms.
For example, clones will be aggregated if vegetative spread is restricted relative to the
distance of seed dispersal. Unlike plumose seeds adapted for wind dispersal, vegetative
propagules often lack traits that facilitate long-distance dispersal. Intermingling of
different clones may occur if plants have runners or lateral roots that can spread up to
several meters in distance or roots that persist dormant in the soil over long periods of
time. Finally, interactions with ecological processes and factors such as gap dynamics,
plant understory composition, competition for pollinators, and soil moisture can impact
the success of seed germination and establishment (Aigner, 2004; Eckert and Barrett,
1993; Jacquemyn et al., 2005; Kudoh et al., 1999). In the clonal woodland herb *Uvularia
perfoliata*, closed canopy habitats contained single genotypes while the open canopy or
gap habitats had a small number of widely spaced and highly intermingled genets (Kudoh
et al., 1999).

Ecological factors impact the spatial arrangement of sexual and asexual offspring
relative to each other by affecting how ramets within a clone are arranged. There are
two types of internal organization of clones that represent extremes of a continuum:
phalanx and guerilla (Lovett Doust, 1981). In the slower spreading phalanx form, there
is a higher density of ramets with shorter and fewer connections between individual
modules. By contrast, in the guerrilla form, outward clonal expansion is fast, and ramets are widely spaced with longer connections between individual modules within a genet. In response to competitive stress and plant density, clonal plants exhibit plasticity in growth forms mediated by changes in resource allocation (Bishop and Davy, 1985; Fischer et al., 2004; Rautiainen et al., 2004; Van Kleunen and Fischer, 2003). For example, inter-specific competition in *Ranunculus reptans* resulted in a 36.9% increased spatial spread with a concomitant reduction in the proportion of flowering rosettes (Fischer et al., 2004). On the other hand, intra-specific competition inhibited clonal spread in *Potentilla anserina* ssp. *egedii* and promoted flower number (Rautiainen et al., 2004).

One interesting example of a species that exhibits both reproductive forms is *Populus tremuloides* or trembling aspen (Michaux) (Brown, 1935). Trembling aspen has the unique feature that its vegetative offspring can disperse at great distances. One example of this characteristic comes from the excavation of a root system where researchers found a single lateral root ran a distance of 31.7m (Buell and Buell, 1959). A dioecious member of the willow family (*Salicaceae*), *P. tremuloides* reproduces sexually in all parts of its range. Using morphological characters to distinguish between sexual offspring and asexual offspring researchers suggest that the harsh semi-arid environmental conditions of the West have severely reduced the probability of seed germination and seedling survival (Mitton and Grant, 1980). With the failure of sexual seedlings, *P. tremuloides* has persisted in the western part of its range through extensive clonal propagation via the production of root suckers (Barnes, 1966; Barnes, 1975; Mitton and Grant, 1996).
There have been many studies examining levels of genetic diversity in natural populations of aspen in different parts of its range including Alberta, Quebec, Ontario, Wyoming, Wisconsin, Minnesota, and Colorado (Hyun et al., 1987; Jelinski and Cheliak, 1992; Lund et al., 1992; Namroud et al., 2005a; Stevens et al., 1999; Wyman et al., 2003; Yeh et al., 1995). Collectively, these studies found high levels of genotypic diversity with very little differentiation among populations, suggestive of extensive sexual recruitment within populations. The suggestion of high levels of sexual recruitment stands in marked contrast to reports of western North American aspen clones occupying between 24.9 acres to 106.8 acres and containing 15,000 to 47,000 ramets respectively (Barnes, 1966; Barnes, 1969; Kemperman and Barnes, 1976).

One reason for the discrepancy is that collection strategies in all of these genetic diversity studies were designed to increase the number of genets sampled by excluding neighboring trees (under 1km). Such sampling designs ignore dynamics within local patches and do not provide insight into whether patches are formed from single or multiple genotypes. An exception is one recent study of a young aspen stand after fire and insect-mediated disturbances in northwestern Quebec (Namroud et al., 2005a; Namroud et al., 2005b). In this study, Namroud et al. (2005a) found a few, small multi-stemmed genets (an average of 3-4 ramets/genet) in older tree cohorts (Namroud et al., 2005a). Clonality in eastern North America, however, is not extensive, and favorable environmental conditions allow seedlings to establish more often than in the West (McDonough, 1985).

Information from smaller-scale geographical studies may provide contrasting results compared to data gathered from larger scales. The goal of this present study is to
provide a fine-scale quantitative description of the frequency and spatial patterning of clonal reproduction in three populations of *P. tremuloides*. To this end, we ask (1) what is the magnitude and spatial pattern of sexual recruitment within aspen patches (2) what is the internal organization of clones within a shared location? (3) what is the relationship structure within a stand and between stands (i.e., are genotypes that share patches full-sibs, half-sibs, or parent-offspring), and (4) can micro site variation in biotic and abiotic factors explain the level of genotypic diversity at a locale? The current study is the first fine-scale study that examines patch composition of aspen populations where clonality is thought to be prevalent.

**MATERIALS AND METHODS**

*Study Species*

*Populus tremuloides*, or trembling aspen, is a dioecious member of the willow family (Salicaceae) that reproduces both sexually and asexually (Perala, 1990). As a dioecious species, clones will consist of either staminate or pistillate flowers. Aspen is capable of asexual reproduction throughout its range by forming adventitious shoots (suckers) from lateral roots. Aspen roots consist of several primordia, which contain meristematic cells or potential apical meristems. Aspen ramets/suckers can be produced as early as age 1 yr although the suckers require high soil temperatures, low moisture, and full sunlight for survival (DesRochers, 2000; Keyser *et al.*, 2005; Schier, 1973). In the early stages of sucker differentiation suckers are suppressed, but disturbances such as fire or herbivory will release these suckers, and each one of them can grow into a distinct tree above-ground.
Study populations and sampling

Absent from previous genetic studies on *P. tremuloides* is information about the number of different genotypes within aspen stands and the spatial distribution of those genotypes. This requires populations where the relative rate of asexual recruitment versus sexual recruitment is high. Ecological factors as well as soil conditions in the western part of aspen’s range reduce seed production, seedling establishment, and survival. Thus we chose three populations in the western part of *P. tremuloides*’ range. Two of the three study sites were located in Waterton Lakes National Park (WLNP), situated in the Rocky Mountains of extreme southwestern Alberta adjacent to the Canada/US border (Figure 1-2). Waterton Lakes National Park was initially established in 1895 as 140 sq km of protected land and officially designated a park in 1932. The park’s oldest stands regenerated after 11 fires between 1633 and 1800 (Barrett, 1996). The first population we studied within the park was an aspen-dominated grassland site in the foothills between Bellevue Hill (BH) and Mount Galwey at an altitude of 1330m, while the second population was situated in a montane region, Red Rock Canyon (RR), at approximately 1900m in altitude. At the latter, aspen grows sparsely on moderate subalpine slopes with mainly southerly aspects and is subject to periodic snow avalanches.

The third population, near Riske Creek, was 40km west of Williams Lake in the Cariboo-Chilcotin grassland of south central British Columbia (Figure 2-1). Riske Creek has a mixed landscape with coniferous forest, aspen-dominated grasslands, and areas with small lakes and ponds. This area was established as rangeland for cattle grazing when it was settled in 1829 (Strang and Parminter, 1980). Fire frequency in the area
occurred at intervals of approximately 10 years between 1759 and 1926 but subsequently ceased due to fire suppression (Daigle, 1996; Strang and Parminter, 1980).

To examine the spatial distribution of sexual and asexual offspring within an aspen patch, we sampled tissue (buds, leaves or secondary phloem) in two ways: on the perimeter of a grove or patch of aspen and along two or three 30-50m transects within a patch. Systematic sampling was used (every 3-5m) as it would best represent the distribution of different genotypes found in that patch. On average, 30-50 individuals were sampled per patch. We sampled 17 patches in Riske Creek (RC), 10 at the base of Bellevue Hill (BH), and 10 in Red Rock (RR). In RC and BH patches were selected randomly from the landscape by dividing up the area into GPS identified squares and sampling squares out of a hat. GPS co-ordinates of the sampled squares were determined and the patch closest to the original co-ordinates was selected for sampling. In Riske Creek, a total of 27 patches were sampled. Of these 27, we selected 17 patches where the majority if not all ramets within the patch had staminate flowers. Only three patches (15, 23, and 25) exhibited ramets with a mixture of staminate and pistillate flowers. The ten patches in Bellevue Hill were randomly selected. Of the ten sampled, four were largely staminate, four were primarily pistillate, and two had a mixture of both staminate and pistillate ramets. In Red Rock we observed no pistillate ramets. No tree less than 1.5m high was sampled, and patches were separated by at least 1 km of terrain lacking aspen trees. Trees on the perimeter and along transects were physically mapped using both a measuring tape and a handheld GPS unit. From the GPS points we calculated the area of a patch using ARCVIEW GIS 3.3 (2002 Environmental Systems Research Insitute). In the Red Rock Canyon population, trees grew sparingly and were interspersed amongst
coniferous trees along a ridge. Every attempt was made to sample Red Rock in the same manner as in the grasslands, but in most cases all the trees in the designated patch was sampled, as the patches were generally small.

To examine possible ecological correlates of clonal reproduction we measured four environment factors for each patch in the different populations: density of ramets per patch, soil moisture regime (SMR), soil nutrient status (SNR), and drainage class (DRAINAGE). A single plot was randomly chosen in a patch. Within this plot, we measured density for each patch by counting all aspen trees and non-aspen trees in a 10m by 10m plot. Using a soil key on soil pits dug within a 1m x 1m quadrat in each patch, we obtained soil the following soil characteristics: SMR, SNR, and DRAINAGE. Soil moisture regime measures the average amount of soil water annually available over several years for evapo-transpiration and can range from driest (very xeric) to wettest (hydric) (Meidinger and Pojar, 1991). Soil nutrient regime is the relative estimate of essential soil nutrients available over several years and ranges from A (very poor) to E (very rich). The seven drainage classes describe soil texture, frequency of wet periods and how effectively soil retains water (Klinka et al. 1989). In addition we surveyed the plant understory composition. The plant community in a patch of aspen was used to provide indirect information on site characteristics (Klinka et al., 1989; Klinka et al., 1996). In a smaller 5m radius circle within each patch percent cover of trees, shrubs, herbaceous species, and moss was assayed.
DNA extraction, amplification and fragment visualization

We isolated DNA using a modified plant tissue protocol (Doyle and Doyle, 1990). Bud or leaf tissue (0.15-0.20g) or bark (0.20g) was homogenized with liquid nitrogen and incubated in 1800uL CTAB isolation buffer (2% CTAB, 1.4M NaCl, 20mM EDTA, 100mM TRIS-HCL pH=8.0, 1% 2-beta-mercaptoethanol) in a 65-70 °C water bath for 45 minutes followed by a chloroform-phenol wash. We used five microsatellite primers originally developed for *Populus trichocarpa* (Tuskan, 2004) to genotype each individual ramet. DNA was subjected to PCR in a 10ul reaction containing 1uL of 10X Buffer, 1.0 uL of 2mM dNTP, 1uL each of forward and reverse primers (1pmol/uL), 0.5 pmol of M13 IRD-labeled primer, and 1 Unit of Taq and 20-50ng of DNA. Samples were amplified with profiles specific for each primer pair, then loaded on a LI-COR 4200 automated sequencer for assay. Microsatellite products were detected by M13 tailed or end-labeled primers. DNA fragments were sized using the software analysis program SAGA (LICOR Inc., Lincoln, NEB) and LICOR size standards (cat. no. 4200-44), labeled with IRD dye.

Genetic Data Analyses

Definition of a genet

Microsatellite loci were used to identify clonemates (ramets sharing the same set of alleles across all five loci) from individuals who differed by more than one locus and were thus inferred to descend from different seeds. When we found a ramet that differed by one allele at one locus but at all other loci shared the same alleles as the most frequent genotype, we considered the allele a potential somatic mutation (see Chapter 3). We
excluded somatic mutants from analyses of diversity. Including somatic changes would cause us to overestimate genetic diversity and genotypic diversity and underestimate the amount of clonal reproduction.

To evaluate the probability that ramets of the same multi-locus genotype belong to the same genet, we modified Aspinwall and Christian (1992) to include the probability that ramets belonging to the same genet can differ by a single allele. Thus, the total probability that ramets sharing alleles belong to the same genet is

\[
\bar{P} = 1 - \left[ \sum_{Q=1}^{N} \prod_{D=1}^{M} \left( \frac{X_{DQ}}{P_D} \right) + \sum_{Q=1}^{N} \sum_{L=1}^{M} \frac{\Delta_{LQ}}{P_L} \frac{\prod_{D=L}^{M} X_{DQ}}{P_D} \right],
\]

where \( Q \) is an individual in the sample, \( N \) is the number of individuals in a population, \( M \) is the number of polymorphic loci in a population, \( X_{DQ} \) is the number of individuals sharing a set of alleles at locus \( D \) as the reference individual \( Q \), and \( P_D \) is the total number of individuals examined for each polymorphic locus. The first part of the equation estimates the probability that an individual is genetically identical at all loci to the focal individual and yet descended from a different seed (i.e., not a clonemate). The second part of the equation considers the possibility that ramets within a clone may match all but a single allele and still belong to the same genet. The term \( \frac{\Delta_{LQ}}{P_L} \) represents the number of individuals differing by one allele at the \( L^{th} \) locus given the total number of individuals examined for \( D^{th} \) locus, \( P_D \). This term is multiplied by the probability that all
other loci are identical, \( \prod_{D=1}^{M} \frac{X_{1D}}{P_L} \). The probability that ramets sharing the same multi-locus genotype belong to the same genet increases with the number of polymorphic loci (Aspinwall and Christian, 1992). Discriminatory power is the highest when most loci are polymorphic.

We obtained frequency distributions of allele frequencies, estimated expected heterozygosity, and performed tests for Hardy Weinberg equilibrium and linkage disequilibrium using GENEPOP v3.4 (Raymond and Rousset, 1995). These standard genetic measures allowed us to obtain an overall genetic description of our three populations. Linkage disequilibrium tests were used to predict if loci were independent. Finally, we performed a contingency table analysis and goodness of fit test to determine if allele frequency distributions were significantly different among our sites. Specifically, we were interested in whether the gene pools of RR and BH were the same or different because these two sites were only separated by a distance of approximately 8-9km. If the allele frequency distributions were significantly different, then we would be confident in calling these two sites different populations. Prior to performing our general genetic analyses, we reduced the dataset to include a single representative of each genotype because clonal reproduction would cause us to consider some genets more than once.

**Internal organization of ramets within genets**

Example plots of genotypes in physical space from each of the three populations can be found in Figure 2-2. If clones followed a phalanx arrangement then the distance between clonemates would be smaller than the distance among different genotypes.
(Burke, 2000). If clones are fully intermingled (guerrilla) then on average there would be no difference between the distances of clonemates versus non-clonemates. To statistically test whether ramets of a clone were aggregated within a patch, we calculated the average pairwise distance among clonemates and compared this to the average pairwise distance between different genotypes.

**Estimation of the magnitude of sexual recruitment within a patch**

We estimated the relative success of sexual vs. asexual recruitment in each of our three populations using the probability of clonal identity $F(r)$ described by the following equation:

$$ F(r) = k \frac{e^{-cr}}{\sqrt{r}} \quad (2) \quad (\text{Harada et al., 1997}) $$

where $r$ is the distance between a pair of ramets within a clone, $k$ is a proportionality constant and $c = \sqrt{\frac{2u}{v\sigma^2}}$ is the rate at which clonal identity declines with distance $r$. The rate parameter, $c$, is a function of the relative success of sexual vs. vegetative reproduction $(u/v)$ and the mean square parent offspring distance in successful vegetative propagation, $\sigma^2$ (Harada et al., 1997). We found an estimate of 4.5m for parent-offspring distance via vegetative dispersal from the literature (Barnes, 1966). Using a likelihood analysis in R version 2.4.1. (The R Foundation for Statistical Computing, 2007), we coupled spatial data (physical mapping of trees) with the identification of genotypes to obtain estimates of $c$ and $k$. Several different starting parameter values were used in the likelihood analysis.
We calculated a number of standard genotypic diversity measures including: the number of distinct multi-locus genotypes observed in the $P^{th}$ patch ($MLG_P$), genotypic richness ($R_P$), genotypic diversity ($D_{obs}$), and a measure of evenness ($E$). Genotypic richness was measured as $R_P = (MLG_P - 1)/(n_p - 1)$, where $n_p$ is the number of ramets sampled for patch, $P$. When all sampled ramets ($n_p$) have a unique genotype then the value of $R_P = 1$, while $R_P = 0$ if all ramets sampled are the same genotype. Genotypic diversity or the probability of sampling two different ramets in a patch is

$$D = 1 - \frac{\sum_{i=1}^{MLG_P} n_i(n_i - 1)}{n_p(n_p - 1)}$$

where $n_i$ is the number of ramets sampled for the $i^{th}$ genotype in the $P^{th}$ patch. Higher values of $D$ correspond to greater levels of genotypic diversity within a patch. The final standard measure used was genotypic evenness, $E = (D_{obs} - D_{min})/(D_{max} - D_{min})$, where $D_{min} = ([MLG_P - 1](2n_p - MLG_P))/[n_p(n_p - 1)]$ and $D_{max} = [n_p(MLG_P - 1)]/[MLG_P(n_p - 1)]$. Evenness ($E$) ranged from 0 for a patch dominated by a single genotype to 1 where all genotypes are equally abundant.

The spatial distribution of sexual vs. asexual offspring

Previous work suggests that sucker recruitment and survivorship occurs at the periphery of a patch where typically there is substantial light, higher soil surface temperatures, and a lack of competition (Barnes, 1966). In contrast, more recent research showed that in response to severe fires, there has been widespread establishment of aspen seedlings into opened clearings (Romme et al., 2005). Thus, it remains unclear in the
case of random mating whether sexual recruitment will occur disproportionately at the
periphery of a patch. To determine where most of the sexual recruitment occurred, we
performed a paired t-test across different patches within each of the two populations (RC
and BH) on the proportion of multi-locus genotypes at the edge of an aspen patch versus
within. Because aspen grew along ridges in small clumps in the montane habitat of RR,
it was not included in this specific analysis but it was included in all other analyses. In
both of the grassland habitats RC and BH, *P. tremuloides* formed dense polygon shaped
patches, which grew in clearly defined patches or groves across the landscape with the
occasional contiguous forested areas.

**Estimation of relationship level among genotypes**

We estimated the likelihood of a particular relationship between pairs of different
clonal genotypes within a patch following Ritland (2000). Comparing two individuals,
there are two pairs of alleles. One set of alleles is derived from female parents, the other
pair from male parents. Thus, for a single microsatellite locus, the probability of a
pairwise relationship between two individuals with the genotypes \((A_iA_j)\) and \((A_kA_l)\) is:

\[
Pr(A_iA_j, A_kA_l) = p_i p_j (2 - \delta_i) (2 - \delta_j) [A_k p_1 + A_k \left( \frac{(\delta_k + \delta_l)p_1 + (\delta_k + \delta_l)p_1}{4} \right) + \Delta_i \left( \frac{\delta_k \delta_l + \delta_i \delta_l}{2} \right)]
\]

(Ritland, 2000) where \(\Delta_i\) is the probability both alleles in the first individual are identical
by descent to alleles in the second individual. For parent-offspring, full-sibs, half-sibs
and first cousins this is 0, \(\frac{1}{4}\), 0, 0 respectively. \(\Delta_k\) is the probability that one pair of alleles
is identical by descent between individuals and the second pair is not. For parent-
offspring, full-sibs, half-sibs and first cousins this is 1, \(\frac{1}{2}\), \(\frac{1}{2}\), \(\frac{1}{4}\) respectively. Lastly, the
probability of no identity by descent at any of the two alleles in the first individual with alleles in the second, \( \Delta_q \), is 0, \( \frac{1}{4} \), \( \frac{1}{2} \), \( \frac{3}{4} \) for parent-offspring, full-sibs, half-sibs and first cousins. To determine the most likely relationship level, we calculated the log-likelihood for each of the five different relationships within and between patches in the three populations, assuming all non-clonemates had the same relationship.

The relationship of micro site variation and genotypic diversity

In order to assess whether micro site differences account for the variation in genotypic diversity across patches within a population, we performed a multiple regression analysis. Our predictor variables included: density of a patch, SMR, SNR, DRAINAGE and plant understory data. The plant understory data consisted of percent cover for >40 different species. To reduce the number of variables considered in this dataset, we performed three separate PCA on the understory data for each of the three populations. The first three principal components from each population’s dataset was then used in the multiple regression analysis. Despite the fact that RR and BH were in the same locale, WLNP, they represented very different habitats (montane vs. grassland). Additionally, genetic differentiation tests found that their allele distributions differed significantly, and so RR and BH could be considered different populations (see Results). No differences in genotypic diversity was found among the three populations (see Results), thus the data was pooled. To account for potential regional differences in environment as there was evidence of population differentiation, we included population by principal component interactions (e.g., PC1xPop) as variables in the multiple regression analysis.
RESULTS

All five loci we assayed were highly polymorphic. This high level of polymorphism led to very high probabilities that ramets of the same multi-locus genotype were members of the same clone. This probability was greater than 0.90 for all three populations, thus we are confident that all ramets sharing a multilocus genotype constitute a genet. In a majority of cases, genotypically distinct ramets differed at most, if not all, of the loci.

Using five microsatellites, we genotyped a total of 612 trees in BH, 691 trees in Riske Creek, and 143 in Red Rock. We found 126 multilocus genotypes across 17 patches covering 105,414sq km of Riske Creek grassland (RC), 141 multilocus genotypes sampled in 10 patches across 4,155sq km grassland at the base of Bellevue Hill (BH) in Waterton Lakes National Park. In the montane environment of Red Rock, Waterton Park (RR), 32 multilocus genotypes were found in 10 patches covering 1,095sq km.

Geometry of clones

In RC and BH the aspen occurred in large distinct patches with few or no aspen trees between patches. We grouped the clones into two groups: multi-stemmed clones, clones with more than one sampled ramet and single-stemmed clones, clones with a single sampled ramet. The average number of sampled ramets per multi-stemmed genet was 11 in RC, 12 in BH, and 7 in RR. The ratio of the number of genets with multiple ramets to single-ramet genets found in RC was 0.800, in the grassland of BH 0.478 and in RR 1.133. If the number of multi-stemmed genets to single-stemmed genets was equal, this ratio would be one. If on the other hand, most genets sampled had on average more
than a single sampled ramet, this ratio would be greater than one. In BH and RC, we found that on average there were a higher number of small-sized clones within a population, while in RR it was this ratio was approximately one. Despite the prevalence of small-sized clones within a population, we found that within a patch a single genet often dominates as shown in Figure 2-3.

We performed a one-sided paired t-test (Ha: $\mu_{\text{same}} - \mu_{\text{dif}} < 0$) across patches in BH and RC comparing the average distance in meters between clonemates to the average distance between two genets. RR was not considered because not all clones were defined patches. In RC, there were six patches where a single genotype was found, while in BH only one patch was found with a single genotype. These patches were excluded from the analysis. Across 12 patches in RC, the average distance between ramets of the same genet was 53.1m (S.E.=11.3), while between different genets it was 260.8m (S.E.=154.1m). Across 9 patches in BH, the average distance between ramets of the same genet was 22.5m (S.E.=5.27m) and between genets it was 31.14m (S.E. = 4.55m). In BH, we found pairs of clonemates were closer together on average than pairs of genets ($t = -2.8506$, df = 8, p-value = 0.01073), while in RC this was not the case ($t = -1.3296$, df = 11, p-value = 0.1066). A paired-sampled t-test assumes the differences come from a normally distributed population of differences (Zar, 1999). We found across patches the differences between clonemate distance and non-clonemate distance in RC did not follow a normal distribution ($W = 0.4687$, p-value = 6.386x10^{-7}). Thus, we employed a nonparametric Wilcoxon paired-sample test on the RC data. Using a non-parametric approach, we found on average clonemates were closer together than non-clonemates ($V = 6$, p-value = 0.006836). Although this suggests that in both of our populations a
phalanx form is more common, we did observe some intermingling of clones in RC (Figure 2-2).

The allele frequency distributions for five different SSR loci are presented in Appendix A-2 (estimating each genotype only once). A contingency table analysis and exact test performed manually by binning alleles showed that for each pair of populations for all loci, both the genotypic and allelic frequency distributions are significantly different, suggesting different gene pools for each of the three different populations. Thus we considered BH and RR separate populations for the rest of the analysis. For each population, basic genetic diversity estimates were calculated, including number of multi-locus genotypes, and expected heterozygosity. Overall, the global test across populations in GENEPOP found no significant associations, thus we treated the five loci as independent for the remaining analyses.

At the patch level, we found RC had the lowest average genotypic richness ($R_p=0.13$, s.e.$=0.036$) and diversity ($D_{obs}=0.43$, s.e.$=0.076$). The range of genotypic richness in this population was $R_p=0.00-0.51$ and the range of genotypic diversity was $D_{obs}=0.00-0.94$. By comparison, BH and RR were more similar in their levels of genotypic richness (BH: $R_p=0.20 \pm 0.056$, RR: $R_p=0.32 \pm 0.12$) and genotypic diversity (BH: $D_{obs}=0.52 \pm 0.11$, RR: $D_{obs}=0.57 \pm 0.11$). The range in genotypic richness for BH was 0.00-0.49 and for RR was 0.00-0.29. The range in genotypic diversity for BH was 0.00-0.96 and for RR was 0.00-0.77. An ANOVA, however, found no significant differences among populations in genotypic richness ($F_{2,34}=1.932$, p-value$=0.160$) or in genotypic diversity ($F_{2,34}=0.619$, p-value$=0.544$). All populations had similar levels of
evenness \( (\bar{E} \approx 0.4) \) (see Table 2-1). No patch within a population had an evenness score of one where every individual is represented by equal numbers of ramets. Genotypic diversity measures for the patch level are provided in Appendix B-2.

*Estimation of the magnitude and spatial pattern of sexual recruitment.* Based on the decline in clonal identity with distance, the relative success of sexual to asexual recruitment \( (u/v) \) was small in RC \( (0.065 \pm 0.083) \) but slightly larger in BH \( (0.245 \pm 0.226) \). The range of values for \( u/v \) in RC was \( 3.08 \times 10^{-7} \) to 14.8, while in BH \( u/v \) ranged from \( 2.32 \times 10^{-10} \) to 2.14. Estimates of \( c \) could not be obtained on patches where a single genotype dominated because in these cases there was no detectable sexual recruitment (Appendix B-2: Table 1-B-2 and 3-B-2). In the montane habitat of Red Rock, Waterton National Park, aspen grew along ridges in small clumps, thus we did not estimate relative success of sexual to asexual offspring.

In both grassland habitats, we performed a paired t-test on the proportion of the total number of multi-locus genotypes in the patch found at the edge versus within a patch, the results are shown in Figure 2-4. The number of multi-locus genotypes per ramet sampled was significantly higher on the perimeter of a patch than in the centre of the patch (BH: \( t = 2.441, \text{df}=8, p\text{-value}=0.040 \), RC: \( t = 2.283, \text{df}=11, p\text{-value}=0.043 \)). Thus, seeds either arise more often at the edge of a patch, or are more likely to establish there, or both.
Estimation of Relationships

For all pairwise comparisons of multi-locus genotypes within a population, we obtained probabilities for five different relationship types. These pairwise comparisons were subdivided into two categories: within and between patches. Total log likelihoods for each type of relationship were obtained by summing the log-transformed probabilities for each pairwise comparison. Across all three of our populations, we found less than five comparisons where parent-offspring relationships had probabilities greater than zero. This may be because we sampled patches where most if not all ramets had staminate flowers making parent-offspring relationships less likely. Within these predominately male patches it is entirely possible that female genotypes were recruited into the patch, yielding some parent-progeny pairs, however, this relationship type was excluded from further analysis.

Table 2-2 shows the individual log-likelihoods for the different relationship types in all three populations. The individual log-likelihoods are not meaningful and the models are not nested, thus we converted the log-likelihoods into AIC and examined the Delta AIC (\( \Delta_i = AIC_i - AIC_{min} \)) (Burnham and Anderson, 2002). Within and between patches in RC, pairs of genotypes were likely unrelated, as all other relationships had a \( \Delta_i > 10 \) (Table 2-2). In contrast, genotypes sharing a patch were inferred to be related at the level of first cousins in RR and BH, but unrelated between patches.

The relationship of microsite variation and genotypic diversity

We performed three separate PCAs on the plant composition data for each of the three populations. Only the first three principal components from each population were
kept. From RC, the first three principal components accounted for 37.5% of the variation in the original data, in RR the three axes captured 80.1% of the original variation, and in BH 55.6% of the variation was accounted for by the first three components. For a description of these principal components see Chapter 3 (Results). Although among populations there were no significant differences in diversity measures, there was substantial variation across patches within a population for genotypic richness and diversity (Appendix B-2).

Ecological studies of seedling establishment and survivorship have found germination and emergence of seedlings are correlated to measures of soil moisture, soil temperature, nutrient status, competition, and the presence of certain plant understory species (McDonough, 1985; Erwin et al., 2001; Romme et al., 2005). Thus, we were interested in determining whether patch environment within a population might explain the level of variation found in genotypic diversity measures.

As no differences in genotypic diversity were found among populations, we pooled the data from the different populations including population/region as a main effect. We found no significant explanation of the variation in either genotypic diversity or genotypic richness by any of the three principal components extracted from the plant understory data or by density in a patch, soil moisture, drainage, and nutrient status. In summary, the nonrandom distribution of genets in the population was not explained by any of the microsite habitat variables we measured.

We also found no relationship between variation in clonal spread with density of aspen in a patch \((F_{3,15}=1.565, \ p\text{-value}= 0.2392, \ R^2_{\text{adj}} = 0.0861)\). Clonal spread was
measured as the ratio of the average distance in meters between sexual offspring and average distance between clonemates among the sampled individuals in a patch.

**DISCUSSION**

Local forest patches in western populations of *P. tremuloides* consist largely of multiple genotypes suggestive of high levels of sexual recruitment. Nevertheless, most patches were dominated by a single genotype, and in one population we found several patches (five out of 17) consisting of a single genotype. Clones consisted of a mixture of both guerilla (i.e., intermingling of clones) and phalanx (i.e., clumped or aggregated ramets) growth forms, though the phalanx form predominated. When seedlings were recruited into the populations, successful sexual establishment occurred largely on the edges of a patch. Among patches, genets were unrelated although in two of the populations when two genotypes shared a patch, the most likely relationship was first cousins. Some aspen patches sampled in this study were largely male, thus this level of relationship within a patch is unlikely to compromise population mean fitness via mating among relatives.

Across 21 clonal plant species, Ellstrand and Roose (1987) found an average genotypic diversity of 0.62 (range of $D=0.00-0.98$) and 0.68 for evenness (range of $E=0.00-0.97$). In addition, they reported a study of *P. tremuloides* where $D$ and $E$ were both one, indicating each individual sampled was a different genotype (Cheliak and Dancik, 1982). These high levels were corroborated by other genetic studies of clonal diversity in aspen (Hyun, 1987; Jelinski and Cheliak, 1992; Lund, 1992; Mitton, 1996; Tuskan, 1996; Stevens et al., 1999; Namroud et al., 2005). Some of these studies,
however, were conducted in the east, where sexual recruitment is high. By contrast, despite the fact that we found multi-clonal populations, the populations we surveyed had lower genotypic richness (range of $R_F=0.13-0.32$), diversity (range of $D=0.43-0.57$), and evenness ($E=0.41$) (Table 2-1).

Furthermore, estimates of the relative success of sexual to asexual offspring ($u/v$) are consistent with a much higher level of asexual recruitment than previously reported. In Riske Creek, this estimate was $0.065$ (S.E. $±0.083$) and in Bellevue Hill, Lake Waterton National Park it was $0.245$ (S.E. $±0.226$). This suggests that for each seed successfully recruited into Riske Creek, $\sim 15$ suckers are produced, while in Bellevue Hill this number is approximately four. Accurate estimates of the relative success of sexual versus clonal reproduction are contingent on the validity of the assumptions of Harada et al. (1997). This method assumes that the sampling scale is not smaller than a minimum vegetative dispersal distance, that local adaptation doesn’t explain the spatial distribution of clones, and that clones exhibit a clumped distribution. Finally, this equation also assumes that seed dispersal is random over the scale of the patch. Some clones in RC were intermingled and sexual offspring were not randomly dispersed over the patch, leading us to overestimate of $c$ and consequently $u/v$. Estimates are also dependent on accurate values of the mean-square parent-offspring vegetative dispersal distance ($\sigma^2$), which is likely to vary across habitats.

Eriksson (1989) described two possible models of genet dynamics which may apply to populations of aspen clones at different successional stages. Under Initial Seedling Recruitment (ISR), there is an initial establishment of a large cohort of seeds followed by a substantial loss of poorly adapted genets. After a period of time, surviving
genets increase in size through extensive clonal reproduction and come to dominate local populations. In this seedling recruitment model, there is no further seedling establishment after the initial colonization. This pattern has been clearly demonstrated by studies of aspen following extensive fires where there is widespread seedling establishment (Keyser, 2005; Romme et al., 2005). While mortality was high among the initial cohort of seedlings (~25%), it declined such that no losses were reported 11 years after the fire (Romme et al., 2005).

In the alternative model Repeated Seedling Recruitment (RSR) genets are continuously recruited into the local population. Under ISR, no differences between the edge and internal patch genotypic diversity are expected. Ongoing sexual recruitment largely at the edges of a patch, as was found in this study, is more consistent with RSR. Both regions where we sampled experience very little in the way of severe fires. In Riske Creek, fire frequency in the area occurred at intervals of approximately 10 years between 1759 and 1926 but subsequently ceased due to fire suppression (Daigle, 1996; Strang and Parminter, 1980). In Waterton Lakes National Park fires were frequent between 1633 and 1940 but efficient fire exclusion reduced the frequency substantially (Barrett, 1996). Thus, our populations may be representative of long-term trends in the demographic and genetic structure of aspen populations.

Many studies of clonal plants assume that patches consist of a single genotype (Ellstrand and Roose, 1987b and references therein). This, however, underestimates the amount of genetic diversity present at small geographic scales. As we have shown estimates of genotypic diversity and levels of clonal reproduction are highly sensitive to both the scale and intensity of sampling. Furthermore, the non-random spatial
distribution of clonal versus sexual offspring implies that the observed levels of
genotypic diversity will depend on the sampling strategy over space. Elucidating fine-
scale genetic structure in clonal plant populations requires a recognition that patterns
emerge from the collective assembly of the smaller scale structural units.
Figure 2-1. (Colour plate) A map showing the location of the three *P. tremuloides* populations. Below the map, enlarged areas showing the location of the patches within the populations. Patch locations are in latitudes and longitudes.
Figure 2-2. Examples of patches from three populations of *P. tremuloides* illustrate the presence of multiple genets (different symbol shapes represent different genotypes) in *P. tremuloides*. 

a. In Riske Creek (RC) a total of 691 ramets were sampled. 

b. In Bellevue Hill, the grassland habitat of Waterton Lakes, we sampled 612 ramets, and 

c. In Red Rock, the montane environment of Waterton Lakes, we sampled 143 ramets.
**Figure 2-3.** A single genotype predominates in a patch. From each of the three populations, we provide examples of a patch. The different multi-locus genotypes are simply ordered alphabetically. The y-axis gives an estimate of the number of ramets that would be found in that patch if all ramets of a patch were sampled. The number of ramets was estimated using area, density, and the proportion of the sampled ramets represented by a genotype. Each of the single bars represents a different multi-locus genotype coded in uppercase letters.
**Figure 2-4.** The number of multi-locus genotypes per ramet sampled was significantly higher on the perimeter of a patch than in the centre of the patch (BH: $t=2.441$, df=8, p-value=0.040, RC: $t=2.283$, df=11, p-value=0.043). Of the total number of multi-locus genotypes (MLG) found in a patch, we compared the average number found on the edge compared to the average number found on the inside of a patch for BH and RC. The total number of MLGs sum to a value greater than one because in some patches the same MLG was found at the perimeter and inside a patch. The bars represent the standard error around the mean.
Table 2-1. Estimates of clonal and genetic diversity across five microsatellite loci in three populations of *P.tremuloides* are presented. The number of ramets (*N*), the number of distinct multi-locus genotypes (*MLG*), average number of alleles observed (*A*) are given. Genotypic richness (*R_p*), genotypic diversity (*D_obs*), and evenness (*E*) are averaged across patches in each population. We also provide average expected heterozygosity (*H_e*) for each of the populations.

<table>
<thead>
<tr>
<th>Population</th>
<th><em>N</em></th>
<th>MLG</th>
<th><em>A</em></th>
<th><em>R_p</em></th>
<th><em>D_obs</em></th>
<th><em>E</em></th>
<th><em>H_e</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Riske Creek (RC)</td>
<td>691</td>
<td>126</td>
<td>19.8</td>
<td>0.13</td>
<td>0.43</td>
<td>0.41</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.036)</td>
<td>(0.076)</td>
<td>(0.083)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Red Rock (RR)</td>
<td>143</td>
<td>32</td>
<td>12.0</td>
<td>0.32</td>
<td>0.57</td>
<td>0.40</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.13)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Bellevue Hill (BH)</td>
<td>612</td>
<td>141</td>
<td>21.5</td>
<td>0.20</td>
<td>0.52</td>
<td>0.40</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0560)</td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.0003)</td>
</tr>
</tbody>
</table>

S.E. are shown in parentheses.
Table 2-2. The log-likelihoods for the different relationship types in all three populations are presented. The most likely relationship is italicized. For example, genotypes within patches at Riske Creek (RC) are inferred from the log-likelihood values to be unrelated. In contrast, genets within a patch are likely to be first cousins in Red Rock (RR) and in the grassland habitat of Bellevue Hill (BH). In all three populations multi-locus genotypes between patches are largely unrelated. We examined how likely other relationship types were by calculating Akaike criterion (\(AIC_c\)) and determining the Delta AIC (\(\Delta AIC = AIC_i - AIC_{min}\)). Values of \(\Delta AIC\) greater than 10 indicate unlikely models.

<table>
<thead>
<tr>
<th>Relationship Level</th>
<th>LogL (RC)</th>
<th>AIC(_c) (RC)</th>
<th>LogL (RR)</th>
<th>AIC(_c) (RR)</th>
<th>LogL (BH)</th>
<th>AIC(_c) (BH)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Patches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-Sibs</td>
<td>-284.4</td>
<td>578.8</td>
<td>-41.77</td>
<td>93.6</td>
<td>-4359.6</td>
<td>8729.1</td>
</tr>
<tr>
<td>Half-Sibs</td>
<td>-250.9</td>
<td>511.8</td>
<td>-39.15</td>
<td>88.3</td>
<td>-2982.1</td>
<td>5974.2</td>
</tr>
<tr>
<td>First Cousins</td>
<td>-209.6</td>
<td>429.3</td>
<td>-25.58</td>
<td>61.16</td>
<td>-2417.6</td>
<td>4845.2</td>
</tr>
<tr>
<td>Unrelated</td>
<td>-175.4</td>
<td>360.7</td>
<td>-32.75</td>
<td>75.5</td>
<td>-2691.9</td>
<td>5393.8</td>
</tr>
<tr>
<td><strong>Between Patches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-Sibs</td>
<td>-3.19x10^4</td>
<td>6.4x10^4</td>
<td>-1751.7</td>
<td>3513.4</td>
<td>-3.07x10^4</td>
<td>6.1x10^4</td>
</tr>
<tr>
<td>Half-Sibs</td>
<td>-1.58x10^4</td>
<td>3.2x10^4</td>
<td>-879.9</td>
<td>1769.8</td>
<td>-1.65x10^4</td>
<td>3.3x10^4</td>
</tr>
<tr>
<td>First Cousins</td>
<td>-8842.6</td>
<td>1.8x10^4</td>
<td>-577.2</td>
<td>1164.4</td>
<td>-9975.0</td>
<td>2.0x10^4</td>
</tr>
<tr>
<td>Unrelated</td>
<td>-5572.3</td>
<td>1.1x10^4</td>
<td>-563.8</td>
<td>1137.6</td>
<td>-6535.2</td>
<td>1.3x10^4</td>
</tr>
</tbody>
</table>
REFERENCES


46


Investigation of the bottleneck leading to the domestication of maize. PNAS 95: 4441-4446.
Fernando, D. D., and D. Cass, 1997 Developmental Assessment of Sexual Reproduction
selection on growth form and its plasticity in a clonal plant. Journal of
Evolutionary Biology 17: 331-341.
Fisher, M. J., 1928 The Morphology and Anatomy of the Flowers of the Salicaceae II.
Frey, B. R., V. J. Lieffers, Landh, auml, S. M. usser et al., 2003 An analysis of sucker
patterns of aspen dieback: mechanisms and knowledge gaps. Canadian Journal of
Forest Research 34: 1379-1390.
Frumkin, D., A. Wasserstrom, S. Kaplan, U. Feige and E. Shapiro, 2005 Genomic
Fry, J. D., Keightley, P. D., Heinsohn, S. L. and Nuzhdin, S. V., 1999 New estimates of
the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. 
PNAS 96: 574-579.
Fry, J. D. H., Stefanie L., 2002 Environment dependence of mutational parameters for
Gabriel, W., and R. Burger, 2000 Fixation of clonal lineages under Muller's ratchet.
Evolution 54: 1116-1125.
Garcia-Dorado, A., and A. Caballero, 2000 On the Average Coefficient of Dominance of
Gardner, S. N., and M. Mangel, 1997 When Can a Clonal Organism Escape Senescence?
Gill, D. E., L. Chao, S. L. Perkins and J. B. Wolf, 1995a Genetic Mosaicism in Plants and
Gill, D. E., L. Chao, S. L. Perkins and J. B. Wolf, 1995b Genetic Mosaicism in Plants and


Johnston, M. O., and D. J. Schoen, 1995 Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. Science 267: 226-228.


Swoboda, P., B. Hohn and S. Gal, 1993 Somatic homologous recombination in planta: The recombination frequency is dependent on the allelic state of recombining sequences and may be influenced by genomic position effects. Molecular and General Genetics MGG 237: 33-40.


Udupa, S., and M. Baum, 2001 High mutation rate and mutational bias at (TAA)n microsatellite loci in chickpea (Cicer arietinum L.). Molecular Genetics and Genomics 265: 1097-1103.


Chapter 3: CAN CLONE SIZE SERVE AS A PROXY FOR CLONE AGE? AN EXPLORATION USING COALESCENCE THEORY*.

*A version of this manuscript will be submitted for publication.
Ally D, Ritland K, and Otto, SP. Can clone size serve as a proxy for clone age? An exploration using coalescent theory
The definition of life span, while trivial in unitary organisms, is more complicated in long-lived clonal plants. The death of an individual is, at least in theory, escapable because plant clones can continually produce new physiological modules (ramets) by asexual processes. (Gardner and Mangel, 1997; Orive, 1995). This makes clone age difficult to estimate. To circumvent this issue, previous estimates of clone age in plants couple the mean growth rate per year of shoots with the area covered by the clone (Escaravage et al., 1998; Reusch, 1998; Steinger, 1995; Tyson, 1998; Vasek, 1980).

Analogous to tree-ring dating, this method assumes the area covered by a clone is a perfect circle (phalanx growth) and that area ($\pi r^2$) increases with the square of time. Using this method, one clonal plant, a Creosote bush, *Larrea tridentata*, was estimated to be 11,700 years old (Vasek, 1980). Growth in a perfect circle, however, is only one pattern of expansion that plant clones exhibit. Alternative geometrical arrangements such as, the intermingling of different clones, termed guerrilla growth, also occur (Ye et al., 2006). The geometric arrangement of clones ultimately depends on the relative pattern and rate of ramet recruitment, establishment probability, and mortality (Eriksson, 1989).

Moreover, clonal architecture is affected by the ecological conditions under which asexual reproduction occurs (Klimeš and Klinešová, 1999). These ecological conditions include spatial and temporal heterogeneity in resources, density and spacing of ramets, and whether among-ramet interactions are competitive or facilitative.

Clonal growth is achieved in a variety of ways including rhizomes, stolons, bulbils, or lateral roots. Unlike sexual reproductive maturity which occurs between 10-20 years of age, asexual reproductive maturity is possible at the age of one year. Clonally
produced offspring are called “ramets”. In *Populus tremuloides*, clonal growth is akin to a branching pattern because each ramet develops a large distal and small proximal root (Figure 3-1). These lateral roots give rise to suckers that arise from adventitious meristems or pre-existing primordia on roots. Suckers develop into independent, physiological units, which in turn produce their own lateral roots and thus, give rise to additional ramets (Schier, 1973). During this repeated cycle of ramet birth, development and death, somatic mutations have the opportunity to occur. Through time, ramets within a clone will accumulate mutant lineages independently making the clone a mosaic of differently aged mutant and non-mutant cell lineages (Figure 3-2a). Furthermore, individual ramets may be themselves be mosaic, although the number of cell divisions in the root system of *P. tremuloides* are likely to be much greater than within the shoot itself.

The tacit assumption in previous attempts to estimate lifespan is that the age (in years) of the genet or clone can be measured from some area-based size metric obtained from the collection of ramets in a clone. The only other known alternative to assess age is to use radiocarbon dating of clones like *Larrea*, where all aboveground mass (shoots and branches) of the clone remain present. For some systems this may be appropriate because all ramets are alive and attached to the genet, e.g. *Carex* (Harris and Lovell, 1980; Steinger, 1995; Vasek, 1980). For other clonal plant species, like *P. tremuloides*, this may not be the case. The ‘virtual’ genetic individual can continue to exist through dormant root systems even if all of its shoots or ramets or suckers do not (Maini, 1965a; Maini, 1965b; Peet, 1981). Estimates of clone age made using a size measure assume that within and across clones in a population the birth-death process of ramet recruitment
is constant through time. In addition the method assumes that on average shoots have a constant and equal annual growth rate (cm/yr). Finally and most importantly, it implicitly assumes that clone size increases with time. But knowledge of how size changes during the lifespan of a clone is difficult to obtain, and it is unlikely that clone size always increases with time. Rather, environmental constraints may often limit, for example, the physical area covered by a clone. Furthermore, temporal changes in resources or pathogen virulence can cause a contraction in clone size during the lifespan of the clone, as is the case for trembling aspen when there is a stand dieback (Frey, 2004; Hogg, 2002). Such changes in size over time will result in the uncoupling of age and size (Hughes, 1984; Tanner, 2001).

An alternative approach to measuring clone age is to approximate the number of mitotic cell divisions that have occurred since the clone was established from a sexually produced propagule, i.e., a seed (Klekowski, 1997). If clones are a collection of mutant lineages then, one proxy for the number of mitotic divisions that have occurred since all mutant lineages descended from a single common ancestor, is neutral genetic divergence of a clone. Using an ontogenic ‘molecular clock,’ we explore this alternative estimate of the time to the common cellular ancestor (Figure 3-2b and 3-2c).

Consider two models of clonal growth: a constant population size (equilibrium model) and a sudden expansion in population size (non equilibrium model or a star phylogeny). The constant population size model (Wright-Fisher) characterizes the clone as composed of $N$ diploid ramets replicating each generation. Offspring ramets are sampled (with replacement) from $N$ diploid parental ramets who die each generation. Thus clone size is constant and finite over time. Under this equilibrium model of clonal
growth, the expected time to coalescence between any two alleles is $2N$ generations and
the probability of a neutral mutation, $\mu$, in either one of the two ramet lineages
is $2\mu$ (Hudson, 1990a). Thus, a pair of ramet lineages, differ by $\theta = 4N\mu$ mutations
(Figure 3-2b). In the second model of clonal growth, relationships among ramets can be
represented by a “star phylogeny.” Using this rough approximation of clonal growth, all
lineages (ramets) are independent replicates of the time to the most recent common
ancestor (Figure 2c). Here, the probability of a neutral mutation in either one of two
ramets remains $2\mu$ but the time until coalescence becomes $T_{CCA}$, (time to the common
cellular ancestor, the seed), the height of the genealogy or the clone age (Rosenberg and
Hirsh, 2003). Thus, any pair of ramets, under a star phylogeny, are separated by $2\mu T_{CCA}$
differences.

There are two ways to estimate differences among ramets within a clone: the
average number of pairwise differences for the $k^{th}$ clone, $\pi_k$, and the number of
polymorphic loci of the $k^{th}$ clone, $S_k$. The first estimator, $\pi_k$ is the average number of
pairwise differences in alleles between two ramets within a clone. In a population of
multiple clones, for the $k^{th}$ clone,

$$\pi_k = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} s_{ij}$$  \hspace{1em} (1) \hspace{1em} (Slatkin, 1996; Wakeley, 2007)

where $n$ is the number of sampled ramets and $s_{ij}$ represents the number of differences
counted between ramet $i$ and ramet $j$ at a single locus. For the equilibrium model, the
average number of pairwise differences, $\pi_k$ is an estimate of the expected number of differences separating two ramet lineages, $\theta = 4N\mu$. Under the non-equilibrium model, the average pairwise genetic distance among ramets of the $k^{th}$ clone, $\pi_k$ estimates differences separating any two ramet lineages, $2\mu T_{CCA}$.

A second way to estimate the expected number of differences within a clone, with a constant population size, uses the number of polymorphic sites in the $k^{th}$ clone, $S_k$, but standardizes it by the geometric series of the $i^{th}$ lineage given $n$ sampled ramet lineages (Tajima, 1989a):

$$\hat{\theta}_k = \frac{S_k}{\sum_{i=1}^{n-1} \frac{1}{i}}$$  \hspace{1cm} (2) (Tajima, 1989a).

Under the non-equilibrium model, the total number of polymorphic sites, $S_k$, is an estimate of the total number of differences among all ramet lineages, $n\mu T_{CCA}$, where $n$ is the number of sampled ramet lineages.

In two natural populations of $P. tremuloides$ we examined 14 microsatellite loci for somatic changes within clones. First, we compared how well these two different models of population growth fit the observed data. We expected that if the equilibrium model was most appropriate, a 1:1 relationship between $\pi_k$ and $\frac{S_k}{\sum_{i=1}^{n-1} \frac{1}{i}}$ would be expected. If on the other hand, clonal growth is best approximated by a star phylogeny or non-equilibrium conditions, then a 1:1 relationship between $\pi_k$ and $S_k \cdot \frac{2}{n}$ (rearranging
\[ \pi_k = 2\mu T_{CCA} \text{ and } S_k = n\mu T_{CCA} \] is expected. The purpose of this comparison was not to facilitate a reconstruction of the genealogical history of ramets within the clone. Instead, we wanted to determine which of the two different clonal growth models would be most appropriate in order to select a measure of genetic distance among ramets within a clone. Given these two different models of clonal growth, we then asked which size metric (area, diameter, etc.) best explains the variation in the number of somatic mutations. If size is linearly related to age, we might expect that larger sized clones (as measured by area, perimeter, or maximum distance between two ramets) will possess more somatic mutations per locus. We assume that mutations arise at the same rate across clones and that the loss of mutant lineages is also the same across clones within a population.

Finally, we explore the consistency of this relationship by examining two populations with contrasting environments. If there is a general trend across these two populations, it would indicate that a particular size metric can be used as a convenient way to estimate clone age. Documented reports suggest that clone size is quite variable in the western half of *P. tremuloides*’ range ranging in size from 3.8 acres to 106.8 acres (Kemperman and Barnes, 1976). Thus, the opportunity for a large number of somatic cell generations through time make aspen an ideal organism in which to examine molecular estimates of clone age.

**MATERIAL AND METHODS**

*Data Collection*

We collected foliage from 878 trees in two different populations of the dioecious species *Populus tremuloides* from Riske Creek, British Columbia (RC) and Red Rock, Waterton Lakes National Park, Alberta (RR). At Riske Creek, *P. tremuloides* grows in
discontinuous, clearly defined, patches or groves across the landscape with the occasional contiguous forested areas. Trees in both populations were sampled in two ways: on the perimeter of a grove or patch of aspen and systematically along two or three 30-50m transects within a patch. On average 30-50 individuals were sampled per patch. No tree less than 1.5m in height was sampled, and only patches separated by at least 1 km of terrain lacking aspen trees were used. Trees on the perimeter and along transects were physically mapped using both a measuring tape and a handheld GPS unit. In addition to collecting foliage and GPS measurements on all transect trees, we measured height and diameter at breast height (DBH). The heights of trees were measured using a clinometer and tape measure, and diameters at breast height were measured with a steel DBH tape. Finally, an increment core was taken from a sampled of ramets belonging to each genotype.

In the Red Rock Canyon population, trees grew sparingly and were interspersed amongst coniferous trees along a ridge. Every attempt was made to sample in the same manner as in the grasslands, but in most cases all the trees in the designated patch were sampled. A total of 719 individuals in 17 patches were assayed at Riske Creek. A total of 152 individuals in 10 patches were sampled at Red Rock.

Criteria for Somatic Mutations

We examined 14 microsatellite loci for somatic changes. Microsatellites are ideal loci to use for estimating clone age for two reasons. First, the high levels of polymorphism are reflective of high rates of mutation, which are on the order of $10^{-3}$ to $10^{-6}$ per sexual generation (Chakraborty et al., 1997; Ellegren, 2004; O'Connell and
Ritland, 2004; Schlotterer et al., 1998; Schug et al., 1997; Thuillet, 2002; Udupa and Baum, 2001; Vazquez, 2000). High mutation rates provide higher resolution for determining the age of a clone. Secondly, microsatellite length changes during DNA replication via the mechanism of slipped strand mispairing. Hence, as is the case with other mutations, length polymorphisms are directly linked to cell division (Frumkin et al., 2005). Furthermore, microsatellite somatic mutations, previously demonstrated in aspen clones (Tuskan et al., 1996), will include length changes arising from simple replication errors, slipped strand mispairing or homologous recombination (Puchta et al., 1994; Swoboda et al., 1993). While most microsatellite mutations involve one or two repeats, some mutational events are multi-step changes (Di Rienzo et al., 1998; Renwick et al., 2001). In addition to facilitating insertion or deletion of repeats, homologous mitotic recombination can cause the loss of heterozygosity, allowing homozygotes at a locus to appear within otherwise heterozygous clones. Studies suggest that recombination events occur both early and late in plant development and are found in all plant organ types: flowers, roots and shoots (Puchta et al., 1994). The frequency of homologous somatic recombination among repetitive DNA ranges from 1 event in $10^5$ to 1 event in $10^7$ cells (Puchta et al., 1994). In plants mitotic recombination per generation is of the same order of magnitude as meiotic recombination (Puchta et al., 1994).

For this study, alleles were scored as somatic mutations when an individual ramet in a clone differed by one allele at one locus but at all other loci shared the same alleles as the most frequent genotype. Thus, we accepted alleles as mutants regardless of the change in repeat number. Our conservative approach to scoring microsatellites does
not allow for multiple mutations in the same lineage. While a ramet is limited to a maximum of one mutational event, clones can have more than one somatic change.

Eight of the loci are known to map to seven of the 19 possible different linkage groups in the *Populus* genome (Cervera, 2001; Yin *et al.*, 2004). For the other six loci, we have no mapping information. PCR was performed on all ramets in both populations to identify potential somatic mutations. In all PCR reactions we used a hi-fidelity AmpliTaq DNA polymerase (Roche) with no additional dNTPs or MgCl₂. The maximum number of PCR cycles was 30, and in a 10uL reaction we used 1.0uL of 10X PCR buffer (Roche) (MgCl₂: 1.5mM) with 0.2mM dNTP. In the cases where PCR failed, we also performed several additional reactions altering the conditions. Failed PCR reactions could be because of the usage of non-focal microsatellite primers. Twelve of the 14 primers used were originally designed on *Populus trichocarpa* and *Populus deltoides* (Cervera, 2001; Tuskan, 2004; Yin *et al.*, 2004)(see Table 3-1). Evolutionary divergence between *P. tremuloides* and *P. trichocarpa* (section: Tacamahaca) and *P. deltoides* (section: Aigeros) could allow for mutations to arise in flanking primer regions producing null alleles, i.e., no PCR fragment. A failed PCR reaction was not counted as a somatic mutation, and the locus was excluded from the analysis for that clone.

*Ramet population growth models*

Using microsatellites, we measured the number of polymorphic sites, $S_k$, by calculating the proportion of polymorphic loci for the $k^{th}$ clone where $n$ was the total number of sampled ramets in the clone. For the measure, $\pi_k$, we measured differences between all pairs of ramets within a clone across multiple microsatellite loci. This was
then divided by the total number of unique pairwise comparisons among sampled ramets in a clone to get the average pairwise estimate. To determine which population growth model, the Wright-Fisher or a star phylogeny, best represented the genetic data, we examined under which model the two estimates showed a one to one relationship. We expect that a star phylogeny would be more appropriate to the data for two reasons. First, clones are initiated by a single seed and thus do not have a constant ramet population size over time. Secondly, we employed the restriction of one somatic change per lineage. In order to show that this restriction was not why the data followed a star phylogeny, we performed a sensitivity analysis to determine if accepting two changes within a single ramet lineage (relaxed dataset) affected the outcome (see Results).

Relationship between clone size and time

To determine if clones grew in a perfect circle, we described clonal growth by plotting the degree of invagination against elongation. Degree of invagination (DI) and elongation (E) best capture the irregular pattern of clonal growth. We compared the perimeter and diameter, respectively, to that expected from a perfectly circular clone of the same area. Thus the degree of invagination (DI) is measured as

\[ DI = \frac{P_{\text{obs}}}{2 \cdot \sqrt{\pi \cdot \text{Area}}} \]  

where \( P_{\text{obs}} \) is the observed perimeter. When \( DI=1 \), the clone’s perimeter matches a perfect circle, while \( DI>1 \), indicates the clone has a perimeter more extensive than
expected from a circle. Elongation measures the extent to which a clone resembles an ellipse, quantified by

\[ E = \frac{D_{\text{max}}}{2 \sqrt{\frac{\text{Area}}{\pi}}} \]  

(4)

where \( D_{\text{max}} \) is the maximum distance between any two ramets. When \( E > 1 \), the clone exhibits a stretched form.

If mutations accumulated as the clone grows through time, then a linear relationship between an estimate of genetic variation and some measure of clone size is expected. Three measures of clone size were explored: area (A), perimeter (P), and maximum distance between any two ramets in a clone (\( D_{\text{max}} \)). To calculate perimeter of a clone and clone area, we used the software ArcView GIS 3.3 (2002 Environmental Systems Research Insitute). ArcView GIS 3.3 creates polygons from the GPS data and then sums the lengths of all the edges of the polygon. The value of the length segments in the polygon was calculated using the standard Euclidean distance, where the physically mapped co-ordinates of two different ramets (at the edge of a clone) were the x and y. To obtain \( D_{\text{max}} \), we estimated all possible pairwise physical distances (m) separating any two ramets and then ordered the distances. From these pairwise distances between any two ramets, we selected the maximum distance. These size metrics were then log-transformed to meet the assumption of normality. Cases where two clones shared the same patch were not an issue as each sampled ramet of a genotype was physically mapped. We analyzed the relationship between size and clone age for our two
populations separately because it has been previously shown (Ally et al. unpublished) that key features of these environments including density, soil moisture, and nutrients differ.

To select the best predictor of the neutral mutation load, the three predictors (logArea, logDI, logE) were included in a stepwise selection process in a linear regression model. We followed a stepwise selection process, which entered and removed terms until the model with the lowest Akaike Information Criterion (AIC) was obtained. Linear model evaluation included examining scaled residual error, goodness of fit tests, and AIC criterion. To determine if some environmental variable may be responsible for size variation, we obtained an assay of site quality by measuring the percent cover of trees, shrubs, herbaceous plants, and moss, lichens, and bryophytes in each of the different patches. This vegetation analysis used a single sampling plot with a radius of 5m per aspen patch. Surveys were conducted during the months of May and July 2003. Where possible, all non-woody herbs and shrubs were identified to species level. In a few cases where habitats were similar and more than one species of a genus was found, we collapsed species into genus-level groups. All statistical tests and analyses were performed using R version 2.3.1. (THE R FOUNDATION FOR STATISTICAL COMPUTING 2006).

Estimates of clone age

Clone age or the time to common cellular ancestor, $T_{CCA}$ can only be estimated when clonal growth approximates a star phylogeny. This is because the average pairwise genetic distance among ramets of the $k^{th}$ clone, $\pi_k$, is an estimate of $2\mu T_{CCA}$. Given a

71
mitotic mutation rate, clone age or \( T_{\text{CCA}} \) can be obtained, using the equation 
\[
T_{\text{CCA}} = \frac{\pi_k}{2\mu} .
\]

Conversely, if clone age was known, a mitotic mutation rate can be estimated. We provide two different estimates of the per locus mitotic mutation rate: a per cell generation estimate and a per year estimate. We also provide an estimate of clone age in cell generations and in years.

An estimate of the per cell generation mitotic mutation rate was obtained by counting the number of mutational events that occurred during the growth of a clone in physical space. This clonal spread was represented by the sum of all possible pairwise distances (\( m \)) separating any two sampling points within a clone. A mutational event was counted only once even if it occurred in several ramets. We coupled the number of mutational events detected in the total meters of growth for a clone with an estimate of root growth of 9.5mm per day (Coleman, 1996) and cell cycle duration. This estimate of root growth was a maximal rate during ~20 days in the growing season when most root production occurs (Coleman, 1996). We were not able to find a species-specific estimate of cell cycle duration for \textit{Populus tremuloides}, thus we used an estimate of 18hrs obtained from cortical cells in \textit{Arabidopsis thaliania} (Beemster and Baskin, 1998). Multiplying the number of mutations in total meter growth of a clone by the following factor:

\[
\frac{0.0095 m}{\text{day}} \cdot \frac{1 \text{ day}}{24 \text{ hr}} \cdot \frac{18 \text{ hr}}{\text{cell cycle}} = \frac{0.0071 m}{\text{cell cycle}} \quad (5)
\]
converts mutations per meters growth into mutations per cell cycle or cell division. The length of the mitotic cell cycle, however, may be affected by the amount of heterochromatin, suggesting that use of cell cycle parameters from annuals might underestimate cell cycle duration in a longer-lived species (Bennett, 1972; Nagl, 1974). As an alternative estimate to account for higher levels of heterochromatin, we used a cell cycle estimate from *Pinus banksiana* of 25.7hrs (Miksche, 1967). Using the formula $T_{CCA} = \frac{\pi_k}{2\mu}$, we then estimated clone age in cell generations. As root growth is both a function of elongation and division (Beemster and Baskin, 1998), the effect of not including elongation is likely to make the age in cell generations an overestimate.

As an alternative method, we used a maximum and minimum clone age to determine upper and lower bounds on the per year mitotic mutation rate. Pollen records suggest *P.tremuloides* was first present in southern British Columbia (location of Riske Creek) approximately 8000 years BP (Alley, 1976; Cawker, 1983). Thus, we assumed the clone with the most neutral divergence, $\pi_{\text{Largest}}$, was a maximum of 8000 years old. Setting $T_{CCA} = 8000$, we obtained the lower bound on the per year mitotic mutation rate $(\mu_{\text{Lower}} = \frac{\pi_{\text{Largest}}}{2 \cdot 8000 \text{yrs}})$, which can be used to estimate the maximum possible ages of our clones in years. To determine an upper bound on the mitotic mutation rate we used tree ring data. Because increment cores were taken from individual ramets within the clone, we could estimate the age of the oldest sampled ramet in each clone using tree ring counts. We averaged the ages of the oldest ramet across clones and coupled this with averaged values of $\pi_k$. Setting the average age of the oldest clone to $T_{CCA}$, allowed us to
place an upper bound on the mitotic mutation rate \((\mu_{\text{upper}} = \frac{\pi_{\text{RC}}}{2 \cdot Av\text{Yrs}})\), from which we could determine the minimum ages of the clones in years. We excluded the clone with the highest value of \(\pi_k\) because we were interested in determining the minimum clone age. In addition, although observed minimum values of \(\pi_k\) were zero, due to a limited number of loci studied, we included these values in the estimate of the average value of \(\pi_k\). A similar procedure was used in Red Rock to obtain upper and lower bounds for the mitotic mutation rate, where \(P.\ tremuloides\) pollen was found approximately 6600 years BP in Waterton Lakes, the location of our Red Rock population (Bujak, 1974).

**RESULTS:**

*Somatic mutations*

We genotyped a total of 715 ramets in Riske Creek and found a total of 24 somatic changes across 14 microsatellites. In Red Rock, a total of 147 ramets were genotyped for 14 microsatellite loci, and we found 5 somatic changes. Before describing the relationship between the level of neutral divergence within a clone and some measure of size, we discuss the overall number and type of microsatellite changes across the two populations to show that these mutations are not artifacts of PCR. In both populations somatic mutations shared by a cluster of neighboring ramets were found. It is unknown which ramet had the first mutational event, however, these mutation clusters likely do not represent multiple mutational events but a mutation inherited by nearby ramets from their shared ancestor.
Of the eleven one-step changes, eight were an increase while three were a decrease in repeat number. Of the thirteen somatic mutations where the change was more than single repeat, six of these were 2-step changes. Eight of the thirteen k-step (>1 step) changes were decreases and five were increases. The five remaining were cases involving a clone that was otherwise heterozygous at a locus but the individual ramet was a homozygote. Other than their association with specific linkage groups, we could not find map information indicating the distance of these markers to the centromere. These changes were scored as most likely a result of a homologous recombination event during mitosis (Table 3-2). The number of repeat changes will not only be determined by the number of repeats and repeat type, but how many timesteps have occurred, in this case how many cell generations or years between mutations. One study of microsatellite mutations in maize over 11.3 generations found 95% dinucleotide repeats changes were either 1 or 2 step changes (Vigouroux et al., 2002). In this study when only dinucleotide loci were considered, we found over 80% of the somatic changes were 1 or 2 step events.

In general, Taq DNA polymerase is expected to have a base pair substitution error rate that is a function of the starting population size (amount of DNA), the number of cycles, and the amount of MgCl₂ and dNTPs. For the concentrations we used, the error rate of Taq is 1.33x10⁻⁶ point mutations per base pair per duplication (Cline et al., 1996). This per base pair error is more likely to create an increase or decrease in one base pair in the length of the microsatellite. Given the number of cycles or duplications used in our PCR was 30, the probability of getting an increase in the length of the microsatellite by one bp, three times in a row was 6x10⁻¹⁴ mutations per bp. This calculation assumes
mutations were independent, however, this may not be true if there are aspects in the DNA sequence that it make it prone to error, i.e. hairpin loop formation. In order to discriminate between a true allele change due to a somatic change and PCR error we performed two additional PCRs on the ramets with the potential mutations. To maintain a consistent sizing across different gels, we also performed two additional PCRs on a clonemate where no mutations were observed. Furthermore, we used high quality DNA template, i.e., young tissues, reducing the potential allele dropouts, allelic competition, and high amounts of stuttering (Dewoody et al., 2006). With this in mind, we provide five lines of evidence to suggest why the somatic changes we observed are not artifacts of AmpliTaq errors. First, studies which have examined the error rate on mononucleotides and dinucleotide repeat sequences found PCR amplification introduces systematic decreases, most of which were a loss of a single repeat (Clarke et al., 2001; Walsh et al., 1996). We found approximately equal numbers of changes that resulted in either an increase (13) or decrease (11) in the number of repeats. Of the single repeat changes, we found 8 increases and only 3 decreases in repeat number. Secondly, alleles were amplified three independent times and were only counted if amplified consistently. Thirdly, over half of the mutations occurred at only a subset of the microsatellite loci (PMGC575, ORPM276, PMGC510, W19U). Although longer repeats are likely to have higher error rates, at the longest microsatellite, W19U, reported to be a tri-nucleotide repeat of 28 repeats long, we observed only three somatic changes, one of which is likely the result of homologous recombination. Fourthly, we can rule out contamination because none of our negative controls showed any evidence of products of the same
length as the mutants. Lastly, to mitigate scoring errors, we reanalyzed our data at two separate timepoints both with the software and by visual inspection.

**Ramet population growth models**

As illustrated in Figure 3-3, the slope of a regression of $\pi_k$ on $S \cdot \frac{2}{n}$ (star phylogeny) was nearly 1 ($\beta_{\text{star}} = 1.09 \pm 0.33$, H$_0$: $\beta_{\text{star}} = 1$, p-value=0.765), while the slope of a regression of $\pi_k$ on $\hat{\theta}_k$ (constant population size) was not ($\beta_{W-F} = 0.48 \pm 0.15$, $\beta_{W-F} = 1$, p-value=0.0014). Thus, a star phylogeny was more consistent with the genetic relationships among clones than an equilibrium model. The sensitivity analysis, where we accepted two changes in any one lineage, did not change the results ($\beta_{\text{star}} = 1.09 \pm 0.33$, p-value=0.86, $\beta_{W-F} = 0.48 \pm 0.15$, p-value=0.0005). This implies that $T_{CCA}$ might well be estimated by $\frac{\pi_k}{2\mu}$. We thus proceed to ask whether $\pi_k$ is related to the size of a clone.

**Relationship between clonal divergence and size: Riske Creek**

A plot of the degree of invagination (DI) and elongation (E) (Figure 3-4) showed most clones do not grow as a perfect circle. In fact, most clones were highly convoluted in shape with only a single clone exhibiting growth in a perfect circle. A correlation analysis showed that area and perimeter were positively correlated ($\rho = 0.92$, p-value=1.02x10^{-8}) as were area and the maximum distance between two ramets, $D_{\max}$ ($\rho = 0.63$, p-value=0.0028). There was, however, no relationship between perimeter and $D_{\max}$ ($\rho = 0.55$, p-value=0.0113) after a Bonferroni correction ($\alpha = \frac{0.05}{3} = 0.02$).
A stepwise procedure found no single predictor was sufficient to explain the variation in square-root transformed $\pi_k$. The model with the lowest AIC criterion is typically identified as the ‘best’ model. The model with the lowest AIC criterion was a model with only the intercept (AIC=-44.81). The full model with all three predictors ($\log$ (Area) + $\log$ (Perimeter) + $\log$ ($D_{\max}$)) had the highest AIC score of -39.5 and was not significant ($F_{3,16}=0.1993$, p-value=0.90). Typically the DeltaAIC, 

$$\Delta_i = AIC_i - AIC_{\min},$$

is used to provide evidence for a model where the $AIC_{\min}$ is the value of the best model. The model with all three predictors had a $\Delta_i=5.26$ indicating the full model had considerably less support. For models with individual predictors only, the DeltaAIC was between 1.3 and 1.6, suggesting little ability to discriminate between reduced models. None of these models, however, were significant (Table 3-4).

We log transformed all size measures and square-root transformed $\pi_k$ to meet the assumption of normality of the residuals for the linear regression. Thus, our data do not support a relationship between clonal divergence and clone size.

To determine if clonal growth could be explained by ecological correlates, a principal component analysis using a correlation matrix performed on 54 environmental variables. We found these 54 variables reduced to eight axes that captured 74.3% of the variance in the original data. We examined the relationship between $\log$ (Area), $\log$ (DI), and $\log$ (E) against these eight principal components. Neither PC1 nor PC2 showed any relationship to $\log$ (Area), but PC3, which captured 10.48% of the variation, had a significant relationship ($\beta=-0.61$, s.e. = 0.23, t-value= -2.598, p-value= 0.0182) with $\log$
transformed area ($F_{1,18}= 6.749, R^2= 0.27, AIC=64.28$). A plot of log (Area) against PC3 can be seen in Figure 3-5.

We found 33.0% of the variation in log (DI) was explained by both PC1 ($\beta =-0.31$, s.e.= 0.13  $t=-2.328$  $p$-value=0.0325 ) and PC3 ($\beta =0.31$, s.e.=0.13, $t= 2.378$  p-value=0.0294 ) and the linear regression of log (DI) against these two principal components was significant ($F_{2,17}= 5.607$,  p-value: 0.0135). To determine what ecological factors were represented by PC1, PC2 and PC3 we examined the loadings. After a Bonferroni correction for multiple comparisons, PC1 was significantly negatively correlated to percent cover of Agrostis ($\rho=-0.61$, p-value=0.002), Homalothecium aeneum ($\rho=-0.75$, p-value<0.0001), Hieracium ($\rho=-0.675$, p-value<0.0001), Peltigera canina($\rho=-0.75$, p-value<0.0001), Pleurozium schreberi ($\rho=-.61$, p-value<0.0001), Rhytidiopsis robusta ($\rho=-0.63$, p-value=0.0013), Rosa($\rho=-0.73$ p-value=0.0001) and significantly positively correlated with Potentilla anseria ($\rho=0.66$, p-value=0.0006). While PC3 was significantly correlated to the percent cover of Antennaria microphylla ($\rho=0.55$, p-value=0.007).

Relationship between clonal divergence and size: Red Rock

In Red Rock, there was no significant correlation between area and perimeter ($\rho=0.69$, p-value=0.0878) or between area and $D_{\text{max}}$ ($\rho=0.75$, p-value=0.0524). We did, however, find a significantly positive correlation between perimeter and maximum distance between two ramets ($\rho=0.993$, p=6.6x10^-6) after a Bonferroni correction ($\alpha = \frac{0.05}{3} = 0.02$).
We also found no linear model using any of the predictors was sufficient to significantly explain the variation in $\pi_k$. The model with the highest AIC was a model with the intercept only (-11.91: see Table 3-4). The model with the lowest AIC was the full model (-16.35). Models with a single predictor had a $\Delta_i \leq 2$. Using the DeltaAIC criterion only, a model with all predictors was better than a model with no predictors at all. However, despite the lowest AIC, this full predictor regression model was not significant ($F_{3,3}=3.444$, p-value: 0.1685, adjRsq= 0.55). One caveat is we excluded any clones which had fewer than 5 ramets from any analysis leaving us with a sample size of 5 clones. Thus for this population, we had very few degrees of freedom in our analysis. Additionally, we found no relationship between any environmental variables and size or shape metric in this relationship.

Estimates of clone age

Using the rate of root growth in mm/cell generation (equation 5) and the total estimated root growth within a clone, we found an average per locus per cell generation mitotic mutation rate of $4.8 \times 10^{-7}$ (with a cell cycle duration of 18hr: S.D=9x10^{-7}) and $7.0 \times 10^{-7}$ (with a cell cycle duration of 25.7hr: S.D=1.3x10^{-6}) for the Riske Creek clones, while in Red Rock we estimated a per cell generation mutation rate of $5.3 \times 10^{-6}$ (with a cell cycle duration of 18hr: S.D=1.3x10^{-5}) and $7.6 \times 10^{-6}$ (with a cell cycle duration of 25.7hr: S.D=1.9x10^{-5}). Using the per locus per cell generation mitotic mutation rate of $4.8 \times 10^{-7}$, the age of our oldest clone in Riske Creek, was 83,000 cell generations. Among
clones where we obtained non-zero estimates of $\pi_k$, estimates of clone age using $\frac{\pi_k}{2\mu}$ ranged from 2,600 to 83,000 cell generations in Riske Creek, with the average clone age in cell generations at 11,000 (S.D.=18,000). In Red Rock, however, using the mitotic mutation rate of $5.3 \times 10^{-6}$, we found the oldest clone was 2500 cell generations and clones ranged in age from 1300 to 2500. The average clone age in cell generations was 1200 (S.D.=1100).

Using a maximum clone age of 8000 years, more time will have passed, making the mitotic mutation rate a lower bound. We estimated of this lower bound as $\mu_{\text{Lower}} = 5 \times 10^{-6}$ per locus per year with maximum possible clone age ranging between 250-8000 years old. Using the minimal clone age based on tree ring data, less time has passed and many more mutations are likely to have occurred, making the mitotic mutation rate estimate an upper bound. The upper bound on the mitotic mutation rate was $\mu_{\text{Upper}} = 2 \times 10^{-5}$ per locus per year with the minimum possible clone ages ranging between 30-900 years old in Riske Creek. In Riske Creek, the average of the oldest ramet across all clones was 77 years and the average $\pi_k$ estimate was 0.0067. In Red Rock, the upper and lower bounds around the mitotic mutation rate per year were $\mu_{\text{Upper}} = 7.0 \times 10^{-5}$ and $\mu_{\text{Lower}} = 2 \times 10^{-6}$. In Red Rock, the average of the oldest ramet across all clones was 72 years, while the average in Riske Creek, the average of the oldest ramet across all clones was 77 years, while the average $\pi_k$ estimate was 0.0103. In Red Rock, clones were at a minimum of between 100-200 years of age and a maximum of between 3000 and 6600 years old.
DISCUSSION

The assumption of a positive linear relationship between clone age and clone size has not been explored in any detail in the clonal literature. Our data illustrate the complexity underlying this relationship. Despite variation both in clone size and genetic distance within a clone, no simple relationship emerged between clone size and age. We found no repeatable pattern across our populations in terms of the relationship of either shape or size to variation in the number of somatic changes. This result is not surprising given that how a clone occupies physical space is not predictable by time alone but depends on the genotype’s competitive ability given the context of the environment.

Clonal growth models

We examined two different growth models (Wright-Fisher and star phylogeny) to describe the demography of ramet populations within clones. Our data suggest neutral genetic divergence within clones is adequately described by a sudden expansion model of growth (Figure 3-3). This model allowed us to estimate ages in cell generations and in years for the clones in our data.

Nevertheless, several caveats are in order. Clearly a star phylogeny cannot be a perfect descriptor of the history of the clone, because we have evidence of shared mutations among neighboring ramets. Instead, we argue that the ancestry is close to star like and that \( \pi_k \) can be used to estimate the time to a common cellular ancestor in cell generations (\( T_{CCA} = \frac{\pi_k}{2\mu} \)). We have used \( \pi_k \) rather than \( S_k \) as a surrogate of age because the former is less sensitive to the existence of clusters of related lineages. An additional
caveat is that we have assumed an infinite sites model when measuring $S_k$, implicitly assuming that we could detect each mutational event. With high rates of mutation at microsatellite loci, it is possible to have parallel mutations, thus reducing the number of segregating sites expected under the infinite sites model (Bertorelle and Slatkin, 1995). In addition, there is a strong downward mutation bias in long microsatellites increasing the possibility of changing to the same repeat number (Wierdl et al., 1997). Although on a longer time scale (millions of generations) microsatellite loci suffer from multiple mutations, we are using shorter timescales (hundreds of generations) where this scenario is less likely. One study showed that estimates of $\pi_k$ were not sensitive to change in mutation model (infinite-sites to a finite-sites) (Bertorelle and Slatkin, 1995), providing further reason to prefer $\pi_k$ as a measure of clone age.

Finally, the ability to detect somatic mutations may be influenced by field sampling strategy and different clonal growth patterns (birth-death processes). If the birth of new ramets occurs primarily at the perimeter of a clone, while death occurs in the centre, then the ability to detect somatic mutations will be greater in perimeter ramets. Habitat heterogeneity in a patch involving such factors as herbivory, competitive interactions, soil temperature, light and nutrient availability and soil moisture will affect sucker initiation, growth, survival and density. For example higher soil temperatures, which may occur at the edge of a patch, could stimulate early initiation of suckers, while very dry or water-saturated soils will reduce sucker initiation (Maini and Horton, 1966; Schier, 1982). Thus, estimates of age might well be influenced by how and where sampling occurs.
Estimates of the mitotic mutation rate and clone age

Clones ranged in age at Riske Creek from a possible minimum of between 30-900 years old to a possible maximum of between 250-8000 years of age. In Red Rock, on the other hand, clones ranged in age at a minimum of 100-200 years of age or at a possible maximum between 3000 and 6600 years old. Most reported per year mutation rate estimates lie closer to the lower bound of the per year mitotic mutation rate found in this study ($\mu_{\text{Lower}} = 5 \times 10^{-6}$), implying that clones are likely quite old. Our genetic data correspond with the pollen records, implying that clones at Red Rock are probably younger than those in Riske Creek. After the glaciers retreated 10,000 yrs BP, *Populus tremuloides* was found in the pollen assemblage zone in western interior lake samples around 8,000 BP (Alley, 1976; Cawker, 1983; MacDonald and Ritchie, 1986). On the other hand, in Alberta, glacier retreat is dated at about ~8000 BP, and pollen cores from Waterton Lakes National Park (Red Rock) date aspen establishment to the late Holocene, ~6600 years BP (Bujak, 1974). Furthermore, given the variance in clone age, in both populations there is continual recruitment of new seeds into the population.

We have presented estimates of the somatic mutation rate in two different units of time. The per locus per cell division rates were between $4.8 \times 10^{-7}$ (RC) and $5.3 \times 10^{-6}$ (RR), while the per year estimates had lower and upper bounds of $\mu_{\text{Lower}} = 5 \times 10^{-6}$ and $\mu_{\text{Upper}} = 5 \times 10^{-4}$. Both these estimates incorporate somatic mutation and mitotic recombination in somatic cells. It is difficult to find comparable microsatellite mutation rates whose units are in *per cell generation* (doubling time for one cell to become two) for plants, let alone for clonal organisms. There are, however, several estimates of *per sexual generation* (from seed to reproductively mature adult) rates available. For example, in *Zea mays*
(subsp. *mays*) across 142 loci, the rate was between $1 \times 10^{-3}$ and $5 \times 10^{-4}$ (Vigouroux *et al.*, 2002) and across 10 loci the per locus per generation rate in *Triticum turgidum* it was $2.4 \times 10^{-4}$ (Thuillet, 2002). A similar estimate was found across 8 microsatellite loci in *Thuja plicata*, where somatic mutations arose on average between $3 \times 10^{-5}$ and $4 \times 10^{-3}$ per locus per generation (O’Connell and Ritland, 2004). Finally, two studies on *Drosophila melanogaster* found a much lower microsatellite mutation rate of $5 \times 10^{-6}$ (Vazquez, 2000) and $6 \times 10^{-6}$ per locus per generation (Schug *et al.*, 1997). Both studies suggested one reason for the lower mutation rates in Drosophila were the shorter length of microsatellites making them more stable (Schug *et al.*, 1997).

Our estimate of the rate at which somatic mutations arise per cell division (RC: $4.8 \times 10^{-7}$ and RR: $5.3 \times 10^{-6}$) was comparable to estimates found in the literature, *Pinus strobus*: $1 \times 10^{-7}$ and $5 \times 10^{-7}$ (Cloutier *et al.*, 2003). In this study, because no somatic mutations were observed, the zero term for the Poisson distribution was used, where $\lambda = 2\mu\kappa$. In this study, $\kappa$, was the number of cell divisions that occurred between sampling positions within a tree. To estimate $\kappa$ the researchers divided the mean cell diameter, measured in the shoot meristem and mature pith tissue, by the linear distance between two sampling points (Cloutier *et al.*, 2003). The use of medial pith cells was justified because they originate close to the apical meristem, are constant in size, and remain relatively undifferentiated (Scofield, 2006). But in assuming that the number of cell generations is exactly the number of cell lengths minus one, this method implies medial pith cells are the products of a single, irreplaceable, dividing cell in the apical meristem. While this may be true in some tissues, such as primary roots, where cell file number is conserved and apical initials follow a stringent organization (Rost and Bryant,
1996). It is not the case in other tissues, such as lateral roots, where cell file numbers are more variable and the pattern of cell division is less regulated (Dolan et al., 1993). Furthermore, it is unknown if apical initials live forever or are replaced by proximal cells in the quiescent centre and at what rate replacement might occur. Finally, changes in the organization of primary root meristematic cells with the age of the plant have been observed in some species of the Asteraceae, for example *Helianthus annuus* (Dolan et al., 1993).

Applying the method of Cloutier et al. (2003; and subsequently Scofield, 2006), we illustrate the limitation of an implicit linear growth assumption. If we use a longitudinal estimate of a hybrid aspen xylem fiber (seedlings) of 16.21 μm and or a longitudinal estimate of columnar epidermal cells from *Populus trichocarpa* embryos of 90μm (Barlow, 2002; Sundberg, 1983) to determine the total number of cells resulting from root growth within a clone, we find a range of per locus per cell generation mitotic mutation rate of $5 \times 10^{-10}$ to $9 \times 10^{-9}$. Using this mitotic mutation rate, we can estimate the age in cell generations of the oldest clone (i.e., with the most neutral genetic divergence $\pi_{\text{Largest}} = 0.079$). This makes the age of the oldest clone in Riske Creek 4 million to 80 million cell generations. Using an estimate of the cell cycle duration for a long-lived woody species of 25.7hr, this translates into approximately between 13,000 and 200,000 years old, an age that is incongruent with the palynological records. By comparison, using a mutation rate obtained under a model of exponential cell division ($4.8 \times 10^{-7}$ per locus per cell generation), where a progressively larger number of cells divide at an approximately constant rate, this same clone is ~83,000 cell generations or ~240 years old. An exponential model of cell production is still problematic because it assumes a
progressively larger number of cells divide at an approximately constant rate. A small pool of apical cells within the structured plant meristem divide (Lyndon, 1998). Furthermore, the number of dividing apical cells is determined by cell size, the size of the meristem, the tissue type (shoots or roots), and the size and rate of the elongation zone (Beemster and Baskin, 1998). Neither linear nor exponential growth models are likely to be correct.

Typically, when estimating the rate at which mutation arises within a population, the unit of time used is a sexual generation. A sexual generation is defined as the mean age to reproductive maturity. This unit of time is most relevant because natural selection acts on new mutations in terms of their per generation impact on relative fitness. While straightforward for purely sexual organisms, the unit of time is more complicated in clonal organisms because of the multiple levels at which selection can act. If fitness is defined as the relative contribution of mitotic offspring to the next generation then a per cell division or per ramet generation may be more appropriate. Moreover, for comparisons of mutation rates between a clonal plant and a sexual one requires knowledge of the number of cell generations found in a single sexual generation. Translating these estimates, however, requires knowledge of the number of cell generations in one sexual generation. There are a lack of biologically realistic models to describe cell growth and production in plant modules.

Relationship between clonal spread and divergence

We found no support for a linear relationship between genetic divergence within a clone and size in either population. Our data implies that how a clone spreads through
physical space has little relationship with age, but may, in fact, be affected by competition and herbivory (Figure 3-4). In Riske Creek, patches with a higher percent cover of *Antennaria microphylla* had a smaller clone area but were more convoluted in shape, while those with low *Antennaria* cover had clones with larger areas. Typically, *Antennaria microphylla* when found in association with *P. tremuloides* stands, indicate stands that have been heavily grazed by cattle and deer (Burke, 1989). Much of the Cariboo-Chilcotin especially around Riske Creek has a history of grazing by livestock (cattle and horses) and big horn sheep (*Ovis canadensis californiana*) (Hooper and Pitt, 1996).

Clone age is important to ascertain because it informs different aspects of ecological and evolutionary theory on organisms with complex life histories. In order to assess how clonal plant populations increase or decrease in size through time, demographic parameters including growth and mortality rates for the genet are required (Tanner, 2001). Furthermore, mechanisms that arise in a clonal species in response to temporal and spatial heterogeneity may depend on the timescale over which it experiences the environment. In environments where only certain genotypes can grow, asexual reproduction, clonal expansion and/or growth may simply be a strategy that overcomes the two-fold cost of sex (Peck *et al.*, 1998). On the other hand, clonal growth may be a plastic response to temporal and spatial environmental heterogeneity that increases the efficiency of resource acquisition and reduces mortality risk (Charpentier, 1999). Finally, the relative strength of intra-organismal selection (Klekowski, 2003) depends on understanding genet fitness, which in turn requires knowledge of genet size and age.
This study is the first to characterize the association between neutral genetic variance and clone size, motivated by the desire to have a quick and easy proxy to assess clone age. Our results indicate that no simple size metric can be used to estimate clone age and more labor intensive estimates of genetic divergence are needed. Specifically, distinguishing between plant growth models (linear vs. exponential) will resolve estimates of the mitotic mutation rate making age estimates more reasonable. Additionally, empirical work which examines how field sampling might affect clone age estimates is needed. For example, if asexual offspring are solely recruited at the edges of a forest patch, then sampling within a grove of aspen will underestimate clonal divergence. As more information becomes available on clonal growth, development, architecture and responses to habitat, we may improve our estimation of the lifespan of plant clones.
Figure 3-1. A hydraulic excavation of part of a *Populus tremuloides* clone shows the pattern of interconnection among ramets. When ramets die, roots decay and the surviving ramets may be fragmented from the rest of the clone. From Barnes (1966).
Figure 3-2. **a.** Clonal growth through time from a common cellular ancestor, the seed. As the genet grows, there are repeated bouts of ramet birth and death possibly with mutations, resulting in a clone becoming a genetic mosaic. We represent the different somatic mutations carried by ramets in the shape and colour of circles. Without knowing the original genotype of the seed, it is impossible to determine the true ancestral lineage within a clone. For simplicity’s sake, we designate the most frequent genotype as the ancestral lineage (open circle). The genet is a collection of ramets of varying physiological ages (numbers beside the circles). In *P. tremuloides*, ramets often remain connected to each other via lateral roots (solid lines) but sometimes, these root connections degrade (dashed lines), fragmenting the individual. **b.** If the age of the somatic mutations corresponded to the age of the ramet and the ancestor was known, then we might surmise the genealogy. Despite the presence of mutations of different ages, clonal growth will still produce a genealogy where the internal branches are shorter than the external ones (Slatkin, 1996) In aspen a tree is capable of asexual reproduction as young as one year of age. Thus, the ancestral and derived ramet are indistinguishable and ancestral nodes may still be present at time of sampling. **c.** Star phylogeny approximation.
b. 

\[ T_{CCA} \]

time of sampling

\[ T_{CCA} = \text{time to most common cellular ancestor} \]
**Figure 3-3.** A comparison of two different growth models for Red Rock (solid circles) and Riske Creek (open circles). If the model fits perfectly, the observed relationship between two estimates of divergence (solid) would fall on the expectation of a 1:1 relationship (dashed).  

**a.** Under a Wright-Fisher model of constant population size, the slope of this relationship is $0.478 \pm 0.145$ ($F_{1,25}=10.83$, p-value: 0.002971, $R^2=0.3023$).  

**b.** Under the star phylogeny approximation, the slope of this relationship is $1.099 \pm 0.329$ ($F_{1,25}=11.17$, p-value: 0.002613, $R^2=0.3089$).
Figure 3-4. Clone shape does not follow a perfect circle (DI=1 and E=1) in Riske Creek. This deviation may be largely due to increased perimeter (convolution) as a consequence of herbivory.
Figure 3-5. In Riske Creek, the presence of herbivory changes the area and morphology of a clone. a. Clone area significantly decreases with increasing cover of *Antennaria microphylla* ($F_{1,18}=6.749$, $p$-value=0.018, $R^2=0.27$). b. The extent of invagination increases with increasing cover of *Antennaria microphylla* ($F_{2,17}=5.607$, $p$-value=0.013, adj$R=0.33$).
Table 3-1. A table describing the linkage group and repeat motifs of the 14 microsatellite loci used in the study.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Linkage Group (P. trichocarpa)</th>
<th>Repeat Motif (observed in P. trichocarpa)</th>
<th>#Alleles Observed</th>
<th>Size Range Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMGC 486 (Pop3)</td>
<td>3</td>
<td>(GA)$_{n.a.}$</td>
<td>20</td>
<td>132-298</td>
</tr>
<tr>
<td>PMGC 510 (Pop4)</td>
<td>10</td>
<td>(GA)$_{n.a.}$</td>
<td>21</td>
<td>145-213</td>
</tr>
<tr>
<td>PMGC 575 (Pop 7)</td>
<td>1</td>
<td>(GA)$_{n.a.}$</td>
<td>16</td>
<td>172-230</td>
</tr>
<tr>
<td>PMGC 2274 (Pop15)</td>
<td>3</td>
<td>(GA)$_{n.a.}$</td>
<td>10</td>
<td>129-147</td>
</tr>
<tr>
<td>PMGC 2658 (Pop29)</td>
<td>13</td>
<td>(GA)$_{n.a.}$</td>
<td>15</td>
<td>237-277</td>
</tr>
<tr>
<td>WPMS17</td>
<td>7</td>
<td>(CAC)$_{15}$</td>
<td>11</td>
<td>139-204</td>
</tr>
<tr>
<td>ORPM 276</td>
<td>19</td>
<td>(TA)$_{6}$</td>
<td>18</td>
<td>128-228</td>
</tr>
<tr>
<td>GCPM 2768</td>
<td>2</td>
<td>(GA)$_{11}$</td>
<td>13</td>
<td>193-217</td>
</tr>
<tr>
<td>WPMS19 (lower)</td>
<td>?</td>
<td>unknown</td>
<td>19</td>
<td>178-250</td>
</tr>
<tr>
<td>WPMS19 (upper)</td>
<td>?</td>
<td>(CAG)$_{28}$</td>
<td>22</td>
<td>109-286</td>
</tr>
<tr>
<td>GCPM 1065</td>
<td>?</td>
<td>(AC)$_{9}$</td>
<td>12</td>
<td>145-209</td>
</tr>
<tr>
<td>PMGC 2731 (pop33)</td>
<td>?</td>
<td>(GA)$_{n.a.}$</td>
<td>29</td>
<td>150-254</td>
</tr>
<tr>
<td>PTR-2</td>
<td>?</td>
<td>(TGG)$_{8}$</td>
<td>14</td>
<td>213-246</td>
</tr>
<tr>
<td>PTR-3</td>
<td>?</td>
<td>(TC)$_{11}$</td>
<td>19</td>
<td>198-258</td>
</tr>
</tbody>
</table>
Table 3-2. Somatic mutations detected across 14 microsatellite loci in two populations
of *Populus tremuloides* (Riske Creek and Red Rock).

<table>
<thead>
<tr>
<th>Clone</th>
<th>Locus</th>
<th>Repeat type</th>
<th>Allele Change (1 or k-step) Or Homologous recombination (HR)</th>
<th>Mutant Ramet Alleles (bp)</th>
<th>Most Frequent Genotype (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>+ 1 step</td>
<td>- 1 step</td>
<td>+ k step</td>
</tr>
<tr>
<td>RC24-1</td>
<td>PMGC 575</td>
<td>di</td>
<td>2</td>
<td>x</td>
<td>178</td>
</tr>
<tr>
<td>RC9-1</td>
<td>PMGC 575</td>
<td>di</td>
<td></td>
<td></td>
<td>194</td>
</tr>
<tr>
<td>RC9-3*</td>
<td>PMGC 575</td>
<td>di</td>
<td>2</td>
<td>x</td>
<td>139</td>
</tr>
<tr>
<td>RC17-3</td>
<td>PMGC 575</td>
<td>di</td>
<td></td>
<td></td>
<td>176</td>
</tr>
<tr>
<td>RC23-1</td>
<td>PMGC 575</td>
<td>di</td>
<td></td>
<td></td>
<td>184</td>
</tr>
<tr>
<td>RC25-2</td>
<td>PMGC 575</td>
<td>di</td>
<td>3</td>
<td></td>
<td>184</td>
</tr>
<tr>
<td>RC15-1</td>
<td>PMGC 510</td>
<td>di</td>
<td>x</td>
<td></td>
<td>163</td>
</tr>
<tr>
<td>RC22-2*</td>
<td>PMGC 510</td>
<td>di</td>
<td>2</td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>RR4-1</td>
<td>PMGC 510</td>
<td>di</td>
<td></td>
<td></td>
<td>153</td>
</tr>
<tr>
<td>RC13-1</td>
<td>W19U</td>
<td>tri</td>
<td>x</td>
<td></td>
<td>229</td>
</tr>
<tr>
<td>RC16-1</td>
<td>W19U</td>
<td>tri</td>
<td></td>
<td></td>
<td>235</td>
</tr>
<tr>
<td>RR1-1</td>
<td>W19U</td>
<td>tri</td>
<td></td>
<td></td>
<td>238</td>
</tr>
<tr>
<td>RC17-3</td>
<td>O276</td>
<td>di</td>
<td>x</td>
<td></td>
<td>212</td>
</tr>
<tr>
<td>RC17-3</td>
<td>O276</td>
<td>di</td>
<td></td>
<td>9</td>
<td>194</td>
</tr>
<tr>
<td>RC20-1</td>
<td>O276</td>
<td>di</td>
<td>x</td>
<td></td>
<td>214</td>
</tr>
<tr>
<td>RC1-1</td>
<td>PMGC 486</td>
<td>di</td>
<td>2</td>
<td></td>
<td>136</td>
</tr>
<tr>
<td>RC3-1</td>
<td>PMGC 486</td>
<td>di</td>
<td>x</td>
<td></td>
<td>224</td>
</tr>
<tr>
<td>RC4-1</td>
<td>PMGC 2274</td>
<td>di</td>
<td>2</td>
<td></td>
<td>137</td>
</tr>
<tr>
<td>RR9-1</td>
<td>PMGC 2274</td>
<td>di</td>
<td>x</td>
<td></td>
<td>142</td>
</tr>
<tr>
<td>RC20-1</td>
<td>PMGC 2731</td>
<td>di</td>
<td></td>
<td>x</td>
<td>218</td>
</tr>
<tr>
<td>RC20-4</td>
<td>PMGC 2731</td>
<td>di</td>
<td></td>
<td>x</td>
<td>214</td>
</tr>
<tr>
<td>RR3-1</td>
<td>PTR-2</td>
<td>tri</td>
<td>x</td>
<td></td>
<td>218</td>
</tr>
<tr>
<td>RC13-1</td>
<td>PTR-2</td>
<td>tri</td>
<td></td>
<td>x</td>
<td>228</td>
</tr>
<tr>
<td>RC15-1</td>
<td>G2768</td>
<td>di</td>
<td>x</td>
<td></td>
<td>209</td>
</tr>
<tr>
<td>RC24-1</td>
<td>G2768</td>
<td>di</td>
<td>5</td>
<td></td>
<td>201</td>
</tr>
<tr>
<td>RC25-1b</td>
<td>W17</td>
<td>tri</td>
<td></td>
<td></td>
<td>154</td>
</tr>
<tr>
<td>RC13-1</td>
<td>PMGC 2658</td>
<td>di</td>
<td></td>
<td></td>
<td>251</td>
</tr>
<tr>
<td>RC15-18</td>
<td>PTR-3</td>
<td>di</td>
<td>x</td>
<td></td>
<td>230</td>
</tr>
<tr>
<td>RR9-1</td>
<td>G1065</td>
<td>di</td>
<td></td>
<td></td>
<td>165</td>
</tr>
</tbody>
</table>

*Not included in any further analysis because the number of ramets sampled for that clone was below five.*
Table 3-3. a. Different linear models for the relationship between $\pi_k$ and clone size metrics area, perimeter (P) and maximum distance between two ramets ($D_{\text{max}}$) for two populations of aspen. b. The anova table for the full model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fstat</th>
<th>df</th>
<th>AIC</th>
<th>Estimates (s.e.)</th>
<th>Log(Area)</th>
<th>Log(P)</th>
<th>Log($D_{\text{max}}$)</th>
<th>adjRsq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RISKE CREEK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept Only</td>
<td></td>
<td></td>
<td></td>
<td>0.072 (0.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=0.00030</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Model</td>
<td>0.20</td>
<td>3,16</td>
<td>-39.54</td>
<td>-0.059 (0.19)</td>
<td>-0.0038 (0.039)</td>
<td>0.026 (0.077)</td>
<td>0.0045 (0.026)</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>p=0.90</td>
<td></td>
<td></td>
<td>p=0.76</td>
<td>p=0.92</td>
<td>p=0.74</td>
<td>p=0.87</td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>0.51</td>
<td>1,18</td>
<td>-43.37</td>
<td>0.00083 (0.101)</td>
<td></td>
<td>0.0098 (0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.48</td>
<td></td>
<td></td>
<td>p=0.99</td>
<td></td>
<td>p=0.48</td>
<td></td>
<td>0.028</td>
</tr>
<tr>
<td>Model B</td>
<td>0.64</td>
<td>1,18</td>
<td>-43.50</td>
<td>-0.046 (0.14998)</td>
<td></td>
<td>0.023 (0.028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.44</td>
<td></td>
<td></td>
<td>p=0.76</td>
<td></td>
<td>p=0.44</td>
<td></td>
<td>0.034</td>
</tr>
<tr>
<td>Model C</td>
<td>0.32</td>
<td>1,18</td>
<td>-43.16</td>
<td>0.0204 (0.092)</td>
<td></td>
<td>0.011 (0.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.58</td>
<td></td>
<td></td>
<td>p=0.83</td>
<td></td>
<td>p=0.58</td>
<td></td>
<td>0.018</td>
</tr>
</tbody>
</table>

<p>| <strong>RED ROCK</strong> |       |     |      |                  |           |       |                     |        |
| Intercept Only |       |     |      | 0.088 (0.032)    |           |       |                     |        |
|                |       |     |      | p=0.032          |           |       |                     |        |
| Full Model     | 3.444 | 3,3 | -16.35| -0.950 (0.50694) | 0.056 (0.028) | 0.53 (0.32) | -0.47 (0.307) | 0.55   |
|                | p=0.1685 |     |      | p=0.16           | p=0.14   | p=0.202 | p=0.22             |        |
| Model A       | 5.725 | 1,5 | -15.25| -0.097 (0.081)   | 0.032 (0.013) |           |           | 0.53   |
|                | p=0.06218 |     |      | p=0.28           | p=0.06   |           |           |        |
| Model B       | 4.398 | 1,5 | -14.32| -0.22 (0.150)    | 0.060 (0.029) |           |           | 0.47   |
|                | p=0.09008 |     |      | p=0.20           | p=0.090  |           |           |        |
| Model C       | 4.389 | 1,5 | -14.32| -0.13 (0.108)    |           |           | 0.052 (0.025) | 0.47   |
|                | p=0.09033 |     |      | p=0.27           |           |           | p=0.090  |        |</p>
<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>pvalue (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RISKE CREEK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogArea</td>
<td>1</td>
<td>0.0028</td>
<td>0.0028</td>
<td>0.46</td>
<td>0.507</td>
</tr>
<tr>
<td>LogP</td>
<td>1</td>
<td>0.00066</td>
<td>0.00066</td>
<td>0.11</td>
<td>0.747</td>
</tr>
<tr>
<td>LogD$_{\text{max}}$</td>
<td>1</td>
<td>0.00018</td>
<td>0.00018</td>
<td>0.030</td>
<td>0.865</td>
</tr>
<tr>
<td>Residual</td>
<td>16</td>
<td>0.098</td>
<td>0.0061</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RED ROCK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LogArea</td>
<td>1</td>
<td>0.023</td>
<td>0.023</td>
<td>7.11</td>
<td>0.0758</td>
</tr>
<tr>
<td>LogP</td>
<td>1</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.84</td>
<td>0.428</td>
</tr>
<tr>
<td>LogD$_{\text{max}}$</td>
<td>1</td>
<td>0.0075</td>
<td>0.0075</td>
<td>2.38</td>
<td>0.221</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>0.0095</td>
<td>0.0032</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Chapter 4: **ON THE ESTIMATION OF CLONE AGE IN THE PRESENCE OF AGE-STRUCTURE AND DENSITY-DEPENDENT GROWTH**.

*A version of this will be submitted for publication.

INTRODUCTION

The demographic history of a population changes the structure of gene genealogies, subsequently affecting extant patterns of genetic variation. Previous research, for example in humans and domesticated maize, for example have made retrospective inferences about past demographic events using genetic variation at mitochondrial DNA and microsatellites (Eyre-Walker et al., 1998; Rosenberg and Hirsh, 2003; Slatkin and Hudson, 1991; Stumpf and Goldstein, 2001; Thuillet et al., 2005). This type of molecular genetic data is ideal because different demographies leave specific signatures on the distribution of site differences between pairs of sampled DNA sequences and on the distributions of allele frequencies (Rogers and Harpending, 1992; Slatkin and Hudson, 1991). For example, episodes of population growth result in unimodal peaks in the frequency of pairwise site differences with the height of the distribution corresponding to the timepoint where the expansion occurred. By comparison, population size reductions generate L-shaped distributions where most pairs exhibit few if any site differences (Rogers and Harpending, 1992).

Likewise, in clonal plants, which consist of a collection of clonal lineages distinguished by somatic mutations, similar inferences regarding clonal history might be made. Each lineage within a clone is represented by a reiterated module or ramet produced by mitotic growth, and each lineage has its own life history and physiological age. The patterns of survivorship and reproductive ability among all lineages will generate a unique genetic signature in the clone as a whole. By using a coalescent framework, neutral genetic divergence accumulated in the ramets during the lifespan of a clone may be used to
approximate the time to the most recent common cellular ancestor, the seed or the age of the clone \(T_{CCA}\) (Chapter 3).

Neutral divergence among ramets within a clone may be estimated by the number of pairwise differences in alleles between any two ramets within a clone, \(\pi_k\). Under a Wright-Fisher model of population size, \(\bar{\pi}\), estimates the expected number of differences separating two lineages/ramets, \(4N_c\mu\) (Tajima, 1989b). If the mutation rate, \(\mu\) and effective population size, \(N_e\) are known, then the time to most recent common ancestor can be estimated. This is not true for a star phylogeny model, where the time to a common cellular ancestor in the case of a clone, \(T_{CCA}\), i.e. clone age, is estimated using the average number of pairwise genetic differences (Slatkin, 1996), \(\bar{\pi}_k\), among ramets of the \(k^{th}\) clone and the mitotic mutation rate, \(\mu_e\), in the following formula

\[
T_{CCA} = \frac{\pi_k}{2\mu_e}
\]

(for details on this derivation see Chapter 3).

This model, however, assumes that gene lineages within a clone follow a star phylogeny. For clones, a star phylogeny is produced when the number of ramets within a clone grows (for details on clonal growth see Chapter 3). It is the expansion in ramet numbers that creates a gene genealogy that is star like, where the internal branches are short and the external ones are relatively long (Slatkin and Hudson, 1991). Exponential growth is a likely model of growth for a clone in its early years of growth because it is a single seed that initiates the clone.
The assumption of a star phylogeny representing relationships among ramets, however, may be problematic if clones grow slowly relative to the turnover rate of ramets. Rosenberg and Hirsh (2003) showed that unless a population grew such that $N \cdot r$ grew to exceed $10^5$, where $r$ is the intrinsic rate of growth and $N$ is the population size, then a star phylogeny assumption resulted in a downward bias in the pairwise estimator, $\hat{\pi}$. They further demonstrated that for a population of constant size, the assumption of a star phylogeny will lead the pairwise estimator, $\hat{\pi}$, to underestimate the time to a common ancestor by 50%. If, on the other hand, growth is fast, then coalescent events will mostly be near the root, hence producing less bias in the estimator. Thus, clone age estimation may be affected by the population dynamics and demography of its modules.

Populations are rarely, if ever, at a constant size or in a constant state of exponential growth. For many plant communities natural disturbance is a mechanism that fundamentally alters densities and age-structures of populations over time and space. One long-lived clonal plant for which this mechanism is particularly important is trembling aspen or *Populus tremuloides*, a disturbance-dependent species. Fire-mediated disturbances have been shown to impact the age-structure of a stand, competitive interactions with conifers, and the dynamics of clonal reproduction (Keyser *et al.*, 2005; Shepperd, 2001). Fires can have a widespread impact on the landscape, for example the Yellowstone fires in 1988 burned nearly 700 000 ha (Stevens *et al.*, 1999). Not only does fire impact mortality but trembling aspen exhibits periods of accelerated differential mortality in response to higher ramet densities, resulting in waves of extinction that
reduce the overall density of juveniles or suckers in a clone (DeByle, 1985; Navratil, 1990). Moreover, there is evidence that ramet mortality may be tied to the combined effects of insects and fungi, which are spatially variable across a clone (Steneker, 1976). This evidence then implies that there is high turnover rate of the ramets within clones.

If the overall population size of ramets within a clone fluctuates substantially and widely as might happen under cyclic disturbance regimes like fires, and if somatic mutations accumulate as a clone ages, then clones within a population might be expected to exhibit a high variance in clone age estimates (or $T_{CCA}$) even if they belong to the same age cohort. This is because genetic drift governs the loss of neutral variance within a clone, resulting in a high level of genetic variation in some clones but very little in others. Clones, such as $P.tremuloides$, however, may retain neutral variance because even if ramets are destroyed by fires, the roots remain alive but dormant (Barnes, 1966; DesRochers and Leiffers, 2001). If most of the genetic variation is generated via cell divisions within the roots and the shoots or aboveground mass dies, genetic variation can be regenerated upon ramet regrowth. On the other hand, if the root systems are damaged, then age-structure might serve as an alternative mechanism by which clones may retard the loss of neutral genetic variance. By partitioning the variation among different age classes and given sufficient overlap among the generations, genetic variation can persist for longer periods of time (Ellner and Hairston, 1994).

There have been numerous explorations of how molecular data (DNA sequence) is impacted by different models of population growth including a constant population size,
exponential growth, and logistic growth (Austerlitz et al., 1997; Di Rienzo et al., 1998; Ramos-Onsins and Rozas, 2002; Rienzo and Wilson, 1991; Rogers and Harpending, 1992; Slatkin and Hudson, 1991). Additional demographic characteristics like age-structure, however, have not been explored. Age-structure is characteristic of organisms with long-lived life history stages like trees. Previous work has demonstrated that age-structure can influence the partitioning of genetic variation in populations (Epperson and Allard, 1989; Epperson and Alvarez-Buylla, 1997; Schnabel and Hamrick, 1990). One study found spatial genetic structure present among young progeny but absent in older age classes, which showed little evidence of spatial correlations among allele frequencies (Epperson and Allard, 1989). The apparent discrepancy was explained by age-specific responses to density dependence. Thus, any estimation of genet age must consider how temporal structure and demography of ramets within a clone (clone growth through time) interact to affect current levels of genetic variation.

Here we use a modified version of an individual-based, genetically explicit stochastic simulation program, Nemo (Guillaume and Rougemont, 2006), to determine if age-structure might impact estimates of $T_{CCA}$ in ramet populations subject to density dependence. We consider the age-structure or age classes of ramets within a clone. Specifically, our goal is to determine how age-structure affects the pairwise estimator, $\hat{\pi}_k$, as a predictor of genet age. We focus on $\hat{\pi}_k$ because this estimator is used in this thesis to estimate genet age in natural populations of *Populus tremuloides* (Chapter 3 and Chapter 5), assuming genealogical relationships among ramets within a clone are star-like.
METHODS

In Figure 1-4 we outline the life cycle events and the transitions of the Leslie matrix used to model the population dynamics of ramets within clones and the pattern of neutral mutation accumulation. Simulations were run for 50,000 iterations and pairwise differences among a sample of ramets within a clone was measured at 40 microsatellite loci. Ramets within a clone were sampled randomly. In our simulations, microsatellites followed a j-step mutation model. To estimate \( \pi_k \), for the \( k^{th} \) replicate, we averaged the pairwise differences at 20 microsatellite loci across all sampled ramets (N=40). Although four possible pairwise comparisons are possible among the alleles between any two ramets, only two of those are relevant to our estimate of \( \pi_k \). This is because we compared alleles that were maternally inherited and alleles that were paternally inherited between any two ramets. Thus, the maximum number of differences that any two individuals can have is 40. We thus divide the total number of differences by 40 so that \( \pi_k \) measures the probability that two ramets differ in an allele inherited from the same parent. We present the results from 10 clonal replicates. In addition, if all ramets within a clone died, because of demographic stochasticity then a new genotype was reseeded. Thus, clone age represents the actual age of the surviving genet.

We compared two different demographic models: one with and one without age structure. In the model without age-structure, we started the simulations at carrying capacity (N=10000), every ramet in the clone reproduced but then immediately died. The simulations were of a single clone with N ramets, however, each ramet could reproduce. Thus every iteration is equal to 1 year. In the alternative model of age-structure, we sub-
divided it into two phases of growth: exponential and a stationary phase. To define the bounds of exponential growth, we determined the point during the simulation that carrying capacity was reached. A stable age distribution was reached on average by iteration 3410 (Figure 2-4) while carrying capacity was reached at iteration 4300. Thus, we considered the first 4300 iterations the exponential growth phase while after iteration 4400 was considered the stationary or density-dependent growth phase. The process of population regulation selected a number of individuals from each age-class to survive, proportional to the fraction of the total population each age-class comprised. These ramets made up the clone’s population in the next generation maintaining the total number of ramets within a clone at a constant population size of N=10,000.

The unit of time is not straightforward in our simulations because of the presence of age-structure. For our simulations under the model without age-structure, all individuals reproduced and died after a single iteration, thus a generation or the mean age at reproduction is one year. For age-structured populations, however, the mean age at reproduction for a ramet is not one year because ramets can continue to reproduce as they age (Charlesworth, 1994). Further, in a growing population, reproductive value among different age classes is not equivalent because ramets that have offspring earlier will leave more descendant ramets than ramets reproducing later. Because we are investigating the properties of age-structure, we used generation time as our unit of time because it is biologically more meaningful than an iteration. Thus, following Charlesworth (1994) we used the following formulae to calculate a generation time for the exponential growth phase \( T_g \) and the density-dependent phase \( T_s \),
\[
T_g = \sum_{0}^{500} (x + 1) \cdot e^{-r(x+1)} l(x) \cdot m(x + 1) \quad (1)
\]

\[
T_s = \frac{\sum_{0}^{500} x \cdot l(x) \cdot m(x + 1)}{\sum_{0}^{500} l(x) \cdot m(x + 1)} \quad (2)
\]

where \(x\) is the age of the ramet, \(r\) is the intrinsic rate of growth \((r = \ln \lambda, \lambda\) is the leading eigenvalue of our Leslie matrix (Figure 1b), \(l(x)\) is the schedule of survivorship, and \(m(x)\) is the fecundity schedule. We chose 500 years as an upper limit because it has previously been demonstrated that some ramets of *Populus tremuloides* can live to be 300 years old (Chapter 4). Thus, during the growing phase of our age-structured clones, one generation equaled \(~64\) yr, while during stationary phase it equaled \(~80\) yrs (Charlesworth, 1994).

In order to compare the age-structured stationary growth model to clonal replicates without age-structure, we rescaled time and the mutation rate (see Figure 1-4). One generation in the age-structured populations is equivalent to 80 iterations, while one generation in the clones without age-structure equals one iteration. *Populus tremuloides* populations in the western part of its range (British Columbia and Alberta) are unlikely to be older than 8000 years old or 100 generations (MacDonald and Ritchie, 1986).
RESULTS AND DISCUSSION

As expected (Rosenberg and Hirsh, 2003), under exponential growth, \( \hat{\pi}_k \) increases linearly with the true age of the genet, making it a reasonable predictor of clone age (Figure 3-4a). This was true regardless of the mutational scenario, i.e., mutations in both adults and clonal offspring or mutations in clonal offspring only. A linear regression of \( \hat{\pi}_k \) on true genet age showed that when mutations were allowed to accrue both in adult ramets (shoots) and in their offspring (roots), the slope of the relationship was \( \beta_A = 9.58 \times 10^{-5} \) (S.E. = 2.59 \times 10^{-6}, p-value < 0.00001). This was not significantly different when mutations arose only in the offspring (roots) where the slope was \( \beta_B = 5.30 \times 10^{-5} \) (S.E. = 3.19 \times 10^{-6}, p-value < 0.00001). The difference in slopes is explained by the difference in the opportunity for mutation to generate pairwise differences among ramets. Early in the lifespan of a clone when most ramets arise from young ramets, the two scenarios are not different. However, as the clone ages and differences accumulate, the two scenarios diverge and the pairwise estimate, \( \hat{\pi}_k \) in clones where mutations accumulate only in lateral roots grows at a smaller rate (Figure 3-4a). A higher mitotic mutation rate in the root-only scenario will reduce the difference over time between the two mutational scenarios.

In order to determine if estimates of average pairwise distance is affected by demography, we compared the slope of a linear regression for ramet populations with age structure at two different phases of growth: growing (~ for 75 generations) and stationary (>75 generations). The 95% confidence intervals around the two slopes overlapped,
indicating that there was no detectable difference between the slope under an exponential growth model (95% C.I: 4.73x10^{-3} and 6.18x10^{-3}) and the slope of the linear regression for the first 1500 generations of density-dependent regulation (95% C.I: 4.25x10^{-3} and 5.30x10^{-3}). Frequency distributions of pairwise differences for populations experiencing exponential growth showed different signatures to those obtained under a stationary phase. Our distributions for a growing population were largely L-shaped, indicating more ramet pairs were identical, a signature found in populations which have undergone a population reduction (Rogers and Harpending, 1992). Plant clones are initiated by a single individual, the seed. Thus, it was not surprising to find most ramets were identical at most loci. In the stationary phase, most ramet pairs had accumulated some differences between them and exhibited one or several peaks (Figure 4-4b). This is unlike previous work which showed a distinct unimodal peak under exponential growth, however, exponential growth in those simulations were maintained for a long period of time while we were only able to maintain a growing phase for on average 4400 iterations or approximately 70 generations (Slatkin and Hudson, 1991).

To determine if age structure has an impact on average pairwise differences accumulated within a clone, we compared the frequency distributions of pairwise differences among ramets within a clone between the model with age-structure and without age-structure size. In the model without age-structure, early on a majority of ramet pairs are identical and distributions are L-shaped, but by generation 6000 most replicates exhibit a single peak. This indicates that the majority of ramet pairs are different at most of the loci (Figure 4-4a). Under age-structure and density regulation individual replicates exhibited a
ragged distribution with multiple peaks by iteration 500,000 (generation 6250). The raggedness of the distribution arises because allelic differences among ramet pairs are accumulating slowly over time (Figure 4-4b). In the age-structured model even at generation 6250 some ramet pairs were still identical while only 30% of the ramet pairs sampled were found to be different at less than half the alleles. These differences between the model with and without age structure are not a result of age-structure but more an artifact of the simulation. As a result of stochasticity in the simulations, early on some clones went extinct and had to be reseeded, thus resetting the time to zero and creating a lag. If these simulations were run for longer periods of time, eventually the frequency distributions of age-structured clones will look similar to the ones without age-structure.

We found minor differences between clones without age-structure and clones with age-structure in the linear fit of the relationship between $\hat{\pi}_k$ and the true genet age (Figure 4-5). The curvilinear relationship seen in Figure 4-5b is simply an artifact of the limited number of generations for the simulation run without age structure. With more than 6000 generations, this curvilinear pattern would disappear and the pattern would resemble Figure 4-5a. With no age-structure, a linear model explains very little of the variation in true genet age ($F_{1,412} = 0.8195$, p-value=0.3658, $R^2 = 0.002$). Similarly, we found the linear model was not a good representation of the relationship between the average pairwise difference and true genet age for clones with age-structure after generation 3000 ($F_{1,232} = 3.509$, p-value=0.0623, $R^2 = 0.011$). It is clear that the age-structured populations, although at a demographic equilibrium, require several million more iterations to reach a
mutation-drift equilibrium. Unlike clones with age-structure undergoing density-dependent growth, clones with no age-structure, reach an equilibrium value of $\hat{\pi}_k$. If age structured populations reached genetic equilibrium, as is the case in the model without age structure, there would be little to no discriminatory power in the pairwise estimator because older clones do not show larger values of $\hat{\pi}_k$ than younger clones. However, populations of $P. \text{tremuloides}$ clones used in previous work (Chapter 3 and Chapter 5) are unlikely to be older than 8000 years (Alley, 1976), thus a genetic equilibrium would be a biologically unlikely scenario.

Previous work has showed that the presence of age-structure can change how mutational variance is partitioned within a clone (Ellner and Hairston, 1994; Epperson and Allard, 1989; Epperson and Alvarez-Buylla, 1997; Schnabel and Hamrick, 1990). To explore whether there were significant differences among slopes in the different age classes for the relationship between $\hat{\pi}_k$ (over clonal replicates) and genet age, we performed an analysis of covariance. As before we found a significant relationship between the average pairwise estimator (averaged across replicates) and genet age ($F= 12786.912$, $p<2\times10^{-16}$). Age class did not seem to matter as the slopes of the different age classes were not significantly different ($F$-value=0.8845, $p$-value=0.4503). In addition, we found that for all age classes, variance increased with increasing genet age and the rate of increase was not distinguishable among age classes ($F=0.4099$, $p$-value=0.7461). The variance, shown in Figure 4-6, is created by sampling within and between ramets. This is because some sampled pairs exhibit no differences at microsatellite allele sets, while other lineage pairs are different at many alleles. In our simulations each age class had the
same growth rate thus resulting in an equal rate of increase in variance in each age class. One interesting avenue of exploration might be to investigate cases where the stable age distribution is approached more slowly to see how unequal growth rates among age classes affect the rate of increase in variance.

In the face of partial extinction or high mortality, as is the case in *P. tremuloides*, it is probably root dormancy or longevity of individuals that buffers the complete loss of genetic variation in the clone (Ellner and Hairston, 1994). Despite density-dependent responses like self-thinning, which result in high juvenile or sucker mortality, the presence of long-lived adults or dormant roots acts like a “storage container” of mutational variance. Thus, providing these lineages with the opportunity to persist through time in the older age classes or underground and allowing them to be passed on to the next generation. We found no effect of age class on the estimate of \( \hat{\pi}_k \) averaged over clonal replicates suggesting that sampling any age class of a clone would still lead to accurate estimates of age (Figure 4-6).

Estimates of clone age may be complicated by spatially explicit plastic growth responses to density dependence (Hamilton *et al.*, 1987). In one study of *Ranunculus reptans*, increasing intra-specific competition among clones resulted in a higher allocation to sexual reproduction than asexual reproduction, corresponding to clones growing in an intermingled growth form (guerilla formation) (Lovett Doust, 1981; Van Kleunen and Fischer, 2003). Alternatively, density dependence may result in a shift from sexual reproduction to vegetative reproduction (Fischer *et al.*, 2004) and a concomitant change...
to a phalanx growth habit, which recruits new ramets largely at the perimeter.

Differences in growth habit could potentially lead to differences in how lineages are distributed in a clone and ultimately estimates of age. If mutations occurred only in offspring and phalanx clones were sampled only inside the clone, it might lead to an underestimate of the true level of variation present in the clone. A spatially explicit model could tease apart how growth form and sampling strategy interact to affect estimates of clone age.

In conclusion, we have demonstrated that a star-shaped genealogy is an appropriate assumption in the estimation of clone age on age-structured clones subject to density-dependence and exponential growth.
Figure 4-1. **a.** The life cycle used in the simulations. Each patch was seeded with a single genotype. The sequence of events in the life cycle were asexual reproduction or cloning, followed by aging of ramets (survivorship in the current age class or transition to the next), and finally density-dependent regulation. The carrying capacity for all clones was set to N=10,000 ramets. Density-dependent regulation was carried out by randomly sampling 10,000 ramets, based on the proportion of the total population that each age class comprised. Two scenarios of mitotic mutation were explored. In **A** somatic mutations were allowed to accumulate in both the shoots and the roots, while in **B** somatic mutations arose only in the offspring, i.e., roots. In the model without age-structure, the mitotic mutation rate was set to \( \mu_c = 1.44 \times 10^{-6} \) per iteration. In the model with age-structure, we used the mitotic mutation rate, \( \mu_c = 1.152 \times 10^{-4} \), to correct for generation time differences. We used only the generation time at carrying capacity because of the limitations in the simulation. **b.** The Leslie matrix used for the simulation describes the population dynamics for each of the age classes. The total number of ramets produced by clonal reproduction was \( N_{tot,r} \) and was the product of the number of ramets at time \( t \) in each non-offspring age classes (\( N_{t,i} \)) and its associated fecundity (\( m_i \)). The total number of ramets in each age class was the number remaining in the current age class (\( s_{i,t} \)) plus those moving from the previous age class to the next age class (\( s_{i-1,t} \)). Note that only individuals that were old enough would move to the next class (e.g., 19 → 20 yr), because within a class individuals could age. Age classes were arbitrarily chosen.
a.

\[ N(t=0) = 1 \]

\[ \mu_c \]

\[ N_{tot,i} = \sum_{i>0} N_{t,i} \cdot m_i \]

\[ A & B \]

\[ E[N_{tot,s,i}] = \sum N_{r,i} \cdot s_{ii} + N_{r,i-1} \cdot s_{i-1,i} \]

\[ \sum \cdot + \cdot = \sum_{i} N_{r,i} \cdot s_{ii} + N_{r,i-1} \cdot s_{i-1,i} \]

b.

\[ m_0 = 0 \]

\[ S_{0,i} = 0.04 \]

\[ m_1 = 1.2 \]

\[ S_{1,2} = 0.90 \]

\[ 1-19\text{ yrs} \]

\[ S_{11} = 0.90 \]

\[ m_2 = 2.5 \]

\[ S_{2,3} = 0.96 \]

\[ 20-79\text{ yrs} \]

\[ S_{22} = 0.96 \]

\[ m_3 = 10 \]

\[ 80\text{ yrs} \]

\[ S_{33} = 0.99 \]
Figure 4-2. In our simulations 10 clonal replicates grow exponentially until a stable age distribution (SAD) is reached on average at iteration 3410 (range: 1200 to 7000 iterations). Carrying capacity is reached on average at 4300 iterations but ranges from 2200-7500 iterations. These clonal replicates were reseeded several times (resetting time to t=0 for genet age) until a single genotype was established. Here we plot the result from a single clonal replicate. The population was censused before regulation, thus the number of offspring individuals is above the carrying capacity of N=10,000.
Figure 4-3. A plot of $\hat{\pi}_k$, the genome-wide pairwise estimate of age against true genet age. We averaged the raw data both for $\hat{\pi}_k$ and genet age over 10 clonal replicates. 

a. During exponential growth, $\hat{\pi}_k$ is a reasonable predictor of clone age regardless of the mutational scenario explored, i.e., somatic mutations can arise in adults during growth and in clonal offspring and in clonal offspring only (see Figure 4-1). We rescaled the x-axis using the generation time which for stationary age-structured clones is approximately ~80 iterations. The mitotic mutation rate used in both cases was $\mu_c = 1.44 \times 10^{-6}$ per iteration.

b. Under density dependence with four age classes c. In the model without age-structure, population size is held constant at $N=10,000$. The ramet population in each clone consists only of one class.
Figure 4-4. Frequency distributions of average pairwise diversity at 20 microsatellite loci. A total of 10 clonal replicates were used in the simulations but here we present three sample replicates at one timepoint, i.e., Generation 6000. The right hand panel was corrected for generation time differences. a. Without age-structure. b. Age-structured clones with density-dependent growth. We compared pairs of ramets for both alleles across 20 loci and then averaged the number of differences across the total number of ramet pairs.

Average number of pairwise differences among sampled ramets
**Figure 4-5.** **a.** For clones without age-structure, $\hat{\pi}$ and genet age exhibit no relationship ($F_{1,412}=0.8195$, $p=0.3658$, $R^2=0.002$). **b.** Under age-structure, however, this relationship is explained by a linear model for only the first 3000 generations during exponential growth (solid line: $\beta_s=3.11x10^{-3} \pm 1.3x10^{-4}$, $F_{1,214}=567.4$, $p<2.2x10^{-16}$, $R^2=0.725$). After 3000 generations, the slope of the line is no longer significant ($F_{1,232}=3.509$, $p$-value=0.0623) and the regression model is a poor fit (dashed line: $\beta_s=-4.3x10^{-4} \pm 2.28x10^{-3}$, $R^2=0.011$). Here we have plotted the raw data from all 10 clonal replicates. We sampled 40 ramets within a clone every 100 generations and estimated the number of pairwise differences at 20 microsatellite loci among these sampled ramets. $\hat{\pi}$ was obtained by averaging the number of pairwise differences across all sampled pairs.
**Figure 4-6.** a. In age-structured clones with density dependence (600 generations), sampling only from one age class has little impact on the estimate of true genet age. b. Variance in the number of pairwise differences at 20 microsatellite loci among sampled ramets increases in all age classes with the true genet age. We present only the case where mutations occurred in both adults and offspring.
REFERENCES


Chapter 5: A genomic deleterious mutation rate for the loss of sex in a natural population of *Populus tremuloides*.

*A version of this manuscript will be submitted for publication.
Ally D, Ritland K, and Otto, SP. A genomic deleterious mutation rate for the loss of sex in a natural population of *Populus tremuloides*. 
Mutations are fundamental to evolution. Despite their importance, estimates of the rate and effects of mutation in nature are rare. Our knowledge about mutation rates and effects are restricted to small organisms with short generation times, often under controlled, benign laboratory conditions (Keightley and Caballero, 1997b; Kibota and Lynch, 1996; Mukai, 1964; Schultz, 1999; Shaw et al., 2000). Traditionally, mutational parameters are obtained by assaying fitness of derived lines from mutation accumulation experiments on organisms like *E. coli*, *Drosophila*, and *S. cerevisiae*. Here we report a novel approach that couples marker-based estimates of clone age with sexual fitness data to infer the rate and effect of mildly deleterious mutations for a woody tree species in the wild. We observed a significant decline in sexual fitness with increasing clone age in a natural population of aspen, *Populus tremuloides*. Mutations reduced relative sexual fitness in clonal aspen populations by about $0.12 \times 10^{-3}$ to $1.01 \times 10^{-3}$ per year. The deleterious mutation rate for sexual fitness in a natural population of aspen is estimated to be $U_{ML} = 2.43 - 9.95$ per diploid genome per year. Mutations that affect sexual fitness have an average effect of $\bar{a}_{ML} = 6.8 \times 10^{-4}$, suggesting most mutations are of small effects. By taking advantage of clonality, we obtained an estimate for the genomic deleterious mutation rate and an average effect size for the most widely distributed tree in North America, *Populus tremuloides*. Furthermore, the decline in sexual function with clone age, likely due to the accumulation of deleterious mutations, is evidence that clonal organisms are vulnerable to the effects of senescence.
Estimates of deleterious mutation rates in long-lived plants are rare. Rarer still are estimates of mutation rates and the magnitude of mutational effects on fitness in natural populations (Charlesworth, 1989). Yet mildly deleterious mutation rates are central to theory accounting for the evolution of sex (Kondrashov, 1988), ecological specialization (Cooper and Lenski, 2000), and small population viability (Gabriel and Burger, 2000). Despite almost two decades of research, we know of only one mutation rate estimate (7.4x10^{-3} per haploid genome per generation) for a woody tree species, obtained on wild populations of *Rhizophora mangle* from lethal mutants (Klekowski and Godfrey, 1989). This method assumed that highly selfing populations are at a mutation-selection balance, and required direct information on the genetic identity of parents and offspring, thus limiting the generality of the method. A second method, developed for highly self-fertilizing annual plant populations, yields estimates of mutation rates from measures of inbreeding depression (Charlesworth et al., 1990). This method assumes variation is maintained solely by a mutation-selection balance and lineages have historical levels of inbreeding, unlikely to be true and difficult to evaluate in natural populations. Unlike many herbaceous annuals and perennials, most woody tree species possess attributes like dioecy, outcrossing, an extended life cycle, a large stature, and many more cell divisions separating the germ line from the zygote (Petit and Hampe, 2006). The last two attributes make it impractical to employ any of the described methods to a woody tree species. Yet these same traits are what have led previous researchers to hypothesize that woody tree species have high genomic mutation rates (Klekowski and Godfrey, 1989; Schoen, 2005).
Here we report a novel approach to determine the rate at which sex is lost in the facultatively clonal and dioecious tree, *Populus tremuloides*. Clonal plants can continually produce new physiological individuals (ramets) without undergoing meiosis, thus emulating the process of a mutation accumulation experiment in the wild (see Figure 5-1). In traditional mutation accumulation experiments replicate lines derived from a single ancestor are propagated through time with repeated bottlenecks (Mukai, 1964). This results in an accumulation of weakly to moderate effect mutations that fix due to the action of drift (Lynch and Hill, 1986). By comparing the derived lines to ancestral or control populations, direct evidence of the accumulation of spontaneous mutations is obtained. As mutations accumulate through time lines diverge, resulting in the increase of phenotypic variance. Furthermore, if the majority of fitness effects are negative, mean fitness tends to decline over time.

To sufficiently mimic these MA experiments on a wild population, we had to meet three requirements: an environment where natural selection is relaxed, independently maintained lines with a common origin, and some measure of time. All these conditions were approximated in our experiment. First, during clonal growth, natural selection is relaxed for traits involved in sexual function. Thus, we measured sexual fitness as the mean number of viable pollen grains per catkin per ramet. Second, distantly spaced ramets within a clone should accumulate somatic mutations independently (Figure 5-1). In order to avoid closely related ramets, we sampled systematically at intervals (every 3-5m) along multiple transects. Clone age was estimated using two different units of time: the number of cell divisions since the origin from the zygote and years, by assaying neutral changes within clones at 14 microsatellite
Among a sample of 715 trees across 20 clones in a natural population of aspen, we find a significant directional decline in mean clonal sexual fitness with clone age. We attribute this to mutation accumulation. To rule out alternative factors that may be the underlying cause of a decline in sexual fitness, we examined date of collection (F_{2,94} = 2.243, p-value= 0.1118), tree age (F_{1,93}=0.01306, p-value=0.9093, R^2=0.00014), and inbreeding level as measured by proportion of homozygous loci (F_{1,18}=1.125, p-value=0.3029, R^2= 0.06) and found they did not explain significant variation in sexual fitness (SFigures 5-3 to SFigure 5-5).

Empirical studies suggest that the presence of fungal pathogens and insect herbivory can exert a strong influence on reproductive success (Frey et al., 2003). Thus, morphological characteristics indicating the health of a tree were assayed, and we performed vegetation and site analysis in each patch. Eleven different disease variables were reduced to four composite tree health measures, while plant understory cover and abiotic site variables were reduced to eight environmental indices using principal component analysis (Supplementary Information). To identify a subset of potential predictor variables, we examined the relationship between mean sexual fitness and our abiotic and biotic variables using a correlation analysis. These variables were then put into a stepwise multiple regression analysis. Subsequent model selection was based on AIC criterion, p-values, C_p criterion, and adjusted R^2 values. The final multiple regression model included two predictors: moisture gradient (PC2) and clone age.
When the effects of the environment (moisture gradient) are held constant, sexual fitness declined significantly with clone age (Figure 5-2). Although we have tried to account for confounding factors one of the limitations of studies in the wild is the inability to conclusively rule out other environmental factors such as historical environment.

We estimated the rate of relative sexual fitness decline per cell generation and per year from the untransformed partial regression coefficients for clone age. We found the per cell generation decline was $-1.14 \times 10^{-5}$ (95% CI: $2.04 \times 10^{-5}$ to $1.44 \times 10^{-6}$). Setting the maximum clone age to 8000 years old given the history of glaciation in the area (Alley, 1976), we obtained a lower bound estimate for the rate of decline in sexual fitness per year of $-1.15 \times 10^{-4}$ (95% CI: $-2.05 \times 10^{-4}$ to $-2.46 \times 10^{-5}$). Setting the minimum clone age to 77 years (based on tree ring data for extant ramets), we obtained an upper bound estimate for the rate of decline in sexual fitness per year of $-1.01 \times 10^{-3}$ (95% CI: $-1.81 \times 10^{-3}$ to $-2.17 \times 10^{-4}$) (see Supplementary Information). Thus, mutations cause an average decline in sexual fitness of 0.001% per cell generation or between 0.02-0.1% per year, which reflects the product of the genomic mutation rate, $U$, and the average effect on sexual fitness in a heterozygote, $\bar{a}$. Assuming a constant linear decline, this means it would take ~1000-5000 years to completely lose sexual function with respect to pollen quantity and quality. Further, we estimate clones had a 11% reduction in pollen fitness due to mutation accumulation.

One caveat is that the per cell generation estimate depends on an accurate estimate of the neutral somatic microsatellite mutation rate because as the number of cell divisions between ramets. Every 10-fold reduction in the magnitude of the somatic
mutation rate results in a 10-fold change in the estimate of cell divisions and subsequently the deleterious mutation rate per cell generation (see Supplementary Information). This was not true for the per year measure, which was insensitive to the uncertainty in the neutral mitotic mutation rate. Furthermore, we found the 95% confidence intervals of the per year rate of decline were overlapping, suggesting less uncertainty in this estimate. Thus, we proceed to estimate the minimum genomic mutation rate using the estimate of clone age in years.

We did not find a significant increase in the variance in pollen fitness among ramets, $\sigma^2_M$, within clones as a function of clone age ($F_{1,18} = 1.347$, p-value = 0.26, $R^2 = 0.069$), but this is not uncommon in MA experiments. Several MA accumulation experiments have found significant decreases in mean fitness over time with no statistically significant increase in variance among lines (Fry, 2002; Keightley and Caballero, 1997a; Schoen, 2005; Shabalina et al., 1997). It is possible that environmental variance masks the relatively small changes in variance caused by mutation. Thus, with the maximum clone age at 8000 years old, we found the lower bound on the point estimate of $\sigma^2_M$ was $-1.41 \times 10^{-4}$ (S.E.M. = $1.22 \times 10^{-4}$; 95% CI from $-3.9 \times 10^{-4}$ to $1.14 \times 10^{-4}$). The upper bound was found assuming clones were a minimum of 77 years old, resulting in an estimate of $\sigma^2_M$ of $1.2 \times 10^{-3}$ per year (S.E.M. = $1.1 \times 10^{-3}$; 95% CI from $-3.5 \times 10^{-3}$ to $1.0 \times 10^{-3}$).

Based on this data, we provide an upper ($U_{BM,Upper}$) and lower bound ($U_{BM,Lower}$) for the minimum genome-wide mutation rate, $U_{min}$, using the Bateman-Mukai (BM) method (Bateman, 1959; Mukai, 1964). Assuming mutations have equal effects, a
minimum estimate of the total deleterious genomic mutation rate and a maximum estimate of the mean mutational effect, $a_{BM}$, are

$$U_{BM} = \frac{2⟨Ua⟩}{σ_M^2} \quad \text{and} \quad a_{BM} = \frac{σ_M^2}{Ua}. \quad (1)$$

Although the best estimates of $Ua$, average decline in fitness, and $σ_M^2$ are typically used to infer $U_{BM}$, a more conservative estimate of the minimum mutation rate is provided by using the lower bound of the 95% CI for $Ua$ (-2.46 x 10^{-5} and -2.17 x 10^{-4}) and the upper bound of the 95% CI (1.14 x 10^{-4} and 1.0 x 10^{-3}) for $σ_M^2$. Doing so, we estimate $U_{BM, Upper} = 1.06 x 10^{-5}$ and $U_{BM, Lower} = 9.4 x 10^{-5}$ per diploid genome per year. As we did not find a significant increase in variance among lines, we did not estimate a maximum effect size using BM. Our estimate of $U_{BM}$ is an underestimate of the genome-wide deleterious mutation rate for several reasons. First, $σ_M^2$ was not significantly different from zero, which would yield an estimated value of $U_{BM} = \infty$. Furthermore, the BM method assumes that mutations are equal in effect size, which causes an underestimate of the genome-wide deleterious mutation rate.

This bias can be avoided by using a maximum likelihood (ML) approach where the shape of the distribution of mutational effects is estimated (Keightley, 1994). Previous ML approaches (Keightley, 1994) and associated C program (mlgenomeu.c) can only be implemented on typical lab-based mutation accumulation experiments, where two time points (t=0 and t=end of MA) are compared for multiple lines originating from a single isogenic ancestor. The assumption of a single isogenic ancestor, however, is not justified with our data because groups of ramets (multiple lines) are seeded by different genotypes. Secondly, each of these genotypes usually has a different clone age. Finally,
a direct comparison to an ancestor or control is not possible, so both the ancestral mean (at time zero), \( M \), and environmental variance, \( V_E \), were estimated from ML. With a few modifications to fit the current data set, we wrote a program in R (version 2.4.1. : R Development Core Group, 2007) that used ML to estimate the genomic mutation rate, \( U_{ML} \), shape parameter, \( \alpha \), and the scale parameter, \( s \). Our likelihood estimate, given the maximum age of a clone is 8000 years old, was \( U_{ML, Lower} = 2.43 - 9.85 \) per diploid genome per year and the average effect size of, \( a_{ML} = 6.8 \times 10^{-4} \) (see Supplementary Information for details on likelihood estimation).

Although we have assumed that natural selection on sexual function is relaxed during asexual reproduction, mutations deleterious to one aspect of fitness may fix because they benefit some other aspect of fitness. Thus, mutations that confer poor sexual function to a clone may proliferate through enhanced clonal fitness. If selection is facilitating the loss of sex through the action of antagonistic pleiotropy, then we would expect to see a trade-off between sexual fitness and asexual fitness (Cooper and Lenski, 2000; Dorken et al., 2004). We looked for evidence of this trade-off at two levels: ramet and clone. First, we asked whether mutations accumulating in ramet lineages reduced sexual fitness while increasing overall ramet growth. We performed a correlation analysis on ramet sexual fitness and volume growth per year (cubic cm/yr) but found no evidence of a relationship (\( \alpha \) level : 0.016, \( t = 1.2097 \), df = 93, p-value >0.05, \( \rho = 0.12 \)). Nor did we find evidence for a trade-off at the level of the clone between above-ground clonal growth rate and mean clonal sexual fitness (\( t = -0.6 \), df = 18, p-value = 0.556, \( \rho = -0.14 \)). We did observe a negative relationship between clone size and mean clone sexual fitness, but it was not significant after a correction for multiple comparisons was applied.
(t = -2.3902, df = 18, p-value = 0.03, \( \rho = -0.49 \)). Trade-offs due to resource allocation decisions, however, may not be apparent if sexual and clonal structures involved in reproduction do not directly compete for the same resources at the same time (Thompson and Eckert, 2004).

Prior to this study, all MA studies in plants have been conducted on plants that have annual life cycles, e.g., Arabidopsis and Amsinckia (Johnston and Schoen, 1995; Schoen, 2005; Schultz et al., 1999; Shaw et al., 2000). This is the first study, to our knowledge, that has calculated a mildly deleterious mutation rate on a long-lived plant species. Nevertheless, our results fall within the range of mutation rate estimates found in other species (Table 5-2).

There are two reasons for why our rate of fitness decline and the deleterious mutation rate are underestimated in this study. First, if intra-organismal selection (within ramets) (Otto and Orive, 1995) has removed mutations that are deleterious to both sexual and asexual fitness, this will greatly reduce observed mutation rates. Secondly, our measure of sexual fitness, total number of viable pollen grains per ramet, captures only a subset of possible sexual function mutations. Pollen grains perform a diverse array of functions necessary for survival, including rapid growth during pollen tube elongation and cell-cell communication during pistil-pollen interactions (da Costa-Nunes and Grossniklaus, 2003). Our assay would not have detected mutations affecting these functions, only pollen count and viability. Again this would lead to underestimating the true genome-wide rate of mutations deleterious to sexual function.

Higher mutation rates may be expected in long-lived perennials, in part, because of the lack of distinction between the soma and germline (Buss, 1983). For plant clones,
where there is an increase in the number of apical meristems with clonal growth, somatic mutations can accumulate as the clone ages. Although somatic mutations are not immediately life-threatening, a high genetic load ultimately comes at a long term cost to the clone’s sexual function. Hamilton’s assertion that senescence is an inevitable outcome of evolution,(1966) is evidenced by our significant decline in sexual fitness with age. Plant clones have a limited time span within which sexual function remains high. The clones that we examined have lost on average 11% of their sexual fitness, with the oldest clone having only half the pollen viability. In the absence of sex, mutational decay in clonal sexual fitness may play a role in increasing the clone’s risk of ultimate extinction for the genetic lineage.

**METHODS AND MATERIALS:**

*Species description:*

*Populus tremuloides* or trembling aspen, is a member of the Salicaceae, is a dioecious, wind-pollinated tree. All members (ramets) of a clone have either pistillate or staminate flowers. Sexual reproductive maturity is reached between 10-20 years of age. Individual reproductive shoots produce inflorescences (catkins) that often have between 80-100 flowers (Fisher, 1928; Lester, 1963; Nagaraj, 1952). Aspen reproduces clonally by lateral root suckers initiated from lateral root primordia (Barnes, 1969). In aspen, root apical primordia are abundant; in one 10cm root cutting, 200 emerging primordia were found (Schier, 1972).

*Measure of Clone Age:*

We examined 14 microsatellite loci for somatic changes among sampled ramets within a clone. Clones were sampled in two ways: on the perimeter of a grove or patch
of aspen and along two or three 30-50m transects within a patch. Systematic sampling was used as it would best represent the distribution of different genotypes found in that patch. On average 30-50 ramets were sampled per patch. We sampled a total of 715 trees in 17 patches in Riske Creek, in the Cariboo-Chilcotin region of Interior British Columbia. No tree less than 1.5m was sampled and patches were separated by at least 1 km of terrain lacking aspen trees. Trees on the perimeter and along transects were physically mapped using both a measuring tape and a handheld GPS unit. From the GPS points we calculated the area of a patch using ARCVIEW GIS 3.3 (2002 Environmental Systems Research Institute).

A detailed description of how somatic mutations were scored can be found elsewhere (Chapter 3). We estimated clone age using two different units of time: cell generations and years. In both cases, we assumed relationships among ramets can be approximated by using a non-equilibrium model or a “star phylogeny”, where all lineages (ramets) are independent replicates of the time since the most recent common ancestor, the seed. Although this is not strictly true, it is likely to be a reasonable approximation for a growing clonal population especially because we sampled systematically at intervals along multiple transects.

Under standard coalescence theory a population of $N$ diploid individuals replicating every generation, the expected time to coalescence between any two alleles would be $2N$ generations, and the probability of a neutral mutation in either one of the two lineages would be $2\mu$ per generation (Hudson, 1990b). In a star phylogeny, on the other hand, the probability of a neutral mutation in either one of two ramets remains $2\mu$ per generation, but the time until coalescence becomes, $T_{CCA}$, the age of the $k^{th}$
clone. Thus, the average pairwise genetic distance, $\pi_k$, i.e., over all pairs of ramets within the $k^{th}$ clone can be used to estimate $2\mu T_{CCA}$. Thus for the $k$ th clone, $\pi_k$, the average pairwise genetic distance is,

$$\pi_k = \frac{\sum_{l=1}^{N} \sum_{j=1}^{L} \delta_j}{N} \frac{2 \times L}{N} \tag{1}$$

where $N$ is the number of unique pairwise comparisons among different ramets within a single clone, $L$ is the number of loci, $\delta_j$ is an indicator variable for the $j^{th}$ locus and is either 1=alleles different, 0=alleles same. Previous research has shown that somatic microsatellite mutations can be used in the reconstruction of cell lineage trees from human and mice embryos with a very high degree of accuracy (Frumkin et al., 2005). This suggests that the numbers of mutations found at microsatellites might track the numbers of cell divisions reasonably well. Clone age, $T_{CCA} = \frac{\pi_k}{2\mu}$, in cell generations is highly sensitive to the choice of microsatellite mutation rates per cell division, $\mu$. We obtained a mitotic microsatellite mutation rate of $4.06 \times 10^{-7}$ per cell given an estimated number of cell divisions (Chapter 3).

The second method allowed us to obtain an estimate of clone age in years. Pollen records suggest $P. tremuloides$ was first present in southern British Columbia approximately 8000 years BP (Alley, 1976; Cawker, 1983). Thus, to obtain a lower bound for the rate of decline in sexual fitness per year ($\Delta \omega_{\text{lower}}$) in Riske Creek, we used the clone with the most neutral divergence, $\pi_{RC24-1}$, and assumed it was a maximum age of 8000 years old. An upper bound on the rate of sexual fitness decline per year...
(Δω_{Upper}) was obtained using tree ring counts from increment cores. We excluded the clone with the highest value of π_k because we were interested in determining the minimum clone age. Observed minimum values of π_k, however, were zero, due to a limited number of loci studied. Because we could not set the age of the clone to zero, we took the average the remaining values of π_k. We coupled the average value of \bar{\pi} with the average across all clones of all the oldest ramets. Thus, our conversion factor was the average oldest ramet per average estimate of \bar{\pi}. For inclusion in our analysis, clones must have had a minimum of five ramets as assigned by the genetic analysis. This was because clone age estimates are subject to error, especially in clones with few ramets.

Measure of sexual fitness:

We chose to focus on male sexual fitness because the amount and quality of pollen produced is a crucial feature of plant fitness. Secondly, sexual function is a complex trait, likely to be governed by many loci and thus provides a broad target for mutation and drift (Eckert, 2002). It has been shown that mature Arabidopsis pollen grains express between 6,587 to 7235 genes making it a representative sample of the genome (Becker et al., 2003; da Costa-Nunes and Grossniklaus, 2003; Honys and Twell, 2003). Thirdly, the pollen genome is haploid, thus estimates of dominance coefficients are not required nor are deleterious mutations masked in a heterozygous state. Finally, previous work shows that partial or complete male sterility in plants may be attributed to a pre-pollination phenomenon (Fernando and Cass, 1997).
In the spring of 2003, we collected whole catkins from individual ramets (4-6) per clone in two natural populations: Riske Creek and Red Rock, Waterton Lake National Park. Details on the Red Rock population are not included here because this population was comprised of mainly small clones with under five ramets. Every attempt was made to ensure that the catkins had flowers that were fully open and functional anthers were in the two-lobed condition, indicative of the stage, just prior to the shedding of pollen (Fernando and Cass, 1997; Rajora and Zsuffa, 1986). To determine if, at the time of collection, the degree of flower/catkin development affected our estimates of sexual fitness, we noted the state of the catkin. Attempts were made to collect replicate catkins from different parts of the ramet crown. For each ramet, five catkins were collected. Given time constraints and the small size of individual flowers, we did not separate out anthers and suspend them in a mixture of lactophenol-aniline blue as is typical of pollen viability studies. Instead, whole catkins were put immediately into a tube containing lactophenol-aniline blue and a pestle was used to mechanically free the pollen grains from the anthers. Volume of the lactophenol aniline blue, before and after catkin addition was measured. All tubes were randomized. These were then brought back to the lab where pollen counts and estimates of the proportion of viable pollen were assayed by two “blind observers”. A pollen grain that was unstained, collapsed and abnormally shaped was considered non-viable. Pollen count was simply a direct count of all pollen grains using a Neubauer hemocytometer and a microscope (4X objective). With a more powerful microscope (40X), estimates of the proportion of viable pollen grains were made on a standard microscope slide, making three sweeps lengthwise along the slide and counting both viable and non-viable pollen.
Mean sexual fitness of a clone was a composite measure that included pollen viability and pollen count. A histogram plot shows the presence of substantial sexual fitness variation (SFigure 5-3). Although pollen collections from trees were completed within days of each other (Riske Creek: April 17-22 2003), viability and count can be sensitive to timing of catkin emergence and anther dehiscence (Fernando and Cass 1997, Rajora and Zsuffa 1985). We tested to see if sexual fitness variation was explained by the degree of catkin development using an ANOVA and found no effect (Riske Creek:F2,94 =2.5293, p=0.1118, SFigure 5-4). The sexual fitness data was square-root transformed to meet the assumption of normality and constant error variance.

For estimating clone age in total number of cell generations, we assumed most mutations occur during cell division in the roots and assumed cell divisions in the shoots were negligible. If shoot cell division was important then variation in ramet age would explain variation in ramet sexual fitness. To test this we determined ramet age by counting rings on increment cores taken from a sample of trees in the clone. All trees for which reproductive traits were measured had increment cores taken from them but we also sampled several other ramets within the same clone. A simple linear regression found that DBH was a good predictor of ramet age (F1,105=112.550, R²=0.740, p-value<0.0001). Thus to obtain ages for the remaining trees, which did not have an increment core or cookie, we used the following equation:

\[ \text{Ln}(\text{AGE})=1.452 + 0.817\times\text{Ln}(\text{DBH}) \]

Ramet age did not explain a significant amount of variation in ramet sexual fitness (SFigure 5-5).
Measures of asexual fitness

We measured asexual fitness at two levels: ramet and clone. Ramet asexual fitness is essentially incremental volume growth per year. Ramet ages were obtained as described above. Diameter at breast height (DBH) and height (m) were obtained for all sampled trees within the clone using a clinometer and a DBH measuring tape. Using these two measures we were able to calculate the volume of a ramet, \( V \), using diameter \( d \) and height, \( h \) in the following formula for a cylinder:

\[
V = \pi \frac{h}{4} d^2.
\]

We used two measures of asexual fitness for a clone: above-ground height growth and clone size. Although average above-ground clonal growth assumes the rate of above-ground shoot growth (vertical growth) is the same as root growth (horizontal growth), there is some evidence to show that higher leaf area and higher stand density correlates with parental root biomass per ramet (DesRochers and Leiffers, 2001). Our measure for above-ground growth rate assumes that within a clone, growth rate over time is constant and not affected by environmental conditions. Without excavating the entire clone, it is, at present, the only practical method available for estimating clonal growth in a natural population of a long-lived woody perennial.

The second measure of asexual fitness was the extrapolated number of ramets in a clone, clone size. Using ArcView GIS 3.3 (2002 Environmental Systems Research Insitute), GPS co-ordinates obtained on each clone were converted into area and perimeter covered by a clone. Our measure for density was obtained by counting all aspen ramets in a 10mx10m quadrat. Clone size was estimated by extrapolating from the 10mx10m area into the area covered by the clone.
Disease, environment, and principal component analysis:

Empirical studies suggest that plant sexual and asexual reproductive success are affected by the presence of fungal pathogens and insect herbivory (Goss and Bergelson, 2007; Jarosz and Davelos, 1995; Marr, 1998; Osier and Lindroth, 2004; Parker, 1994). Thus any measure of sexual fitness may be affected both by the abiotic and biotic environment of the ramet and/or the clone. For all sampled trees, we measured 11 morphological variables, which have previously been shown to reflect disease status for *P. tremuloides* trees (Brandt *et al.*, 2003; Peterson and Peterson, 1992). These variables were: DBH (cm), HEIGHT (m), number of conks, number of cavities, percent dead branches in crown, presence/absence of sap, number of scars, average length of scar (cm), proportion of leaves scored as eaten, proportion of leaves with a gall, and proportion of leaves exhibiting leaf minor.

A second aspect of the environment affecting plant reproductive success is site quality, as measured by available resources like soil moisture, nutrients, drainage level, light and soil temperature. Environmental variables like moisture vary through time and thus accurate and detailed assessments of site quality can be time consuming, expensive and difficult to obtain. As a proxy for site index, plant assemblages are often used because indicator plant species will often reflect differential resource availability and the presence of competitive interactions. This was true for aspen-dominated communities where understory vegetation was found to be significantly correlated to the site quality. In addition, this same study found that productive sites often had a higher plant understory species richness as measured by the Shannon Index (Chen, 1998; Chen, 2004). Thus, to obtain an assay of site quality we measured the percent cover of trees,
shrubs, herbaceous plants, moss, lichens, and bryophytes in each of the different patches using a sampling plot with a radius of 5m. Surveys were conducted during the months of May and July 2003. Where possible, all non-woody herbs and shrubs were identified to species level. In a few cases where habitats were similar and more than one species of a genus was found, we collapsed species into genus-level groups to reduce the number of variables in the PCA. In addition, from a combination of topographic and soil morphological properties, we obtained data on soil moisture regime, soil nutrient regime, and drainage class (Green and Klinka 1994). These site class characters were recoded into new dummy variables with presence or absence data for the patch.

Results from the principal component analysis:

A principal component analysis using a correlation matrix performed on 12 disease traits identified four axes which captured 73.1% of the variation in the original data. The principal component analysis was performed at the clone level on the averaged ramet data. From the loadings of the variables with each axis, we assigned factor names to these four axes: tree size and pathogen load (PATHLOAD), response to pathogen stress (RESPONSE), number of galls (NGALLS), and signs of tree mortality (MORTALITY). The second principal component analysis using a correlation matrix among 67 variables, 41 of which measured the percent cover of the plant understory in each patch, captured 71.9% of the variance using 8 axes. A sample of factors assigned to different PC axes include: PC1 (HIGH PRODUCTIVITY), PC2 (MOISTURE GRADIENT), PC3 (AMOUNT OF UNDERSTORY LITTER). As is recommended (Dillon and Goldstein, 1984), product moment correlations of $|0.50|$ were used to decide which variables had high loadings on the PC axes.

Consultation with various site identification guides and plant identification books
provided information on the environment and habitats of the different understory plants (Antos, 1996; Green and Klinka, 1994; Johnson, 1995; Shaw and On, 1979). This was then used to determine what factor names to assign to the principal component axes.

Data Analysis:

Mutational parameters were obtained by using both a likelihood approach (Keightley, 1994) and the Bateman-Mukai formula. The Bateman-Mukai formula assumes all mutations have an equal effect (Bateman, 1959; Mukai, 1964). We calculated the minimum genomic deleterious mutation rate, $U_{BM}$, for mutations affecting the number of viable pollen grains per diploid genome per year across a clone. Despite the fact that the sexual fitness data (average number of viable pollen grains per catkin per ramet) is haploid, we are measuring the mean fitness of the diploid ramet. In estimating sexual fitness for a ramet lineage, we averaged over the haploid products from multiple catkins to get an overall measure of ramet sexual fitness. Mean sexual fitness for a clone was the average of all ramets sampled within the clone. The following formula was used

$$U_{BM} = \frac{2(\Delta M)^2}{\sigma_M^2} \quad (1)$$

where $\Delta M$ represents the decline in mean sexual fitness per year. We obtained this estimate from a regression of mean relative sexual fitness on clone age. In a typical MA experiment, $\sigma_M^2$ is estimated from a regression of the among line variance against time.

Phenotypic divergence among lines is a direct result of the accumulation of mutations. In our study, however, each clone represents a single mutation accumulation experiment run for different periods of time, with ramets as the independent replicate lines. Thus, a regression of among ramet variance within a clone against clone age is the best estimate
of $\sigma^2_M$. Replicate catkins, sampled at different locations within the crown of a tree, were averaged and this represented the average sexual fitness for each ramet. Our estimate of $\sigma^2_M$ was obtained from a regression of the within clone variance on clone age and represents an estimate of the increase in genetic variance per year.

**Maximum Likelihood**

In addition to Bateman-Mukai mutational estimates, we employed a maximum likelihood estimation, first developed by (Keightley, 1994). The Bateman-Mukai formula assumes equal effect sizes and thus obtains a downward biased estimate of $U_{ML}$ and an upward biased estimate of the average effect size. Maximum likelihood (ML) estimators on the other hand, relax the assumption of equal mutational effects by modeling the distribution of effects using a gamma distribution Keightley (1994). By assuming equal mutational effects ($\alpha \rightarrow \infty$, where $\alpha$ is the shape parameter in a gamma distribution), we can compare the performance of our Bateman-Mukai estimate to what is obtained from maximum likelihood. With a few modifications to fit the current data set we use the ML approach of (Keightley, 1994) to estimate the genomic mutation rate, $U_{ML}$, parameters of the distribution of effect size $\alpha$, $s$ (scale parameter in the gamma distribution), environmental variance $V_E$, and the ancestral mean, M.

The C program (mlgenomeu.c) developed by Keightley (1994) can only be implemented on typical lab-based mutation accumulation experiments, where two time points (t=0 and t=end of MA) are compared for multiple lines originating from a single isogenic ancestor. Recent extensions made to this likelihood method allow for multiple time points for the same subline but assume a constant mutational effect (Keightley and
Bataillon, 2000). The assumption of a single isogenic ancestor, however, is not justified
with our data because groups of ramets (multiple lines) are seeded by different genotypes.
Secondly, each of these genotypes usually has a different clone age. Finally, a direct
comparison to an ancestor or control is not possible, so both the ancestral mean, \(M\), and
environmental variance, \(V_E\), were estimated by ML.

The observed data is \(Z_{r,c}\), the average sexual fitness (number of viable pollen
grains per catkin) for a ramet within a clone, sampled after the clone has undergone \(c\)
years. The number of accumulated mutations per year in ramets within a clone is
Poisson distributed with the expectation \(\lambda = Uc\). Here, the interval of time over which
mutations accumulate is the estimated clone age, \(c\) (in years), which is assumed to be
shared by all ramets within a clone.

The mutation rate, \(U_{ML}\), is the number of observed mutations per diploid genome
per year. Thus the probability of observing \(j\) mutations in a ramet is:

\[
Pr(X = j) = \frac{\exp^{-Uc} (Uc)^j}{j!}. \quad (2)
\]

As we were dealing with large numbers of cell divisions and thus old clones, we used a
normal distribution to approximate the Poisson for large values of \(\lambda\) \((\lambda = Uc \geq 100)\).
We employed the use of adaptive bins and a continuity correction to take into account the
fact that the Poisson distribution is discrete while the normal distribution is continuous.
For small values of \(\lambda\) \((\lambda = Uc \leq 100)\), we estimated the Poisson probability using
SEquation 2.
Following Keightley (1994), we retained the assumptions that mutations are deleterious, act additively, and have effect sizes, \( a_k \), drawn from a gamma distribution, of the following form:

\[
f(a_k) = \frac{1}{s^\alpha \Gamma(\alpha)} a_k^{(\alpha-1)} e^{-a_k/s}
\]

where \( \alpha \) is the shape parameter and \( s \) is the scale parameter (\( \alpha > 0 \) and \( s > 0 \)). This implies a mean of \( \bar{a} = \alpha \cdot s \) and a coefficient of variation of \( CV = \frac{1}{\sqrt{\alpha}} \). Since the sum of \( j \) gamma random variables is still gamma distributed, we used a gamma distribution to describe the sum of \( j \) effect sizes, \( A_j = \sum_{k=1}^{j} a_k \), where \( E(A_j) = j \cdot \alpha \cdot s \) and

\[
V(A_j) = j \cdot \alpha \cdot s^2.
\]

At large values of the shape parameter or very large numbers of mutations, the mean will increase but the coefficient of variation will tend to its limit of zero, hence approaching a constant effect size model. As we are actually measuring the effect size of mutants presumed to be in heterozygous form in the diploid ramets measured, relative fitnesses are \( 1 \) for the ancestral state, \( \frac{I - A_j}{2} \) for the heterozygote, and \( I - A_j \) for the homozygous mutant.

The observation \( Z_{r,c} \) was assumed to come from a normal distribution with an ancestral mean \( M \) with variance, \( V_E \). Previously, the average probability of observing a phenotype given any number of genotypic and environmental effects was determined using Monte Carlo integration. This was because the likelihood required a high dimensional integral (equation 2: Keightley (1994)) to account for all possible effect sizes of \( j \) mutations for the observed phenotype (drawn from a normal distribution). Monte
Carlo integration is highly demanding of computer time. By using the fact that the sum of gamma distributions is still gamma distributed, we were able to reduce this integral to a single dimension.

In order to account for a observed data point, $Z_{r,c}$, it must be the case that the combined effect of mutations and the ramet’s environmental deviation cause the phenotype of the ramet to depart from the mean by $Z_{r,c} - M$. Given that the ramet has accumulated $j$ mutations, its expected phenotype shifts from $M$ to $M - A_j$, where $A_j$ is the sum total effect of $j$ mutations drawn from a gamma distribution. The environmental deviation must then take the phenotype from $M - A_j$ to $Z_{r,c}$. The likelihood that this occurs is given by the integral:

$$
\int_{A_j=0}^{\infty} \left( \frac{A_j^{\alpha-1} e^{-A_j/j\alpha}}{\Gamma(\alpha)} \right) \left( \frac{e^{-\left(z_{r,c}-(M-A_j)\right)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \right) dA_j
$$

where the first parenthetical term is the probability of observing $A_j$ as the sum total draw from a gamma distribution with $j$ mutations and the second parenthetical term is the probability of observing a normal deviate (with mean 0 and variance $\sigma^2$) of $Z_{r,c} - (M - A_j)$. The overall probability of observing the data is then given by:

$$
L(Z_{r,c}) = p(d|Uc) \frac{e^{-\left(z_{r,c}-(M-A_j)\right)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} + \sum_{j=1}^{\infty} p(j|Uc) \int_{A_j=0}^{\infty} \left( \frac{A_j^{\alpha-1} e^{-A_j/j\alpha}}{\Gamma(\alpha)} \right) \left( \frac{e^{-\left(z_{r,c}-(M-A_j)\right)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \right) dA_j
$$
where $p(j | \lambda)$ is the Poisson probability of drawing $j$ mutations given an expected number of years, $c$. Fortunately, the integral in the above equation can be evaluated, which dramatically speeds up the calculation of the likelihood. The result obtained from evaluating the single integral of a normal distribution multiplied by a gamma distribution will approximate the average probability well for different cases of the parameters:

**Form A:**

$$2^{-\frac{\alpha j}{2}} e^{-\frac{Z_{rc}^2}{2}}\int_0^\infty \frac{\left(\frac{\alpha j}{2}, \frac{1}{s} - \frac{Z_{rc}}{V_E}\right)^2}{\sqrt{\frac{1}{2} (1 + \alpha j)}} \left(\frac{1 + \alpha j}{2}, \frac{3}{2} \frac{1}{s} - \frac{Z_{rc}}{V_E}\right)^2 \frac{1}{s V_E^2} \left(\frac{\alpha j}{2}\right) \]$$

where $M(a,b,z)$ is the confluent hypergeometric function of the first kind, an independent solution to Kummer’s function. When $z>0$ and $V_E > Z_{rc} \cdot s$ then using equation (13.1.3) in Abramowitz (1972) and replacing $(Z_{rc} \cdot s - V_E)$ with $- (s \cdot V_E) \left(\frac{Z_{rc}}{V_E} - V_E^2 / 2\right)$

where $z= \frac{V_E}{2} \left(\frac{1}{s} - \frac{Z_{rc}}{V_E}\right)$ in **Form A** gives a simpler and equivalent integral of **Form B:**

$$2^{-\frac{\alpha j}{2}} e^{-\frac{Z_{rc}^2}{2V_E}} s^{-\alpha j} V_E^{-\frac{1}{2} \cdot \frac{\alpha j}{2}} \left(\frac{\alpha j}{2}, \frac{1}{s} - \frac{Z_{rc}}{V_E}\right)^2 \left(\frac{3}{2}, \frac{1}{s} - \frac{Z_{rc}}{V_E}\right)^2 \frac{1}{s V_E^2} \left(\frac{\alpha j}{2}\right) \]$$

where $U(a, b, z)$ is the confluent hypergeometric function of the second kind.

The two forms were used in the likelihood program but under the following conditions. When the environmental variance is only a very small component of the phenotype ($V_E \leq Z_{rc} \cdot s$), then the original version of the integral (**Form A**) was used.
When \( V_e > Z_{r,e} \cdot s \), the simplified version of the integral provides both consistent and realistic probabilities.

The ML estimates of \( U \) and \( \bar{s} \) were obtained by examining profile likelihoods where a focal parameter was set to a range of values and the likelihood was optimized with respect to all other parameters. We used the simulated annealing optimization routine provided by R (version 2.4.1., R Development Core Group, 2007). Simulated annealing optimization routines are stochastic because they add random noise to the surface, thus avoiding the trap of a local optimum (Bolker, 2007). Thought to be more robust than the Nelder-Mead simplex, this Metropolis algorithm employs the following acceptance rule (Bolker, 2007):

\[
Pr(\text{accept}) = \begin{cases} 
\frac{e^{-\Delta L}}{k} & \text{if } \Delta L > 0 \\
1 & \text{if } \Delta L < 0
\end{cases}
\]

where \( \Delta L \) is the difference in the log likelihoods and \( k \) is a constant (temperature) that is periodically lowered (Bolker, 2007). The likelihood program was written in R (version 2.4.1., R Development Core Group, 2007) and is available upon request.

**Results from the Multiple Regression**

To identify a small subset of potential predictor variables we examined the relationship between mean sexual fitness and our abiotic and biotic variables using a correlation analysis and scatterplots. No correction was made for multiple comparisons because we were simply identifying potential predictors. From this we chose only those predictors, which showed a significant correlation (\( \alpha = 0.05 \)) in two types of analysis: an ordinary least squares regression analysis and a reduced major axis regression. The
ordinary least squares regression assumes that the predictors have all been measured without error. Fitting a regression model when a predictor has error results in a more biased estimate of the regression coefficient, especially if variability in the predictor is more than about a third of the error variability in response variable (McArdle, 2003). Thus we compared the partial regression coefficient obtained for clone age from both types of regressions. Model selection was based on AIC scores, p-values, \( C_p \) criterion, and adjusted \( R^2 \).

A stepwise multiple regression found two variables, CLONE AGE and MOISTURE GRADIENT explained 40.3% of the variation in sexual fitness variation (\( F_{2,17} = 7.412, \ p\text{-value}= 0.0048, \text{AIC} = 404.53 \)). We provide a sample of the rejected models in STable5-3 with the different criteria used for model selection. We selected this model because it had the lowest AIC criterion in comparison to all other models. Secondly, although the AIC criterion between our chosen model (CLONE AGE + MOISTURE) and the single predictor model with CLONE AGE was approximately two, the number of parameters in the CLONE AGE model was greater than \( C_p \) criterion but was not for our chosen model (CLONE AGE + MOISTURE) (STable5-3). Models with no bias will fall below the line \( C_p = p \), where \( p \) is the number of parameters estimated in the regression model. If the model falls above this line then the regression model is interpreted as having bias (Neter, 1996).

Significance is similar in the transformed model where two square-root transformations were performed on CLONE AGE and on SEXUAL FITNESS (STable5-4). On the transformed data, there were no large deviations from normality as observed in a normal quantile-quantile plot (\( W = 0.9806, \ p\text{-value} = 0.9422 \)). Plots of the residuals against each of the predictors revealed no departures from homoscedasticity in the errors.
From a plot of the raw data shown in SFigure 5-6a, we identified a single extreme observation a plot of the residuals against the fitted. The influence of the outlier on the fitted values was examined by examining Cook’s distance and its influence on the regression coefficients by examining the DFBETAS (Neter, 1996). The removal of this point decreased the significance of the regression model and only slightly flattened the slope, but it had no effect on the qualitative results. Because the outlier might well represent an unusually old clone (biologically relevant) and its inclusion in the transformed data satisfied the basic assumptions of linear regression, the outlier was retained for the estimation of the regression coefficient.

Clone age was not correlated to the moisture gradient ($\rho = 0.2419$, $t = -1.1847$, $df = 18$, p-value = 0.2515), suggesting that the regression coefficients corresponding to each predictor reflect an inherent effect of that predictor on the response variable (Neter, 1996). Overall, the regression coefficient for the decline in sexual fitness was $-1.14 \times 10^{-5} \pm 4.25 \times 10^{-6}$ per cell generation. Using the restriction that most clones had to be at least 77 years old (based on tree rings) and at most 8000 years old (based on glacial history), we obtain $\Delta \omega_{\text{Lower}} = -1.15 \times 10^{-4} \pm 4.28 \times 10^{-5}$ and $\Delta \omega_{\text{Upper}} = -1.01 \times 10^{-3} \pm 3.77 \times 10^{-4}$. The regression coefficient reflects an effect of clone age when moisture levels are held constant. A plot of the linear regression of sexual fitness against moisture level is found in SFigure5-7.

When the x-axis is measured with error then to obtain a less biased estimate of the slope, a reduced (standardized) major axis regression (RMA) is a more appropriate analysis and generates a less biased estimate of the slope (Warton, 2006). Model II regression or reduced major axis regression minimizes the sum of the areas of the
triangles formed by vertical and horizontal lines from each observation to the fitted line (Quinn, 2002). In order to obtain a value of the slope which takes into account the error present in clone age we used the SMATR library in R and performed a standardized major axis regression on the transformed data. For example, we found the 95% CI of the least squares estimate per cell generation (95% CI: -2.03x10^{-5} to -2.44x10^{-6}) overlapped with those of the standardized major axis regression (95% CI: -2.35x10^{-5} to -7.60x10^{-5}) but did not overlap with zero, confirming that sexual fitness declines as a function of clone age.

**Inbreeding depression**

Previous empirical data has suggested that early acting inbreeding depression may be high in long-lived organisms (Sorensen, 1999). Secondly, there has been previous evidence that high levels of heterozygosity, as measured by isozymes, was correlated with high growth variability in aspen (Mitton and Grant, 1980). If there is selection against homozygotes, then we might expect to see a negative relationship between time and levels of homozygosity. In addition, if the reduced sexual fitness is due to inbreeding depression then we would expect a positive relationship between levels of homozygosity (a measure of inbreeding) and mean clone sexual fitness. We measured inbreeding level for the clone as the proportion of homozygous loci. As shown in SFigure 5-8, we found no relationship between clone age and the proportion of homozygous loci ($\rho$=0.02665, p-value=0.4556), similarly, inbreeding level, as measured by homozygosity at microsatellites, did not explain any of the variation in sexual fitness ($F_{1,18}=1.125$, p-value= 0.3029, $R^2=0.05881$).
Trade-offs between asexual fitness and sexual fitness

We looked for evidence of a trade-off at two levels: ramet and clone. The ramet is an independent lineage where mutations accumulate, thus reduced sexual fitness may be associated with reduced overall growth. We performed a linear regression of ramet sexual fitness on volume growth per year and found no evidence of a trade-off ($\rho =0.1244$, $t=1.2097$, df=93, p-value>0.05). In addition, we found no relationship between ramet sexual fitness and ramet age (yrs) or incremental height growth of the ramet (cm/yr) (see SFigure 5-9).

We found some evidence for a trade-off at the level of the clone (SFigure 5-9). There was no relationship between above-ground clonal growth rate (average of ramet growth rates) and sexual fitness ($\rho =-0.1400170$, $t = -0.6$, df = 18, p-value = 0.556), but there was a decline in clonal sexual fitness with increasing clone size ($\rho =-0.490842$, $t=-2.3902$, df=18, p-value=0.0280). It was not significant, however, after a Bonferroni correction. Here, our measure of clone size (numbers of ramets) is a function of numbers of ramets belonging to one genotype, density of ramets within a patch, and physical area covered by clone. If real, this negative correlation could reflect mutations with negative pleiotropic effects, but it could also reflect environmental factors that promote clonal growth and retard sexual fitness.

Results from the Maximum Likelihood Analysis

We compared fixed values of the ancestral mean, $M$, while allowing genomic mutation rate $U_{ML}$, scale parameter, $s$, and environmental variance $V_e$ to vary, in order to
determine if mutational parameters were sensitive to different values of the ancestral mean. The maximum likelihood values of $U_{ML}$ were relatively insensitive to changes in the ancestral mean, however, the mean effect size, $\bar{a}$, and the environmental variance, $V_E$, were not (STable 5-5a). We performed a second sensitivity analysis on the environmental variance and found that for two fixed values of the ancestral mean, the genomic mutation rate and average effect size varied drastically (STable 5-5b).

Thirdly, we compared the fit of models with different fixed values of the shape parameter, keeping the ancestral mean constant ($M=35000$ and $M=17259$) but allowing $U_{ML}$, $s$, and $V_E$ to vary. Here we present maximum likelihood estimates based on assuming the clone exhibiting the greatest divergence was 8000 years old ($U_{MLower}$). Estimates of $U_{MLUpper}$ can be obtained by multiplying by a factor of 8.7. By holding the shape parameter to a large value large ($\alpha \geq 100$ so that $CV=\frac{1}{\sqrt{\alpha}}$ is near zero), our likelihood estimate of mutation rate $U_{MLower}=2.85x10^{-5} \ (M=17259)$ and $U_{MLower}=1.31x10^{-4} \ (M=35000)$ was very close to the minimum genomic mutation rate estimated under constant effect size of $UBMLower=1.06x10^{-5}$ per year.

Our maximum likelihood estimate of the average effect size for an ancestral mean of $M=35000$ was $\bar{a}=2.02x10^{-6} \ (\alpha = 1x10^{-4}, \ LogL=-243.37)$ whereas with an ancestral mean, $M=17259$, our maximum likelihood estimate of average effect size went down to $\bar{a}=1.29x10^{-8} \ (\alpha = 1x10^{-6}, \ LogL=-150.63)$. We calculated confidence regions around the average effect size as 2 log likelihood units away from the maximum value. Thus, for the ancestral mean of $M=17259$, the upper confidence region resulted in average effect size no larger than $\bar{a}=2.12x10^{-4} \ (LogL=-152.01: \ see \ STable \ 5-6b)$. While
for the ancestral mean of $M=35000$, the upper confidence region on the average effect size was $\bar{a} = 6.8 \times 10^{-4}$. We could not reject extremely small values of the shape parameter, yielding highly leptokurtic distributions where $\alpha \to 0$. This is because likelihoods changed very little for values of the shape parameter where $\alpha < 0.001$ (Table 5-6). We could, however, reject the constant effects model for both ancestral means because the LogL values were much greater than a factor of 7.4 from the maximum likelihood. For example, with an ancestral mean of $M=17259$, a shape value of $\alpha = 1$ and scale set to $s = 0.0112$ we obtained a value of LogL = -434.69, suggesting that this was a much poorer model. Thus our data implies that the distribution of mutational effects for the loss of average number of pollen grains per catkin per ramet is highly leptokurtic.

Estimated values of $U \to \infty$ as $\alpha \to 0$ and effect size become quite small, thus putting an upper bound on the mutation rate requires some assumptions about the value of $\alpha$. We found the maximum likelihood value for $U_{MLower} = 9.99$ (LogL = -150.63), with a lower confidence region on the estimate of $U_{MLower} = 2.43$ and the upper confidence region for $U_{MLower} = 9.95$. 
Figure 5-1. A genet or clone is initiated by a single genotype or seed. As the clone grows and reproduces asexually, each mitotic offspring (ramet) represents a replicate line. Produced from somatic tissue, these offspring are analogous to daughter cells produced by microbial organisms. With every round of asexual reproduction, somatic mutations can and do accrue resulting in among ramet/line divergence or mosaicism within a single individual. Although some mutations will be shared by neighboring ramets, our sampling scheme avoided close ramets. We represent the different somatic mutations carried by ramets in the shape and colour of circles. Without knowing the original genotype of the seed, it is impossible to determine the true ancestral lineage within a clone. For simplicity’s sake, we designate the most frequent genotype as the ancestral lineage (open circle). The genet is a collection of ramets of varying physiological ages. In *P. tremuloides*, trees are capable of producing new ramets at 1 year of age and ramets often remain connected to each other via lateral roots (solid lines) but sometimes, these root connections degrade (dashed lines), fragmenting the individual.
Figure 5-2. In Riske Creek, for a given the environment, relative clone sexual fitness (average numbers of pollen grains per catkin per ramet) declines significantly with increasing clone age (years). We present both the untransformed and the transformed data on the lower bound ($U_{BM,Lower}$) of the genomic mutation rate per year. To obtain relative sexual fitness we divided the average clone fitness by the estimate of the intercept (15440, S.E.M. ± 1412). Dashed lines represent the 95% confidence intervals.

a. Untransformed data. We found a slope of $-1.15 \times 10^{-4}$ (S.E.M. ± $4.23 \times 10^{-5}$) The 95% CI around the slope are $-2.05 \times 10^{-4}$ to $-2.46 \times 10^{-5}$. Because the outlier might well represent an unusually old clone (biologically relevant) and its inclusion in the transformed data satisfied the basic assumptions of linear regression, the outlier was retained (see Supplementary Materials).

b. The assumptions of regression were best met, however, when the data were square root transformed. On the transformed data, the slope was $-0.0062$ (S.E.M. ± 0.0018) and the 95% CI around the slope are $-0.010$ to $-0.0023$. Both the transformed data and results from the outlier analysis on the untransformed data suggest that there is a significant decline in sexual fitness with clone age.
a. Relative Sexual Fitness vs. Clone Age (years)

b. Square root transformed Relative Sexual Fitness vs. Square root transformed Clone Age (years)
**SUPPLEMENTARY FIGURES:**

**SFigure 5-3.** A histogram showing the variation in mean sexual fitness in the Riske Creek population. Mean sexual fitness in Riske Creek ($N_{ramet}=97$) was 13647.46 viable pollen grains per catkin (s.d. =7834.751).
**Figure 5-4.** An ANOVA showed that state (2 = partially emergent, 3 = mostly emergent, anthers have not yet dehisced, 4 = catkins completely open but anthers have not yet dehisced) of catkin emergence at collection time did not explain sexual fitness variation in Riske Creek. Means ± S.E.M. are shown. Untransformed data: $F_{2,94} = 2.5293$, p-value=0.08513; square root transformed data: $F_{2,94} = 2.243$, p-value=0.1118)
SFigure 5-5. A scatter plot shows that ramet age as estimated from tree ring data does not explain variation in mean sexual fitness (average number of pollen grains per catkin) in Riske Creek (N=97): $R^2=0.00014$, $F_{1,93}=0.01306$, p-value=0.9093.
SFigure 5-6. Results from an outlier analysis on the untransformed data with only clone age in the model. 

a. When a potential outlier (indicated by an arrow) was retained, we found the regression model significant ($F_{1,18}=8.206$, p-value=0.0103, $R^2=0.28$) with a slope of $-1.3 \times 10^{-4} \pm 4.6 \times 10^{-5}$. 

b. With the outlier removed, the regression model is still significant ($F_{1,17}=5.38$, p-value= 0.033, $R^2=0.20$) and the slope was $-2.4 \times 10^{-4} \pm 1.03 \times 10^{-5}$. In the square-root transformed model, the retention of the outlier does not affect normality of the residuals, create non-constant error variance, nor exert undue influence.
**SFigure 5-7.** A linear regression of untransformed sexual fitness against moisture gradient was significant ($F_{1,18} = 4.759$, p value $= 0.043$, $R^2 = 0.1514$).
**SFigure 5-8.** 

a. The proportion of homozygous loci in a clone does not change with time, as measured by divergence ($Rho=0.018$, $p>0.05$, $t=-0.0781$ df=18). 

b. The level of inbreeding, as measured by proportion of homozygous loci, does not explain variation in the average number of pollen grains per ramet ($F_{1,17}=1.025$, p-value = 0.32, $R^2=0.06$).
**SFigure 5-9.** Sexual and asexual fitness measures are only weakly related. a. At the ramet level, where mutations accumulate, there was no evidence for a tradeoff between volume growth (cubic meter per year) and ramet sexual fitness (R=0.09, t = 0.8701, df = 92, p-value=0.385). b. At the clone level, although there is a significant correlation, after a Bonferroni correction where α = 0.015 it is not (R=-0.49, t = -2.3902, df = 18, p-value = 0.03).
Table 5-1. Estimates of regression coefficients for the full model on the raw data are presented. The model includes both predictors: MOISTURE, and CLONE AGE. The data presented here assumes the maximum age of a clone is 8000 years old. The Type I error rate is $\alpha < 0.05$. We provide the estimates for the transformed data in STable 5-4.

<table>
<thead>
<tr>
<th>SEXFIT~</th>
<th>$F_{2,17}$ (p-value)</th>
<th>adj$R^2$</th>
<th>$\beta \pm \text{S.E.M}$</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone Age + Moisture</td>
<td>7.412 (0.0048)</td>
<td>0.403</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>1.009 (± 0.0088)</td>
<td></td>
<td>11.399</td>
<td>2.2x10^-9</td>
<td></td>
</tr>
<tr>
<td>moisture level (pc2)</td>
<td>0.173 (± 0.08)</td>
<td></td>
<td>1.967</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>clone age</td>
<td>-1.15x10^-4 (± 4.28x10^-5)</td>
<td></td>
<td>-2.684</td>
<td>0.0157</td>
<td></td>
</tr>
</tbody>
</table>
Table 5-2. The deleterious genomic mutation rate for sexual function obtained from *Populus tremuloides* is compared to estimates obtained from MA experiments on different organisms for traits related to sexual function.

<table>
<thead>
<tr>
<th>ORGANISM</th>
<th>GENOME SIZE (bp)</th>
<th>ESTIMATE OF U/(\bar{a})</th>
<th>U_{BM}</th>
<th>U_{ML}</th>
<th>UNITS OF THE MUTATION RATE</th>
<th>SEXUAL FITNESS TRAIT</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Populus tremuloides</em></td>
<td>485x10^6</td>
<td>-0.12-1.0x10^{-3}</td>
<td>9.4x10^{-5}</td>
<td>2.43-9.45</td>
<td>per diploid genome per ramet generation</td>
<td>Number of viable pollen grains per catkin per ramet (this study)</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>200x10^6</td>
<td>-0.010</td>
<td>0.12</td>
<td>0.01 to 0.20</td>
<td>per haploid per sexual generation</td>
<td>Egg to adult viability (Fry, 1999) Viability (Avila et al., 2006)</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>140x10^6</td>
<td>-0.004 (n.s.)</td>
<td>0.0003</td>
<td>0.1-0.2</td>
<td>per diploid per sexual generation</td>
<td>Seeds per fruit (Shaw, 2002)</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>140x10^6</td>
<td>-0.00062</td>
<td>0.0024</td>
<td>no ML available</td>
<td></td>
<td>Seed germination</td>
</tr>
<tr>
<td></td>
<td>0.00054</td>
<td>0.0030</td>
<td></td>
<td></td>
<td>Fruit set</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0044</td>
<td>0.054</td>
<td></td>
<td></td>
<td>Seeds/fruit (Schultz, 1999)</td>
<td></td>
</tr>
<tr>
<td><em>Amsinckia spectabilis</em></td>
<td>?</td>
<td>** 0.4 (0-2.89)</td>
<td></td>
<td></td>
<td>per diploid per sexual generation</td>
<td>Total fitness = germination to flowering (Schoen, 2005)</td>
</tr>
<tr>
<td><em>Amsinckia douglasiana</em></td>
<td>?</td>
<td>-0.0055</td>
<td>no BM available</td>
<td>0.098-0.654</td>
<td>0.098-0.656</td>
<td>per diploid per sexual generation</td>
</tr>
<tr>
<td></td>
<td>-0.00135</td>
<td></td>
<td></td>
<td>0.098-0.656</td>
<td>Plant dry weight (Schoen, 2005)</td>
<td></td>
</tr>
<tr>
<td><em>C.elegans</em></td>
<td>97x10^6</td>
<td>-0.03</td>
<td>0.00065</td>
<td>0.026</td>
<td>per diploid per sexual generation</td>
<td>Lifetime productivity (Keightley and Caballero, 1997b) Intrinsic rate of increase (Vassilieva and Lynch, 1999)</td>
</tr>
<tr>
<td></td>
<td>-0.00205</td>
<td>0.0135</td>
<td>0.0242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E.coli</em></td>
<td>4x10^6</td>
<td>?</td>
<td>0.00017</td>
<td>no ML available</td>
<td>per haploid per cell generation</td>
<td>Total fitness (Kibota and Lynch, 1996)</td>
</tr>
<tr>
<td><em>Cryptococcus neoformans</em></td>
<td>19.5x10^6</td>
<td>-0.6761</td>
<td>0.047</td>
<td>no ML available</td>
<td>per haploid per cell generation</td>
<td>Mating efficiency</td>
</tr>
<tr>
<td></td>
<td>-0.2488</td>
<td>0.0013</td>
<td></td>
<td></td>
<td>Filamentation (Xu, 2002)</td>
<td></td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>13x10^6</td>
<td>-3.04x10^{-3}</td>
<td>8.0x10^{-5}</td>
<td>0.00065</td>
<td>per diploid per cell generation</td>
<td>Sporulation rate (Hill and Otto, 2007)</td>
</tr>
</tbody>
</table>

U_{BM}: Bateman-Mukai, U_{ML}: Maximum likelihood **estimated by comparing selfed and outcrossed offspring
**SUPPLEMENTARY TABLES**

**Table 5-3.** The final model (boldface) is compared with the models rejected by the stepwise analysis. We provide F statistics, p-value, AIC criterion, Cp criterion, and adjusted R squared, which was used in model selection. Both sexual fitness and clone age were square-root transformed. We show models with the first three principal components from the environmental variables (PC_e). PC_e2 is MOISTURE LEVEL and with the first two disease principal components (PC_d).

<table>
<thead>
<tr>
<th>SEXFIT ~</th>
<th>F (p-value)</th>
<th>*p</th>
<th>df</th>
<th>AIC criterion</th>
<th>Cp criterion</th>
<th>adjRsq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NONE</strong></td>
<td>-</td>
<td>1</td>
<td>19</td>
<td>411.732</td>
<td>12.376</td>
<td>-</td>
</tr>
<tr>
<td><strong>CLONE AGE</strong></td>
<td>7.676 (0.013)</td>
<td>2</td>
<td>18</td>
<td>406.628</td>
<td>2.154</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>PC_e2 (MOISTURE LEVEL)</strong></td>
<td>4.759 (0.043)</td>
<td>2</td>
<td>18</td>
<td>409.041</td>
<td>8.419</td>
<td>0.17</td>
</tr>
<tr>
<td><strong>PC_e 2+ CLONE AGE (TRANSFORMED DATA)</strong></td>
<td>9.578 (0.002)</td>
<td>3</td>
<td>17</td>
<td>187.983</td>
<td>1.452</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>PC_e 2+ CLONE AGE (UNTRANSFORMED DATA)</strong></td>
<td>7.412 (0.0048)</td>
<td>3</td>
<td>17</td>
<td>404.526</td>
<td>1.452</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>PC_e1+PC_e 2+PC_e 3+CLONE AGE</strong></td>
<td>3.833 (0.0243)</td>
<td>5</td>
<td>15</td>
<td>405.650</td>
<td>3.490</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>PC_d1+PC_d 2+CLONE AGE</strong></td>
<td>3.21 (0.051)</td>
<td>4</td>
<td>16</td>
<td>408.310</td>
<td>5.634</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>PC_d1+PC_d 2</strong></td>
<td>0.1819 (0.835)</td>
<td>3</td>
<td>17</td>
<td>415.310</td>
<td>16.141</td>
<td>0.021</td>
</tr>
<tr>
<td><strong>PC_e1+PC_e 2+PC_e 3</strong></td>
<td>2.409 (0.105)</td>
<td>4</td>
<td>16</td>
<td>410.280</td>
<td>9.665</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>PC_d1+PC_d 2+PC_e1+PC_e 2</strong></td>
<td>0.1819 (0.835)</td>
<td>5</td>
<td>15</td>
<td>415.309</td>
<td>10.129</td>
<td>0.021</td>
</tr>
<tr>
<td><strong>PC_e1+PC_e 2+PC_e 3+PC_d1+PC_d 2+CLONE AGE</strong></td>
<td>2.784 (0.057)</td>
<td>7</td>
<td>13</td>
<td>407.207</td>
<td>7.0004</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*p=Number of estimated parameters in the regression function including the intercept
**Table 5-4.** Estimates of regression coefficients for the square-root transformed data. The model includes both predictors: MOISTURE, and CLONE AGE. The Type I error rate is $\alpha < 0.05$. 

**a.** If clones were a maximum of 8000 years old. 

**b.** If clones were a minimum of 77 years of age.

<table>
<thead>
<tr>
<th>$\sqrt{\text{SEXFIT}} \sim$</th>
<th>$F_{2,17}$ (p-value)</th>
<th>$\text{adj}R^2$</th>
<th>$\beta \pm \text{SEM}$</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLONE AGE + MOISTURE</td>
<td>9.578 (0.002)</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT</td>
<td>1.05</td>
<td></td>
<td>17.362</td>
<td>2.3x10^{-12}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(± 0.0623)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOISTURE LEVEL (PC2)</td>
<td>0.085 (± 0.045)</td>
<td>1.937</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQRT(CLONE AGE)</td>
<td>-0.0062 (± 0.0018)</td>
<td>-3.141</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A sensitivity analysis shows that $U$ is insensitive to a changing ancestral mean value $M$ (average number of viable pollen grains when no mutations have occurred). This was not, however, true of the mean effect size and the environmental variance. The ancestral mean, $M$, was fixed at 7 different values. The genomic mutation rate, $U$, scale parameter, $s$, and environmental variance $V_E$ were optimized for each value of $M$, while the shape parameter was fixed at 1 for all runs. We present likelihood estimates given the range of clone ages was from 140 to 8000 years old. A sensitivity analysis for $V_E$, allowing optimization of the shape parameter, $\alpha$, scale parameter, $s$, and mutation rate, $U$.

### Table 5-5. a. 

<table>
<thead>
<tr>
<th>$M$</th>
<th>$U_{est}$</th>
<th>$\bar{a}_{est}$</th>
<th>$V_{E_{est}}$</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>55304.0</td>
<td>2.76x10^{-3}</td>
<td>0.0065</td>
<td>0.0064</td>
<td>-393.53817</td>
</tr>
<tr>
<td>35000.0</td>
<td>2.76x10^{-3}</td>
<td>0.0073</td>
<td>0.0044</td>
<td>-389.10975</td>
</tr>
<tr>
<td>29500.0</td>
<td>2.76x10^{-3}</td>
<td>0.0072</td>
<td>0.0073</td>
<td>-391.10975</td>
</tr>
<tr>
<td>23553.0</td>
<td>3.82x10^{-3}</td>
<td>0.0085</td>
<td>0.0085</td>
<td>-401.40337</td>
</tr>
<tr>
<td>20510</td>
<td>3.90x10^{-3}</td>
<td>0.0094</td>
<td>0.0094</td>
<td>-413.87845</td>
</tr>
<tr>
<td>15001.0</td>
<td>4.17x10^{-3}</td>
<td>0.0119</td>
<td>0.012</td>
<td>-454.042249</td>
</tr>
<tr>
<td>10651.0</td>
<td>4.54x10^{-3}</td>
<td>0.0143</td>
<td>1.66</td>
<td>-536.16819</td>
</tr>
<tr>
<td>5500</td>
<td>2.98x10^{-3}</td>
<td>0.0263</td>
<td>4.52</td>
<td>-634.97917</td>
</tr>
</tbody>
</table>

### Table 5-5. b. 

<table>
<thead>
<tr>
<th>$V_E$</th>
<th>$M$</th>
<th>$U_{est}$</th>
<th>$\bar{a}_{est}$</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5x10^{-3}</td>
<td>17259.14</td>
<td>2.04x10^{-2}</td>
<td>5.470 x10^{-3}</td>
<td>-89924.84000</td>
</tr>
<tr>
<td>5x10^{-2}</td>
<td>17259.14</td>
<td>8.5</td>
<td>7.37x10^{-7}</td>
<td>-1203.62880</td>
</tr>
<tr>
<td>5x10^{-1}</td>
<td>17259.14</td>
<td>9.95</td>
<td>2.12 x10^{-4}</td>
<td>-152.01582</td>
</tr>
<tr>
<td>5</td>
<td>17259.14</td>
<td>9.93</td>
<td>1.69 x10^{-7}</td>
<td>-247.89563</td>
</tr>
<tr>
<td>50</td>
<td>17259.14</td>
<td>0.17</td>
<td>6.56 x10^{-6}</td>
<td>-547.07702</td>
</tr>
<tr>
<td>5x10^{-3}</td>
<td>35000.14</td>
<td>1.70 x10^{-2}</td>
<td>2.11 x10^{-3}</td>
<td>-176228.37000</td>
</tr>
<tr>
<td>5x10^{-2}</td>
<td>35000.14</td>
<td>8.36 x10^{-1}</td>
<td>2.50 x10^{-5}</td>
<td>-1944.77770</td>
</tr>
<tr>
<td>5x10^{-1}</td>
<td>35000.14</td>
<td>9.98</td>
<td>2.65 x10^{-9}</td>
<td>-243.00265</td>
</tr>
<tr>
<td>5</td>
<td>35000.14</td>
<td>9.05 x10^{-1}</td>
<td>3.35x10^{-5}</td>
<td>-387.49248</td>
</tr>
<tr>
<td>50</td>
<td>35000.14</td>
<td>9.75 x10^{-1}</td>
<td>3.58x10^{-6}</td>
<td>-563.04838</td>
</tr>
</tbody>
</table>
**Table 5-6.** To find the maximum likelihood value of the shape parameter, $\alpha$, we held $\alpha$ to a fixed set of values and allowed our algorithm to optimize $U_{ML}$, $s$ (scale) and $V_E$. The ML estimate of the mutation rate depends on the value of the shape parameter, $\alpha$ and the scale, $s$. We present likelihood estimates on for the case where the oldest clone was assumed to be 8000 years old.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>M</th>
<th>$U_{est}$</th>
<th>$\bar{a}_{est}$</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\infty$</td>
<td>17259</td>
<td>2.8x10^{-5}</td>
<td>&gt;1</td>
<td>-1317.0565</td>
</tr>
<tr>
<td>0.1</td>
<td>17259</td>
<td>0.0206</td>
<td>0.0014</td>
<td>-358.76239</td>
</tr>
<tr>
<td>0.01</td>
<td>17259</td>
<td>0.374</td>
<td>1.06x10^{-4}</td>
<td>-285.2139</td>
</tr>
<tr>
<td>0.0001</td>
<td>17259</td>
<td>2.69</td>
<td>8.87x10^{-6}</td>
<td>-159.21914</td>
</tr>
<tr>
<td>0.00001</td>
<td>17259</td>
<td>2.43</td>
<td>1.29x10^{-7}</td>
<td>-151.05144</td>
</tr>
<tr>
<td>0.000001</td>
<td>17259</td>
<td>9.99</td>
<td>1.29x10^{-8}</td>
<td>-150.63184</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>M</th>
<th>$U_{est}$</th>
<th>$\bar{a}_{est}$</th>
<th>LogL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\infty$</td>
<td>35000</td>
<td>1.31x10^{-4}</td>
<td>&gt;1</td>
<td>-1294.9726</td>
</tr>
<tr>
<td>0.1</td>
<td>35000</td>
<td>0.056</td>
<td>0.013</td>
<td>-317.54699</td>
</tr>
<tr>
<td>0.01</td>
<td>35000</td>
<td>0.239</td>
<td>5.93x10^{-5}</td>
<td>-257.63283</td>
</tr>
<tr>
<td>0.0001</td>
<td>35000</td>
<td>0.743</td>
<td>2.02x10^{-6}</td>
<td>-243.37716</td>
</tr>
<tr>
<td>0.00001</td>
<td>35000</td>
<td>0.019</td>
<td>7.06x10^{-5}</td>
<td>-338.75963</td>
</tr>
<tr>
<td>0.000001</td>
<td>35000</td>
<td>0.020</td>
<td>9.69x10^{-6}</td>
<td>-336.10457</td>
</tr>
<tr>
<td>0.0000001</td>
<td>35000</td>
<td>0.012</td>
<td>8.00x10^{-8}</td>
<td>-355.02513</td>
</tr>
</tbody>
</table>
REFERENCES


Johnston, M. O. & Schoen, D. J. Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. Science 267, 226-228 (1995).


Parker, M. A. Pathogens and sex in plants. Evolutionary Ecology 8, 560-584 (1994).


Chen, H. Y. H., Legare, S. and Bergeron, Y. Variation of the understory composition and diversity along a gradient of productivity in Populus tremuloides stands of


Green, R. N. & Klinka, K. A field guide to site identification and interpretation for the Vancouver Forest Region. (Research Branch of the Ministry of Forests, Victoria, B.C., 1994).


Chapter 6: CONCLUDING REMARKS

“A hen is only an egg’s way of making a new egg.” Samuel Butler (19th Century)

The idea that changes in the germline are the sole means of evolutionary change was first stated by August Weismann in his Doctrine of the Continuity of the Germplasm (discussed in Buss, 1983). With the individual as the principal unit of selection, the soma or sub-organismal units was made insignificant and ultimately irrelevant to the process of evolution. Plants, and in particular, clonal plants illustrate that this doctrine, however, is not universal. Mutations arising in a meristem can be passed during modular growth to the pollen or ova of the ramet. Moreover, the presence of genetic variance within an individual, the abundance and distribution of different genotypes is determined by differential growth (Sutherland and Watkinson, 1986).

Somatic mutation is evolutionarily relevant to plants because the presence of multiple genetic lineages within an organism provides an opportunity for intra-organismal selection to act (Antolin and Strobeck, 1985; Otto and Orive, 1995; Whitham and Slobodchikoff, 1981). This can have the effect of increasing the fitness of an individual by reducing the mutational load through the elimination of deleterious mutations or by fixing beneficial ones (Gill et al., 1995b; Pineda and Fagerstrom, 1999; Whitham and Slobodchikoff, 1981). This thesis demonstrates that an individual plant clone is, in fact, a mosaic both genetically (at neutral loci) and phenotypically (sexual fitness). Furthermore, by taking advantage of genetic mosaicism within an individual, we could estimate the genomic deleterious mutation rate for sexual function. In doing so, we
demonstrated that mosaicism comes at a cost to sexual propagation of the same individual.

In Chapter 2 the groundwork for the study was laid, demonstrating that local forest patches in western populations of *P. tremuloides* often consisted of distinct genotypes, indicating the presence of sexual recruitment. One population had substantial numbers of clonal offspring, as patches of trees in this population were often dominated by a single genotype. In the grasslands where aspen dominates, there is a non-random distribution of genets, i.e., successful sexual establishment largely occurred on the edges of a patch. Although none of the microsite environmental variables measured explained the variation in genotypic diversity, we used composite variables of different plant understory species inferentially. A better measure of the environment is direct measurement of biological traits of the plant species in the aspen understory. Certain plant traits are adaptive and may represent a functional response to environmental features like disturbance. An examination of the relationship between genetic diversity and plant species grouped by common adaptive traits might provide more insight into what governs asexual versus sexual recruitment.

The microsatellite data in Chapter 3 illustrated an individual plant clone is a genetic mosaic. Despite variation both in clone size and genetic distance within a clone, no simple relationship emerged between clone size and time. Our data, however, suggested that neutral genetic divergence within clones is adequately described by a sudden expansion model of growth. This model allowed us to estimate clone age from the level of neutral mosaicism in a clone.
One question resulting from our attempts to estimate the rate at which mosaicism arises within an individual is what is the appropriate measure of time for a clone, i.e., sexual generation, year, ramet generation or cell generation. Typically, when estimating a genomic mutation rate or the rate at which mutation arises within a population, the unit of time used is a sexual generation. A sexual generation is defined as the mean age to reproductive maturity. This unit of time is most relevant because natural selection acts on new mutations in terms of their \textit{per generation} impact on relative fitness. While straightforward for purely sexual organisms, the unit of time is more complicated in clonal organisms because of the multiple levels at which selection can act. If fitness is defined as the relative contribution of mitotic offspring to the next generation then a per cell division or per ramet generation may be more appropriate. Moreover, comparisons of mutation rates between a clonal plant and a sexual one require knowledge of the number of cell generations found in a single sexual generation. There is, however, a lack of biologically realistic models to describe cell growth and production in plant modules.

Chapter 4 demonstrated that a star-shaped genealogy is an appropriate assumption in the estimation of clone age on age-structured clones subject to density-dependence and exponential growth. One avenue for exploration is to examine how repeated waves of extinction that reduce the overall density of juveniles may affect clone age estimates. It is clear that genetic variation within a patch is non-random over space, as shown in Chapter 2, therefore in future it would be useful to explore how the distribution of multiple genotypes in different spatial arrangements interact with sampling strategy to impact clone age estimates.
Finally, Chapter 5 provides the first estimate of a slightly deleterious mutation rate and the distribution of effect sizes for a woody tree species in the wild. Mutations reduced relative sexual fitness in clonal aspen by about $0.12 \times 10^{-3}$ to $1.01 \times 10^{-3}$ per year. The deleterious mutation rate for sexual fitness in a natural population of aspen was estimated to be $U_{ML}=2-10$ per diploid genome per year. We found mutations, while mostly of small effect, resulted in a 11% reduction in a clone’s sexual fitness per mutation and the oldest clone sustained a 50% reduction in total in sexual fitness.

The question of whether the lack of a clear distinction between the “soma” and the “germline” could confer immortality to a clone has been extensively debated in the literature (Gill et al., 1995a; Hamilton, 1966; Orive, 1995; Partridge, 2001). While, Hamilton argued senescence was an inevitable outcome of evolution (1966), many suggest that plants with clonal growth show no senescence (Harper, 1977; Vaupel et al., 2004). One previous model predicted in clonal organisms there would be a greater effect of senescence in terms of a decline in the rate of sexual reproduction rather than an increased probability of mortality (Gardner and Mangel, 1997). This thesis empirically confirms that senescence is a more complex phenomenon and occurs at different rates for sexual and asexual components of fitness.
REFERENCES


Whittle C-A, Johnston MO (2006) Moving forward in determining the causes of mutations: the features of plants that make them suitable for assessing the impact


APPENDIX A-2. Allele frequency distributions for each of the five microsatellite loci across three different populations of *P. tremuloides*.
PMGC 575

Frequency

Riske Creek
Red Rock
Bellevue Hill

Fragment size (bp)
**Appendix B-2.** Genotypic diversity measures for three populations of *P. tremuloides*.

**Table 1-B-2.** Diversity measures for 17 patches in Riske Creek.

<table>
<thead>
<tr>
<th>Patch#</th>
<th>n&lt;sub&gt;p&lt;/sub&gt;</th>
<th>MLG&lt;sub&gt;p&lt;/sub&gt;</th>
<th>Proportion of total MLG (edge)*</th>
<th>Proportion of total MLG (inside)*</th>
<th>c (NegLogL)</th>
<th>Genotypic richness (R&lt;sub&gt;p&lt;/sub&gt;)</th>
<th>Genotypic diversity (D&lt;sub&gt;obs&lt;/sub&gt;)</th>
<th>Evenness (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41</td>
<td>4</td>
<td>0.75</td>
<td>0.50</td>
<td>3.08x10^-7</td>
<td>(-4.384)</td>
<td>0.10</td>
<td>0.46</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>2</td>
<td>0.50</td>
<td>1.0</td>
<td>-</td>
<td>0.03</td>
<td>0.42</td>
<td>0.81</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>6</td>
<td>0.83</td>
<td>0.33</td>
<td>0.0141</td>
<td>(-14.859)</td>
<td>0.15</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>13</td>
<td>0.85</td>
<td>0.23</td>
<td>14.8</td>
<td>(72.379)</td>
<td>0.30</td>
<td>0.57</td>
</tr>
<tr>
<td>6</td>
<td>48</td>
<td>15</td>
<td>0.93</td>
<td>0.13</td>
<td>0.00630</td>
<td>(-25.954)</td>
<td>0.30</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>48</td>
<td>14</td>
<td>0.36</td>
<td>0.64</td>
<td>0.00568</td>
<td>(-14.767)</td>
<td>0.28</td>
<td>0.80</td>
</tr>
<tr>
<td>13</td>
<td>48</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>15</td>
<td>53</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>17</td>
<td>42</td>
<td>6</td>
<td>0.67</td>
<td>0.33</td>
<td>-</td>
<td>0.12</td>
<td>0.46</td>
<td>0.38</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>20</td>
<td>45</td>
<td>6</td>
<td>0.83</td>
<td>0.33</td>
<td>2.01x10^-6</td>
<td>(-11.561)</td>
<td>0.11</td>
<td>0.71</td>
</tr>
<tr>
<td>21</td>
<td>14</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>22</td>
<td>26</td>
<td>2</td>
<td>0.50</td>
<td>0.50</td>
<td>-</td>
<td>0.040</td>
<td>0.52</td>
<td>0.99</td>
</tr>
<tr>
<td>23</td>
<td>55</td>
<td>15</td>
<td>0.73</td>
<td>0.27</td>
<td>1.00x10^-9</td>
<td>(-8.2725)</td>
<td>0.26</td>
<td>0.66</td>
</tr>
<tr>
<td>24</td>
<td>19</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>25</td>
<td>48</td>
<td>6</td>
<td>0.67</td>
<td>0.33</td>
<td>0.510</td>
<td>(100.467)</td>
<td>0.11</td>
<td>0.60</td>
</tr>
</tbody>
</table>

**Mean** - 7.4 0.68 0.42 - 0.13 0.43 0.41

**S.D** - 8.2 0.18 0.23 - 0.15 0.31 0.34

**S.E.** - 1.99 0.051 0.066 - 0.036 0.076 0.083

*Numbers do not sum to one because in some cases a MLG was found both on the edge and inside.
Appendix B-2 Genotypic diversity measures for three populations of *P. tremuloides*.

Table 2-B-2. Diversity measures for ten patches in Red Rock, Waterton National Parks.

<table>
<thead>
<tr>
<th>Patch#</th>
<th>n_P</th>
<th>MLG_P</th>
<th>Genotypic richness ($R_P$)</th>
<th>Genotypic diversity ($D_{obs}$)</th>
<th>Evenness ($E$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>4</td>
<td>0.18</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4</td>
<td>0.16</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>2</td>
<td>0.030</td>
<td>0.059</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>4</td>
<td>0.17</td>
<td>0.38</td>
<td>0.17</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>2</td>
<td>0.083</td>
<td>0.51</td>
<td>0.93</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>3</td>
<td>0.29</td>
<td>0.65</td>
<td>0.62</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>5</td>
<td>1.0</td>
<td>1.0</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>4</td>
<td>0.27</td>
<td>0.77</td>
<td>0.88</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>-</td>
<td>3.20</td>
<td>0.32</td>
<td>0.57</td>
<td>0.40</td>
</tr>
<tr>
<td>S.D</td>
<td>-</td>
<td>1.23</td>
<td>0.37</td>
<td>0.34</td>
<td>0.41</td>
</tr>
<tr>
<td>S.E.</td>
<td>-</td>
<td>0.389</td>
<td>0.12</td>
<td>0.11</td>
<td>0.13</td>
</tr>
</tbody>
</table>
### APPENDIX B-2 Genotypic diversity measures for three populations of *P. tremuloides*.

**Table 3-B-2.** Diversity measures for 10 patches in Bellevue Hill, Lake Waterton National Parks.

<table>
<thead>
<tr>
<th>Patch#</th>
<th>(n_p)</th>
<th>MLG(_p)</th>
<th>Proportion of total MLG (edge)*</th>
<th>Proportion of total MLG (inside)</th>
<th>(c) ((\text{NegLogL}))</th>
<th>Genotypic richness ((R_p))</th>
<th>Genotypic diversity ((D_{obs}))</th>
<th>Evenness ((E))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>23</td>
<td>0.96</td>
<td>0.090</td>
<td>(1.02\times10^{-3}) ((-12.794))</td>
<td>0.31</td>
<td>0.80</td>
<td>0.63</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>2</td>
<td>1.0</td>
<td>0.50</td>
<td>(2.80\times10^{-5}) ((16.151))</td>
<td>0.019</td>
<td>0.037</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>7</td>
<td>0.57</td>
<td>0.57</td>
<td>(6.35\times10^{-7}) ((19.485))</td>
<td>0.11</td>
<td>0.25</td>
<td>0.019</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>9</td>
<td>0.78</td>
<td>0.33</td>
<td>(0.024) ((-34.744))</td>
<td>0.15</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>8</td>
<td>0.67</td>
<td>0.33</td>
<td>(0.98) ((24.087))</td>
<td>0.12</td>
<td>0.42</td>
<td>0.30</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>33</td>
<td>0.67</td>
<td>0.33</td>
<td>(0.034) ((-59.782))</td>
<td>0.49</td>
<td>0.96</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>58</td>
<td>10</td>
<td>0.70</td>
<td>0.30</td>
<td>(0.00054) ((3.0544))</td>
<td>0.17</td>
<td>0.69</td>
<td>0.64</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>11</td>
<td>0.73</td>
<td>0.36</td>
<td>(6.16\times10^{-6}) ((5.3244))</td>
<td>0.17</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>37</td>
<td>0.84</td>
<td>0.19</td>
<td>(0.15) ((-15.833))</td>
<td>0.46</td>
<td>0.88</td>
<td>0.63</td>
</tr>
<tr>
<td><strong>MEAN</strong></td>
<td>-</td>
<td><strong>14.1</strong></td>
<td><strong>0.77</strong></td>
<td><strong>0.33</strong></td>
<td>-</td>
<td><strong>0.20</strong></td>
<td><strong>0.52</strong></td>
<td><strong>0.41</strong></td>
</tr>
<tr>
<td><strong>S.D.</strong></td>
<td>-</td>
<td><strong>12.6</strong></td>
<td><strong>0.14</strong></td>
<td><strong>0.14</strong></td>
<td>-</td>
<td><strong>0.18</strong></td>
<td><strong>0.35</strong></td>
<td><strong>0.34</strong></td>
</tr>
<tr>
<td><strong>S.E.</strong></td>
<td>-</td>
<td><strong>3.97</strong></td>
<td><strong>0.044</strong></td>
<td><strong>0.045</strong></td>
<td>-</td>
<td><strong>0.056</strong></td>
<td><strong>0.11</strong></td>
<td><strong>0.11</strong></td>
</tr>
</tbody>
</table>

*Numbers do not sum to one because in some cases a MLG was found both on the edge and inside*