Phenotypic and Genetic Differentiation Between Sex Chromosome Races of *Rumex hastatulus* (Polygonaceae)

Andrew Joseph Simpson

A thesis submitted in conformity with the requirements for the degree of Master of Science, Graduate Department of Ecology and Evolutionary Biology, University of Toronto

© Copyright by Andrew Joseph Simpson 2014
Phenotypic and Genetic Differentiation Between Sex Chromosome Races of
*Rumex hastatulus* (Polygonaceae)

Andrew Joseph Simpson
Master of Science
Department of Ecology and Evolutionary Biology
University of Toronto
2014

**ABSTRACT**

Wind-pollinated plants often have little genetic structure owing to high gene flow. However, sex chromosomes evolution promotes divergence, potentially leading to population subdivision. *Rumex hastatulus* (Polygonaceae) is a wind-pollinated, dioecious coloniser of open disturbed land with two sex chromosome races (SCRs), which occupy different parts of the species’ range in the southern USA. My thesis investigated phenotypic and genetic differentiation between the SCRs, based on a study of 28 populations sampled throughout the range. A glasshouse experiment demonstrated significant differentiation between the SCRs in several life-history traits, several of which exhibited clinal variation. Analysis of population genetic structure, based on 13 nuclear loci, revealed two distinct clusters corresponding to the two SCRs, with an estimated origin of the North Carolina SCR from the Texas SCR between ~5000–15000 generations ago. My results indicate that the SCRs have developed a moderate degree of phenotypic and genetic differentiation despite ongoing gene flow.
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Spencer Barrett for his many contributions to this thesis and his patience with my pace of writing; his boundless enthusiasm for the all things related to plants was infectious, inspiring, and incisive. I would also like to thank Stephen Wright for providing extensive help troubleshooting population genetics software and interpreting various programs’ outputs his advice was invaluable, and also John Stinchcombe for his help with statistical approaches and excellent advice on rooting out help from other members of the department. My lab mate Ramesh Arunkumar provided glasshouse help and advice on the molecular chapter, and was an excellent person with whom to share an office. Bill Cole provided assistance with glasshouse work and statistical analyses and Josh Hough also helped with glasshouse work and was kind enough to provide the transcriptome data on which my PCR primers were based. I would also like to sincerely thank my family and my partner, Megan Underwood, without whose love and support I never would have finished this thesis. I would also like to thank Alex De Seranno, Adrianna Salcedo, Adam Dobkin, and Alina Usataia for help with sample collection. Stephen De Lisle and Nathaniel Sharp provided advice regarding data analysis. The research in this thesis was funded by an NSERC Discovery grant to Spencer Barrett.
TABLE OF CONTENTS

CONTENTS PAGE
ABSTRACT ii
ACKNOWLEDGEMENTS iii
LIST OF FIGURES v
LIST OF TABLES vi
LIST OF APPENDICES vii
GLOSSARY OF ACRONYMS AND SYMBOLS viii

CHAPTER 1 – GENERAL INTRODUCTION
Effects of life history and mating system on population structure 1
Sex chromosomes as drivers of evolution 2
Common garden experiments 4
Study system – *Rumex hastatulus* 4
Thesis objectives 8

CHAPTER 2 – MORPHOLOGICAL AND PHENOLOGICAL VARIATION
Abstract 9
Introduction 10
Materials and methods
  *Study system* 12
  *Glasshouse experiment* 12
  *Analyses of quantitative trait differentiation* 14
  *Clinal variation in ecologically relevant traits* 15
Results
  *Trait differences between the chromosomal races and genders* 16
  *Multivariate analyses of differentiation between sex* 21
chromosome races and between genders

Clinal variation of ecologically relevant traits 23

Discussion 26

Can the sex chromosome races be distinguished by phenotypic traits? 27

Patterns of sexual dimorphism in the sex chromosome races 28

Correlations of phenotypic traits with environmental variables 29

CHAPTER 3 – MOLECULAR ANALYSIS OF POPULATION STRUCTURE

Abstract 32

Introduction 33

Materials and methods

 Population sampling 36
 Primer design, DNA extractions and Sanger sequencing 37
 Analyses of polymorphism 38
 STRUCTURE analysis 38
 Coalescent simulations 39

Results

 Polymorphism and genetic diversity 40
 Population genetic structure 42
 Demographic history 44

Discussion 45

CHAPTER 4 – GENERAL CONCLUSIONS

Major Findings of the thesis 50

Future Work 51
LITERATURE CITED

SUPPLEMENTARY INFORMATION I
Summary of all populations studied 64

SUPPLEMENTARY INFORMATION II
Trait means and standard errors for each population 66

SUPPLEMENTARY INFORMATION III
Eigen vectors used in the PCA described in Chapter 2 68

SUPPLEMENTARY INFORMATION IV
Population genetic parameters for synonymous and nonsynonymous sites for each population 69

SUPPLEMENTARY INFORMATION V
Code used in R for statistical analyses 70
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NUMBER AND CONTENT</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 – Distribution of <em>Rumex hastatulus</em></td>
<td>5</td>
</tr>
<tr>
<td>1.2 – Male and female plants at different stages of flowering</td>
<td>6</td>
</tr>
<tr>
<td>2.1 – Map of populations used for Chapter 2</td>
<td>13</td>
</tr>
<tr>
<td>2.2 – Proportion of individuals who flowered by population</td>
<td>18</td>
</tr>
<tr>
<td>2.3 – Grand means of phenotypic characters compared between SCRs</td>
<td>19</td>
</tr>
<tr>
<td>2.4 – Mean proportion of total variance explained by different hierarchical levels for quantitative phenotypic traits</td>
<td>20</td>
</tr>
<tr>
<td>2.5 – Three-dimensional scatterplot of a PCA analysis of phenotypic traits</td>
<td>23</td>
</tr>
<tr>
<td>2.6 – Scatterplots of phenotypic traits regressed against latitude</td>
<td>25</td>
</tr>
<tr>
<td>3.1 – Map of populations used for Chapter 3</td>
<td>36</td>
</tr>
<tr>
<td>3.2 – Average values for population genetic parameters calculated for synonymous and non-synonymous sites</td>
<td>41</td>
</tr>
<tr>
<td>3.3 – Values of ΔK based on the Evanno method, as implemented in STRUCTURE Harvester</td>
<td>42</td>
</tr>
<tr>
<td>3.4 – Cluster assignment probabilities for STRUCTURE runs with values of K=2 to K=5</td>
<td>43</td>
</tr>
<tr>
<td>3.5 – Posterior probability estimates of the ancestral θ and coalescence time obtained from MIMAR</td>
<td>44</td>
</tr>
<tr>
<td>TABLE NUMBER AND DESCRIPTION</td>
<td>PAGE</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.1 – Results of two-way mixed-effects ANOVAs on the influence of sex chromosome race, gender, and their interaction on phenotypic traits</td>
<td>17</td>
</tr>
<tr>
<td>2.2 – Results of two-way mixed-effects ANOVAs on the influence of sex chromosome race, gender, and their interaction on the first three principal components comprising seven phenotypic traits</td>
<td>21</td>
</tr>
<tr>
<td>2.3 – Permutational multivariate analysis of variance comparing phenotypic traits between SCRs</td>
<td>22</td>
</tr>
<tr>
<td>2.4 – Multivariate analysis of covariance results investigating the effects of latitude and longitude on phenotypic traits</td>
<td>24</td>
</tr>
<tr>
<td>3.1 – Unique and shared polymorphisms between the Texas and North Carolina SCRs</td>
<td>39</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of <em>Rumex hastatulus</em> populations and their locations</td>
<td>64</td>
</tr>
<tr>
<td>Means and standard error for phenotypic traits measured in each population</td>
<td>66</td>
</tr>
<tr>
<td>Eigenvectors used in the PCA described in Chapter 2</td>
<td>68</td>
</tr>
<tr>
<td>Population genetic parameters for synonymous and nonsynonymous sites for each population</td>
<td>69</td>
</tr>
<tr>
<td>Code used in R for statistical analyses</td>
<td>70</td>
</tr>
</tbody>
</table>
LIST OF ACRONYMS AND SYMBOLS

ANOVA – Analysis of variance
LME – Linear mixed effects model
$F_{st}$ – Wright’s fixation index
MCMC – Markov chain Monte-Carlo
PCA – Principal component analysis
REML – Restricted maximum likelihood
SNP – Single nucleotide polymorphism
$\theta_w$ – Watterson’s theta. Equal to $4N_e\mu$
CHAPTER 1 – GENERAL INTRODUCTION

Evolutionary biologists have long had a fascination with analysing population genetic structure at various geographic scales and taxonomic levels. The reason for this interest is because the evaluation of population structure can provide key empirical tests of evolutionary theory. Wright’s $F_{st}$ (Wright 1943, 1951, 1965) greatly facilitated the analysis of population structure by providing a means of quantifying the extent of genetic differentiation within and among populations. As computing power and molecular techniques have advanced, new methods for examining population structure based on Bayesian algorithms and a range of DNA-based markers have developed (e.g. Pritchard et al. 2001). Irrespective of the methods used, information on population genetic structure is of interest because it can provide insights into the diverse biological and historical factors causing lineage differentiation in plant and animal groups.

Effects of life history and reproductive system on population structure

The 1970s ushered in numerous studies on the patterns of enzyme polymorphism in plant populations, and these often revealed high levels of genetic diversity (Brown 1979). As more species were examined, ecological geneticists began to examine the relations between patterns of genetic variation and ecological and life-history traits. Large-scale comparative surveys allowed researchers to make predictions about population genetic structure based on information regarding the study species’ geographic range, population size and diverse features of life history and reproductive system (e.g. Hamrick et al. 1979). In particular, the pollination and mating system were shown to be influential in affecting population genetic structure and levels of diversity. For example, long-lived, wind-pollinated, outcrossing species, such as many trees, were generally found to maintain high levels of allozyme diversity within populations, but relatively little genetic
differentiation among populations; in contrast, self-pollinating, annual, colonizing species with high rates of self-fertilization most commonly possessed low levels of diversity within populations, but a high degree of differentiation among populations (Hamrick and Godt 1996). As different life-history traits can have contrasting effects on population structure and levels of genetic diversity, the observed patterns will depend on the relative influence of each trait and their interactive effects with historical and demographic factors.

**Sex chromosomes as drivers of evolution**

It is well established that changes in chromosome number or structure play important roles in the evolution of reproductive isolation in flowering plants (Stebbins 1971, Rieseberg 2001, Levin 2002). While there have been extensive investigations of the key role that sex chromosomes play in contributing to lineage differentiation and speciation in animals (Coyne and Orr 1989, Presgraves 2002, Turelli and Moyle 2007), there have been few such studies in dioecious plants that possess sex chromosomes, presumably because plants with sex chromosomes are quite rare among flowering plants, owing to the low frequency of separate sexes (dioecy, see Charlesworth 2002). Despite their infrequent occurrence, plants with sex chromosomes can provide useful insights into the initial stages of their evolution because plant sex chromosomes are generally more recently evolved from autosomes than those of mammals or other animal groups, which have been the subject of extensive genetic studies e.g. *Drosophila* (reviewed in Westergaard 1958, Charlesworth and Charlesworth 1978). Sex-linked genes in plants show evidence of faster rates of evolution relative to autosomes (Filatov and Charlesworth 2002, Whittle and Johnston 2002), suggesting that they may play an important role in the evolutionary divergence of species. Chromosomal regions with faster rates of evolution usually differentiate more rapidly than the genome-wide level of differentiation for divergent lineages.
Speciation proceeds by individuals from divergent lineages evolving barriers to reproduction that reduce genetic exchange between populations. The process generally occurs gradually, owing to a buildup of either prezygotic or postzygotic isolating mechanisms (Barton and Charlesworth 1984; Coyne and Orr 2004). Prezygotic barriers to reproduction include differences in the timing of reproduction, ecology, behaviour, gametic compatibility, and pollinator preferences; whereas, postzygotic barriers include hybrid inviability and infertility, and a breakdown of hybrid fitness in subsequent generations.

Postzygotic reproductive isolation can arise following chromosomal fusions and translocations that cause differences in chromosomal structure among populations (Painter and Stone 1935; Rieseberg 2001). If populations fix different chromosomal arrangements, meiosis cannot proceed normally because of incomplete pairing during metaphase (Baker and Bickham 1986). More recent models of chromosomal speciation propose that chromosomal rearrangements are important for speciation because they reduce recombination in heterokaryotypic genotypes, leading to the accumulation of lineage-specific substitutions (Noor et al. 2001, Rieseberg 2001). Most studies of reproductive isolation and speciation have focused on related species pairs, some of which have been diverged for extended periods of time, or focus on chromosomal inversions rather than translocations (reviewed in Coyne and Orr 2004). These studies are often unable to provide insights into the very early stages of lineage differentiation resulting from changes to chromosomal structure. Investigating species in which there is evidence of population differentiation in chromosomal factors is one way of limiting this problem, as the extent of divergence is likely to be less well developed.
Common garden experiments

Key to understanding lineage divergence is the measurement of phenotypic variation within and among populations that arise from genetic differences. Common garden experiments have had a long and venerable history in plant evolutionary ecology and formed the foundation of the discipline of genecology (Turesson 1922, Clausen et al. 1940, Langlet 1971). A common garden experiment, whether under field or glasshouse conditions, is used to quantify the genetic component of phenotypic variation by growing seed or clonal material sampled from different populations and/or maternal families under controlled environmental conditions. Common garden experiments are particularly useful for examining the patterns of differentiation for species that have widespread distributions and occupy a wide range of environmental conditions including climatic gradients and populations occupying diverse edaphic conditions (Moloney et al. 2009). Understanding the genetic contribution to phenotypic variation is a key element for studies of population structure as it can point to local adaptation as a potential driver of population differentiation.

Study species: Rumex hastatulus

*Rumex hastatulus* is a wind-pollinated, dioecious colonizer of fields and open, disturbed habitats, usually with sandy soils (Fig. 1). It is distributed throughout the southern United States from Texas to Massachusetts and Florida (Fig. 2). There are two distinct geographic sex chromosome races, hereafter SCRs, which also differ in karyotype; the North Carolina SCR, hereafter NC (females = XX, 2n = 8; males = XY₁Y₂, 2n = 9), and the Texas SCR, hereafter TX (females = XX, 2n = 10; males = XY, 2n = 10; Smith 1964). The occurrence of polymorphism for sex chromosome race within a species is unique in *Rumex* and to my knowledge in flowering plants. All other *Rumex* species with sex chromosome are either fixed for XX and XY or the XX and XY₁Y₂ system. Plants of *R. hastatulus* are primarily winter annuals, although
individuals that appear to be short-lived perennials are reported from populations of the North Carolina SCR (Pickup and Barrett 2012, 2013). Individuals overwinter as basal rosettes typically producing several inflorescences in early spring (February – April). The fruit are winged and seeds rely on wind for dispersal.

Figure 1.1 Distribution of *Rumex hastatulus* in North America; the Texas sex chromosome race is shown in blue and the North Carolina sex chromosome race in green. The hybrid zone is shown in purple. Based on data adapted from Smith (1969) and the Biota of North America Project (BONAP).
Figure 1.2 Individual plants of *Rumex hastatulus*: A) male (right) and female (left) at first flower grown under glasshouse conditions. B) male (left) and female (right) in the field at reproductive maturity. Images courtesy of Spencer Barrett A) and Melina Pickup B)

*Rumex hastatulus* has inhabited its current range for at least the last 150 years; although reports from botanists working in the southeastern United States indicate that there may have been rapid population expansion northward between 1810-1830, associated with forest clearing and the spread of arable agriculture (reviewed in Smith 1964). Significantly, there is evidence of natural hybridization and introgression between the two main chromosomal races discussed above (Jackson 1967; reviewed in Smith 1969). The hybrid zone follows the Mississippi river south from northeastern Arkansas and extends from the river’s banks into central Arkansas, central Mississippi, and parts of Louisiana. Little is known about variation or the current extent of the hybrid zone although it has been suggested that hybrids
constitute a third sex chromosomal race known as the Illinois-Missouri Race (reviewed in Bartkowiak 1971), but this has not been studied in any detail.

Crosses of the TX and NC SCR’s produce viable F_1 hybrids, albeit under glasshouse conditions (Smith 1969). The two main SCRs differ significantly in mean values of leaf characters and in the colouration of their stems and fruit but these differences break down in the hybrid zone and do not provide adequate criteria for consistent classification of the races (Jackson 1967 reviewed in Smith 1969). While Smith’s work has established that the TX and NC races hybridize when deliberately crossed under laboratory conditions, the fertility of the F1 hybrids is unknown and the extent of natural hybridization remains an open question.

The XY_Y sex chromosome system has evolved at least twice in Rumex in Europe and North America, and appears to have had a relatively recent origin in R. hastatulus. Based on nuclear intergenic spacer mean distance corrected estimates of Navajas-Pérez et al. (2005), Del Bosque et al. (2011) estimate the split between SCRs occurred approximately 600,000 years ago. These authors have suggested that the XYY system in Rumex has been derived from the XY system. Smith (1964) also suggested, based on a karyotypic analysis of both sex chromosome races, that the XY_Y likely arose from the XY system following two independent translocations, one between the TX race’s Y chromosome a small autosome pair found only in the TX race and the second between the TX race’s X chromosome and the aforementioned autosome pair, resulting in the splitting of the Y chromosome into two. Rumex hastatulus is a good candidate for studying the early stages of lineage differentiation arising from chromosomal rearrangements. It has a short generation time, a clearly delimited hybrid zone, and both XX/XY and XX/XY_Y sex chromosome systems. The plant is also easily cultured under glasshouse, growth chamber and field conditions, growing to flowering in several months. In the following section, I outline the principal objectives of
Chapter 1 – General Introduction

my thesis and give a brief overview of the experiments I used to address them.

*Thesis objectives*

The main objective of my thesis is to investigate genetic and life history differentiation between and within populations of the SCRs of *Rumex hastatulus*. Below, I detail the main questions addressed in the two research chapters.

Chapter 2 investigates the extent of morphological and life-history differentiation among populations of *R. hastatulus* sampled from throughout the species geographical range, by examining the degree and patterns of differentiation between races, populations, and maternal families grown under uniform glasshouse conditions. The experiment also allowed me to examine whether geographical variation in morphological and life-history traits was correlated with environmental variation, particularly climatic variables.

Chapter 3 addresses whether there are differences between the sex chromosome races in population genetic parameters based on a sample of nuclear genes. It also considers the nature and magnitude of genetic differences within and between SCRs. Three specific questions were addressed in this chapter: (1) Is genetic diversity structured within and between the chromosome races? (2) Are populations of the two SCRs that are located close to the boundary more similar genetically than those that are more widely separated? (3) Can information on population genetic structure provide insights into the demographic history of the SCRs and the timing and direction of the split between the races?

Chapters 2 and 3 were written as research papers for publication in the primary scientific literature. As such, there is a certain amount of repetition between them, and with the introductory chapter.
CHAPTER 2 – MORPHOLOGICAL AND PHENOLOGICAL VARIATION WITHIN AND BETWEEN SEX CHROMOSOME RACES OF RUMEX HASTATULUS

ABSTRACT
Geographical patterns of environmental variation can result in clinal adaptation of ecologically relevant traits, including leaf shape, flowering phenology and plant height. Rumex hastatulus is a wind-pollinated, dioecious plant with a wide range across the southern USA. The two main sex chromosome races are largely separated geographically. The races are reported to differ in several leaf characters, but the extent of differentiation for life-history traits is unknown and previous attempts to differentiate the races based on morphology have been inconclusive. This study addressed three main questions: 1) What is the nature and extent of within- and between-race differences for ecologically relevant traits? 2) Is there evidence of sexual dimorphism for morphological and life-history traits? 3) Do gradients in environmental variation correlate with morphological and life-history variation? I investigated these questions using a common garden experiment with 1163 individuals from 28 populations representing both sex chromosome races sampled from across the range of R. hastatulus. I found significant differentiation between the races for flowering time, leaf width, and number of inflorescences. In addition there were significant differences between the sexes for flowering time, leaf length, stem diameter, inflorescence height, number of inflorescences, and the biomass of aboveground vegetative tissue. I also detected within each of the sex chromosome races evidence for clinal variation in stem diameter, leaf length, number of inflorescences, and vegetative biomass suggesting that these traits may be indicative of adaption to local environmental conditions that vary in a continuous manner throughout the geographical range.
Chapter 2 – Morphological and Phenological Variation

INTRODUCTION

Variation in selection across species’ ranges can be detected by observing geographic patterns of variation in ecologically relevant traits (Mayr 1956, Endler 1977, Caicedo et al. 2004). Because plants are sessile, they can experience substantial variation in selective pressures across their range, which can lead to local adaptation and phenotypic differentiation between populations (Joshi et al. 2001, Streisfeld and Kohn 2005). Latitudinal variation in day length, solar angle, temperature, and humidity impose clines of selection on phenotypic traits, notably leaf shape and flowering time (e.g. Stinchcombe et al. 2004, Hopkins et al. 2008). Additionally, longitudinal variation in climate can also cause clines for flowering time (Samis et al. 2012). The interplay of multiple environmental and geographic factors can lead to significant morphological and phenological variation among natural populations; however, most studies of clinal variation in ecologically relevant traits have focused on correlations between a single trait and latitude.

Plants may experience strong natural selection to maximize photosynthesis while minimizing water loss from transpiration. Because leaves are the principal site of photosynthesis, leaf morphology is especially important for influencing rates of transpiration and photosynthesis, and as such they tend to vary along climatic gradients, with leaf size often correlating with mean annual temperature (Givnish 1984, Wolfe 1995, Nunez-Olivera et al. 1996) and decreasing mean annual rainfall (Cowling and Campbell 1980, McDonald et al. 2003). There have been few studies of geographical variation in leaf traits and their association with other life-history traits throughout the entire range of widely distributed species occupying diverse climatic regimes.

Regional selection for a suite of life history and morphological traits can be counteracted by extensive gene flow among populations under selection for different trait expression. The particular suite of life history
traits that characterize a species is important for making predictions about
the extent of gene flow and genetic differentiation among populations
(Hamrick et al. 1979). For example, the pollination system and lifespan of
species are especially relevant, such that wind-pollinated, long-lived species
tend to show less differentiation between populations across their range than
short-lived animal-pollinated species (Hamrick and Godt 1996). Studies on
patterns of genetic differentiation in wind-pollinated plants with annual life
histories have tended to focus on introduced invasive species that are highly
self-pollinating (e.g. *Bromus tectorum*, Mack and Pyke 1983; *Avena barbata*,
Clegg and Allard 1972). Such species often show less variation than native
species because of small founder populations. Thus, relatively little is known
about patterns of genetic differentiation in native, annual, wind-pollinated
species, particularly those that are obligately outbreeding. This knowledge
gap partly motivated the present study.

Here, I examine patterns of morphological and life-history variation
across the geographical range of *R. hastatulus*, an annual, wind-pollinated
dioecious species native to the southern USA. Previous studies of this species
have recognized two main sex chromosomal races (described in Chapter 1 and
see below) that differ to some extent in leaf traits (reviewed in Smith 1969),
but there have been no comprehensive comparisons of the SCR’s throughout
their geographical range, and it is unclear the extent to which the two races
can be reliably distinguished from one another. Moreover, because *R.
hastatulus* is dioecious, patterns of morphological and phenological variation
may also differ between females and males, a phenomenon termed sexual
dimorphism (reviewed in Barrett and Hough 2013, and see Pickup and
Barrett 2012 for *R. hastatulus*). Therefore, I was also interested in examining
this scale of variation in the two SCRs of *R. hastatulus*. This chapter
therefore addresses three main questions: 1) What are the patterns of
morphological and life history differentiation within and among the SCRs; 2)
What traits distinguish females and males, and does this differ between the
SCRs; 3) Is there evidence that geographical variation in morphological and life-history traits correlates with environmental gradients and thus they exhibits clines?

MATERIALS AND METHODS

Study system

*Rumex hastatulus* (Polygonaceae) is a coloniser of fields and open, disturbed ground, with a broad distribution throughout the southern USA from Texas to North Carolina and Florida. There are two main geographic races that also differ karyotypically: the North Carolina (hereafter NC) SCR (females = XX, 2n = 8; males = XY,Y, 2n = 9) and the Texas (hereafter TX) SCR (females = XX, 2n = 10; males = XY, 2n = 10; Smith 1964). Plants overwinter as basal rosettes and produce inflorescences in spring. The fruit are winged and seeds are wind dispersed.

Glasshouse experiment

In January 2012, I germinated 75 field-collected seeds from 15 randomly chosen maternal families from each of 28 populations for a total of 1163 plants of *R. hastatulus*. The location of sampled populations spanned the geographical ranges of both sex chromosome races (Figure 2.1). Further information on the size of individual populations and their sex ratio is detailed in Pickup and Barrett (2013). The seed samples were open-pollinated families collected from the field in April 2010. Seeds were placed in Petri dishes on moist filter paper in a growth cabinet maintained at 4 °C for approximately 48 hours before transferring 60 seeds per population to 3cm pots containing Sunshine Mix #1 (peat moss, coarse perlite, Gypsum, and dolomitic limestone) and Osmocote 14-14-14 in the University of Toronto glasshouse. Transplanting the seedlings to pots was conducted during February 21-22, 2012. The glasshouse was maintained at 28 °C with a 16h light 8h dark cycle, with supplemental light provided by high pressure
sodium lamps. Once establishment was confirmed, seedlings were placed in a complete randomized block design on February 27 – March 2, 2012. Throughout the growth period I recorded the date at which plants first flowered, hereafter flowering time, and the gender of plants that flowered. After 4 weeks I measured the length and width of the longest leaf for each plant. At reproductive maturity (after approximately 10 weeks of growth since transplanting) I measured length and width of the longest leaf, height and stem diameter of the tallest inflorescence, and the number of inflorescences for each plant. From May 10, 2012 to May 20, 2012 I harvested plants and dried them in a drying ovens at 50 °C for two weeks and measured aboveground biomass for reproductive and vegetative tissues.

Figure 2.1 – The geographical distribution of *Rumex hastatulus* population investigated in the glasshouse experiment. Populations from the Texas sex chromosome race are shown with red circles whereas populations from the North Carolina sex chromosome race are shown in blue. Map was generated using ArcGIS online.
Analyses of quantitative trait differentiation between sex chromosome races and genders

All statistical analyses were performed using the R.app GUI 1.53 (6451 Leopard build 64-bit) (R Core Team 2013). I first standardized trait means and variances so that they had means of 0 and unit variance. I checked for normality using the Shapiro-Wilk test (Royston 1995) and by examining quantile-quantile plots. Then, to achieve normality and homoscedasticity, I log-transformed the biomass measurements. Standardizing trait means and variances is useful as it makes it possible to perform meaningful comparisons between variables with different scales and units of measurement (Jongman et al. 1995).

I performed univariate comparisons between plant traits using nested mixed effects ANOVAs and linear regressions. I then partitioned the proportion of variance attributable to each explanatory variable, using generalized linear mixed effects models in lme4 library (Bates et al. 2012), and partitioned the components of variance attributable to maternal family, population of origin, and SCR effects for each of the response variables: time to first flower, length and width of the largest leaf, number of inflorescences, stem diameter, and vegetative biomass. Because gender-specific differences had previously been shown to occur in *R. hastatulus* (Conn and Blum 1981a, 1981b, Pickup and Barrett 2012) I set race and gender as fixed effects and maternal family nested within population, which was in turn nested within race, with races nested in experimental block as random effects. Fitting was done using a restricted maximum likelihood (REML) function because of unbalanced data (Patterson and Thompson 1971). I quantified the proportion of total variance described by predictor variables in the model using the pvals.fnc function in the languageR package (Baayen 2011). This function calculates *P*-values and high probability density intervals for model parameters based on Markov Chain Monte Carlo (MCMC) sampling of the posterior distribution. Models involving experimental block as a predictor
variable were consistently worse than those without it, so the term was removed from analyses.

I performed multivariate analyses using principal component analysis (PCA) and permutational multivariate analyses of variance. Reproductive biomass was omitted from the multivariate analyses because it was highly skewed and did not meet the assumptions of homogeneity of variance. The data set used for PCA analyses consisted of all other variables listed above. Individuals’ sex and chromosome race were used as indicator variables to help visualize relationships in figures. I performed the PCA with scaled means and unit variance to account for the different scales used to measure the traits of interest. For individuals that flowered, I also compared phenotypic traits between SCRs and genders, with family effects nested with populations, and populations nested within race, using a permutational multivariate analysis of variance with 100,000 permutations using the ADONIS function in the VEGAN package (Oksanen et al. 2013).

**Climatic variation in ecologically relevant traits**

Prior to the analyses, I calculated Pearson correlations between latitude and 30-year means (1980-2010) of several climatic variables and between longitude and the same variables. I obtained values for mean annual precipitation, degree days below 0 °C, and degree days above 32 °C from the US National Climate Data Center. Data had not been collected on degree days below 0 °C and degree days above 32 °C for five of the data centres nearest to six TX populations and one data centre nearest to a NC population. As there were not data available for climate monitoring centers within 50 km of some of the populations with missing data, and because latitude and longitude were correlated with the environmental variables, I used latitude and longitude as proxies for environmental variation in subsequent analyses. I performed a multivariate analysis of covariance using the ANOVA function of the car package in R (Fox and Weisberg 2011) to test
for the effects of latitude and longitude on flowering time, leaf length and width, stem diameter, number of inflorescences, inflorescence height, and vegetative biomass, with sex chromosome race and sex included as covariates. I used a MANCOVA to mitigate possible correlations between response variables (Scheiner 2001), and because longitude and race covaried due to the geographical separation of the SCRs. I then performed individual ANCOVAs for the effects of latitude and longitude to examine patterns of clinal variation for individual traits.

I calculated the proportion of reproductive individuals per population. Wild populations of *R. hastatulus* flower in April and their fruit matures in May (Smith 1963). Individuals were considered reproductive if they had begun flowering by the time plants were harvested for biomass measurements in mid-May, which should have given the plants ample time to initiate flowering had they been growing in the field. I used a regression analysis to investigate the relation between latitude and the proportion of individuals initiating flowering for the NC SCR only, because all individuals of the TX race flowered.

RESULTS

*Trait differences between the chromosomal races and genders*

The SCRs of *R. hastatulus* differed significantly for flowering time, leaf width, and the number of inflorescences, with individuals from the TX race flowering earlier and having wider leaves and more inflorescences (Table 2.1, Figure 2.2). The TX race consisted entirely of individuals with an annual life history, as senescence of plants was evident at harvest. In contrast, populations of the NC race varied considerably in the proportion of individuals that flowered in the experiment (Figure 2.3). When the variance of phenotypic traits was partitioned into family, population, and SCR effects,
between-population variation within races accounted for the greatest proportion of variance (Figure 2.4).

Overall, populations of the NC race were more female biased than the TX race (Average percent female NC=56.23%, TX=51.37%; Type II ANOVA: \(F=7.81, \text{df}=1, P=0.01\)). There were significant differences between females and males for several life history and phenotypic traits. Females tended to flower earlier, have longer leaves, taller inflorescences after flowering, fewer inflorescences, thicker stems, and greater vegetative biomasses than males (Table 2.1). The degree of sexual dimorphism for flowering time, width of the largest leaf, and diameter of the largest inflorescence differed in magnitude among SCRs, such that males from the NC SCR tended to flower later than females, whereas both genders flowered at approximately the same time in populations of the TX SCR. Additionally, females of the NC SCR had wider leaves than males, but both genders had approximately equal leaf widths in populations of the TX race.

Table 2.1 – Two-way mixed effects nested ANOVAs on the influence of sex chromosome race, gender and their interaction on flowering time (FT), length (LEN) and width (WID) of the largest leaf, number of inflorescences (NUM), height of the largest inflorescence (HEI), stem diameter (DIAM), and vegetative biomass (BIO) in populations of *Rumex hastatulus*. Maternal family was nested within population of origin, which was in turn nested within chromosome race. Values in bold are significant at \(P<0.01\). See table S2 for means and variances for each population’s traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Sex chromosome race</th>
<th>Gender</th>
<th>Race x Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(F) (P)</td>
<td>(F) (P)</td>
<td>(F) (P)</td>
</tr>
<tr>
<td>FT</td>
<td>4.47 (0.029)</td>
<td>16.59</td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>LEN</td>
<td>2.84 (0.12)</td>
<td>45.46</td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>WID</td>
<td>38.64 (&lt;0.0001)</td>
<td>3.64</td>
<td>0.056</td>
</tr>
<tr>
<td>NUM</td>
<td>9.67 (0.0024)</td>
<td>17.66</td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>HEI</td>
<td>0.0002</td>
<td>0.743</td>
<td>174.95</td>
</tr>
<tr>
<td>DIAM</td>
<td>1.16</td>
<td>0.195</td>
<td>217.85</td>
</tr>
<tr>
<td>BIO</td>
<td>0.95</td>
<td>0.23</td>
<td>221.82</td>
</tr>
</tbody>
</table>
Figure 2.2 – Proportion of individuals within each population of the North Carolina and Texas sex chromosome races of *Rumex hastatulus* that flowered in the glasshouse experiment. The identity of each population is indicated (see supplementary Table S1 for specific localities).
Figure 2.3 – Grand means for three life history traits for the North Carolina and Texas sex chromosome race of *Rumex hastatulus* in the glasshouse
experiment: A) flowering time (days), B) leaf width (mm), and C) number of inflorescences. Bars show ±standard error.

![Graph showing proportion of variance explained by maternal family, population, chromosome race, and residual variance from linear mixed effects ANOVA models.]  

Figure 2.4 – Mean proportion of total variance explained by maternal family, population, chromosome race, and residual variance from linear mixed effects ANOVA models calculated independently for: flowering time, length and width of the largest leaf, diameter and height of the tallest inflorescence, number of inflorescences, and vegetative biomass measured in the glasshouse experiment on *Rumex hastaulus*. Error bars are standard deviations of the grand mean calculated for each level of the model.
Multivariate analyses of differentiation between sex chromosome races and genders

The first three axes of the PCA performed on length and width of the largest leaf, flowering time, inflorescence height, stem diameter, number of inflorescences, and vegetative biomass accounted for 44.80%, 18.25%, and 11.10%, respectively, of the variability and combined accounted for 75.61% of the total variability in the data (Figure 2.5). I included the first three principal components because they capture most of the variation in the data and there are significant differences between SCRs along the second and third axes. The first axis was highly correlated with length of the longest leaf, inflorescence height, stem diameter, and vegetative biomass; the second axis was highly correlated with time to first flower and number of inflorescences, and the third axis was highly correlated with width of the largest leaf (Supplementary table S3). The first principal component differed significantly between the genders, while the second and third principal components differed significantly between the sex chromosome races and genders (Table 2.3). When all traits were included in a nested permutational analysis of variance, there were significant differences between races, genders, and populations (Table 2.4).

Table 2.2 – Results of two-way ANOVAs testing for the effects of sex chromosome race, gender, and their interaction on Principal Components comprising flowering time (FT), length (LEN) and width (WID) of the largest leaf, number of inflorescences (NUM), height of the largest inflorescence (HEI), stem diameter (DIAM), and vegetative biomass (BIO), based on data collected from the glasshouse experiment on *Rumex hastatus*. Values in bold are significant at $P<0.05$ after Bonferroni correction ($P_{corr} < 0.017$).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>1</td>
<td>0.702</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Table 2.3 – Permutational multivariate analysis of variance (PERMANOVA) using the ADONIS function in R comparing time to first flower, length and width of the largest leaf, number of inflorescences, height of the tallest inflorescence, stem diameter, and vegetative biomass between sex chromosome races and genders of *Rumex hastatulus*. Values in bold are significant at *P* < 0.05.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th><em>F</em></th>
<th><em>P</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>1</td>
<td>10.83</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>116.96</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Population</td>
<td>25</td>
<td>5.70</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Race x Gender</td>
<td>1</td>
<td>2.70</td>
<td>0.0885</td>
</tr>
<tr>
<td>Residuals</td>
<td>903</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.5 – Three scatterplots (A) PC2 against PC1, (B) PC3 against PC1 C) PC3 against PC2) of a PCA analysis with scaled means and variances of 928 individuals of *Rumex hastatulus* for seven phenotypic variables (time to first flower, length and width of the largest leaf, number of inflorescences, height
of the tallest inflorescence, diameter of the principal inflorescence’s stem, and vegetative biomass). The Texas sex chromosome race is plotted with open points and the North Carolina sex chromosome race with closed points; females are triangles and males are circles.

Clinal variation of ecologically relevant traits

There were significant correlations between latitude and degree days below 0°C \( (P=0.004; r=\pm 0.60) \) and longitude and degree days above 32°C \( (P=0.002; r<\pm 0.63) \). The results of the MANCOVA indicated that there were significant effects of both latitude and longitude on the suite of traits included in the analyses (Table 2.4). Individual ANCOVAs performed on a trait-by-trait basis revealed that latitude had more of an effect on most traits than longitude. Plants from higher latitudes (more northerly populations) tended to flower later and had longer leaves, thicker stems, fewer inflorescences, and greater vegetative biomasses (Figure 2.6 A-E). There was also a strong latitudinal cline for the proportion of individuals that flowered in the NC race (Figure 2.6 F). All individuals flowered in all of the TX race populations so there was no evidence of a similar cline.

Table 2.4 – Multivariate analysis of covariance of the effects of latitude and longitude on length and width of the largest leaf, stem diameter, height of the tallest inflorescence, number of inflorescences and vegetative biomass of the sex chromosome races of *Rumex hastatulus* grown under glasshouse conditions. Gender and sex chromosome race were included as covariates.

<table>
<thead>
<tr>
<th>Factor</th>
<th>df</th>
<th>SS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude</td>
<td>1</td>
<td>33.15</td>
<td>36.808***</td>
</tr>
<tr>
<td>Longitude</td>
<td>1</td>
<td>29.01</td>
<td>32.214***</td>
</tr>
<tr>
<td>Race</td>
<td>1</td>
<td>43.48</td>
<td>48.279***</td>
</tr>
</tbody>
</table>
Chapter 2 – Morphological and Phenological Variation

Gender

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>17.37</th>
<th>19.281***</th>
</tr>
</thead>
</table>

*P<0.05, **P<0.01; ***P<0.001

A)  
Flowering time (days) vs. Latitude

B)  
Leaf length (mm) vs. Latitude

C)  
Stem diameter (mm) vs. Latitude

D)  
Inflorescence height (mm) vs. Latitude
**Figure 2.6** – The relations between latitude and A) flowering time, B) leaf length, C) stem diameter, D) inflorescence height, E) biomass of vegetative tissue, and F) proportion of reproductive individuals from the glasshouse study of *Rumex hastatulus* populations. A-E) depicts pooled data for both sex chromosome races, whereas F) is based on data from the North Carolina race only. Values for vegetative biomass were transformed to have a mean of 0 and unit variance. Solid lines indicate a significant linear regression for the trait value. Dashed lines indicate 95% confidence intervals.

**DISCUSSION**

In many plant populations, life history and morphological traits are frequently adapted to local climatic and environmental conditions. By conducting a common garden experiment under glasshouse conditions, I was able to reduce the influence of plastic responses to growing conditions that confound surveys of variation among natural populations occurring over a wide geographical area. I used a large sample of populations from across the native range of *R. hastatulus* to identify patterns of genetically-based variation for several ecologically-relevant traits. Of particular importance was to determine the extent to which the two main sex chromosome races of *R. hastatulus* could be distinguished. I also examined variation between the
genders to investigate the extent of sexual dimorphism in the two sex chromosome races. Thus, the research in this chapter set out to address the following specific questions. 1) What are the patterns of morphological and life-history differentiation within and among the sex chromosome races? 2) What traits distinguish females and males, and does this differ between the sex chromosome races? 3) Is there evidence that geographical variation in morphological and life-history traits correlates with variation in environmental conditions? Below I discuss the ecological and evolutionary significance of the main results that were revealed by my common garden study.

**Can the sex chromosome races be distinguished by phenotypic traits?**

Previous work by Jackson (1967 reviewed in Smith 1969) established that the SCRs of *R. hastatulus* differ significantly in the mean values of leaf characters and in the colouration of their stems and fruit, but the extent of this differentiation has been considered insufficient to provide reliable criteria with which to identify specimens of unknown origin. In examining a broader range of variables that included life history, as well as morphological traits, I found that the sex chromosome races of *R. hastatulus* are significantly differentiated for several traits including flowering time, leaf width, and number of inflorescences. The races also differed in terms of their generation length, with the TX race being exclusively annual, while all populations of the NC race had some individuals displaying a short-lived perennial life history. Wild populations begin flowering in early April (Smith 1963), giving roughly four weeks longer to initiate flowering than they would have experienced under field conditions makes it unlikely that the observed differences in life history between the SCRs is an artifact of harvesting too early. These phenotypic and life history differences largely represented quantitative variation that transcended significant population differentiation within each SCR. I detected no obvious qualitative differences between the
SCRs in the glasshouse experiment thus identifying and distinguishing specimens of the two SCRs without knowing where they were collected would be very difficult.

Latitudinal clines for flowering time have been detected for several species, including *Arabidopsis thaliana* (Stinchcombe et al. 2004), *Beta vulgaris* (Van Dijk et al. 1997), and *Brassica rapa* (Franke et al. 2006). Flowering time in *R. hastatulus* was significantly earlier in TX populations and significantly correlated with latitude. Theoretical predictions of the relationship between life history and environmental conditions indicate that an annual life cycle is favoured in regions with a shortened growing season, particularly owing to the cessation of rainfall (Iwasa and Cohen 1989). Thus, the 100% proportion of reproductive individuals observed in populations of the TX SCR and varying rates of flowering in populations of the NC SCR confirms the expectation of an annual life history correlated with growing season length and amounts of precipitation. Rainfall in Texas ceases in early spring but continues throughout the spring and summer throughout the range of the NC SCR.

*Patterns of sexual dimorphism in the sex chromosome races*

In addition to differentiation for flowering time between sex chromosome races, I also found that males and females differed in this trait. Numerous studies of flowering time differentiation between the sexes of dioecious species have found that males flower earlier than females (e.g. Allen 1986, Armstrong and Irvine 1989, Espírito-Santo et al. 2003, Barrett and Hough 2013); however, I found the opposite pattern: females flowered significantly earlier than males, consistent with the results of Pickup and Barrett (2012), but a result that is uncommon for dioecious species (reviewed in Lloyd and Webb 1977). The later onset of flowering in males may be due to the high nitrogen requirement of pollen production of wind-pollinated species (Harris and Pannell 2008). An alternative explanation is that selection for improved
pollen dispersal may result in males delaying flowering until they reach a greater height.

In addition to differentiation between the genders in their time to first flower, I detected significant sexual dimorphism for several morphological traits. Similar to the results of Pickup and Barrett (2012), I found significant differences between females and males for height at seed maturity. Potential explanations for the adaptive significance are explored more thoroughly in the aforementioned paper; briefly, it is probable that the greater height of females promotes more effective seed dispersal. Males and females also differed in the length of their longest leaves, the number of inflorescences produced, the diameter of their largest inflorescence, and the biomass of their vegetative tissue. These differences were generally in the same direction as those seen in other flowering plants (Delph 1999, Osso 2002) and are probably associated with differences between the gender in reproductive function and costs of reproduction.

**Correlations of phenotypic traits with environmental variables**

Including latitude and longitude as covariates is an important consideration in analyses of phenotypic differentiation, as unaccounted geographic clines can invalidate inferences on the adaptive basis of variation. For example, in a meta-analysis of 32 common garden studies of 28 plant species, Colautti et al. (2009) found that including latitude as a covariate in analyses reduced the magnitude and reversed the direction of the effects of life-history and physiological trait differentiation between native and introduced species. Interestingly, I found that adding terms for latitude and longitude to multivariate mixed effects models of trait variation actually increased the effect size of sex chromosome race, but kept the effect size of gender the same. This suggests that the phenotypic differences between SCRs that I detected are due to more than just their geographic separation from one another and are the result of factors yet to be explored. As expected, because of their
spatial separation along a longitudinal axis, the inclusion of an effect of longitude in the models increased the effect size of sex chromosome race to a greater extent than including the effect of latitude.

Clines in plant sensitivity to flowering cues have been detected in *A. thaliana* (Stinchcombe et al. 2005), *Cirsium vulgare* (Wesselingh et al. 1994), and *B. vulgaris* (Boudry et al. 2002), such that plants from higher latitudes require stronger cues, generally in the form of longer exposure times, to respond to vernalization treatments and initiate flowering. This may be responsible for the correlation between latitude and the proportion of reproductive individuals in the NC sex chromosome race. Because response to vernalization tends to require a threshold size below which no response is seen (Klinkhamer and de Jong 1989, Wesselingh et al. 1993), one would predict that the non-flowering individuals should be smaller before the onset of flowering, which is consistent with what I observed. I found that individuals that did not initiate flowering had significantly narrower and shorter leaves before the onset of flowering and significantly longer leaves and greater biomass of vegetative tissue (data not shown).

Leaf size tends to positively correlate with mean annual temperature and correlate negatively with mean annual rainfall (Dolph 1977, Givnish 1984, Wolfe 1995, McDonald et al. 2003). As predicted, I detected larger average leaf width in populations of the TX SCR than the NC SCR, which may be explained by the shorter growing season, warmer daily maximum temperatures and less rainfall experienced by these populations.

My glasshouse experiment documented significant differences in the mean values for a range of morphological, phenological and life-history traits between the SCRs of *R. hastatulus*. These differences were relatively subtle and would not allow identification of specimens to SCR, without knowledge of the origin of the population from which they were collected. Thus, despite occurring over a broad range of climates spanning a large area of the southern USA, there is not a high degree of differentiation for phenotypic
traits in *R. hastatulus*, in comparison with many animal-pollinated species. The lack of strong population differentiation is probably, in part, the result of extensive gene flow as both the pollen and seeds of *R. hastatulus* are transported by wind, and thus likely to be widely dispersed. Another aspect of the ecology of *R. hastaulus* probably also contributes to the absence of strong differentiation in phenotypic traits. Populations occupy similar open, weedy arable habitats regardless of where they occur in the geographical range. Thus, the quantitative variation in morphology, phenology and life history I detected is most likely governed by large-scale environmental gradients in climatic conditions rather than local habitat conditions.
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

**CHAPTER 3 – MOLECULAR GENETIC DIFFERENTIATION AND DEMOGRAPHIC HISTORY OF THE SEX CHROMOSOME RACES OF *RUMEX HASTATULUS***

**ABSTRACT**

Measuring population genetic structure can provide insights into the diverse ecological, evolutionary, and historical factors causing lineage differentiation in plant and animal groups. Life-history traits and karyotypic differentiation influence the nature and extent of genetic structure in species. Obligately outcrossing species with high dispersal ability are predicted to show minimal population genetic structure. However, chromosomal changes between populations can act as barriers to gene flow. I examined the extent of population genetic structure in the wind-pollinated, dioecious, annual plant *Rumex hastatulus* (Polygonaceae), which is distributed across the southern USA. This species has two distinct chromosome races (hereafter SCRs; North Carolina: females = XX, 2n = 8; males = XY, 2n = 9 and Texas: females = XX, 2n = 10; males = XY, 2n = 10). I performed a Bayesian analysis using STRUCTURE of population genetic structure based on DNA sequences at 13 nuclear loci from 28 populations sampled across the range of *R. hastatulus* to examine genetic differentiation. I found that individuals could be divided into two distinct clusters that corresponded with the two SCRs, despite the absence of fixed differences in alleles. The TX SCR had a value of Tajima’s *D* close to zero and more unique polymorphisms than the NC SCR, which had a negative Tajima’s *D*, consistent with an earlier proposal of a fairly recent origin of the NC SCR. Estimates of population genetic parameter values using Polymorphurama suggest an ancestral θ value between 0.001 - 0.0025 and a divergence between the lineages of ~5000 and 15000 generations ago. My results indicate that despite ongoing gene flow between populations there is genetic differentiation between the SCRs and that the NC race is likely derived from the TX race.
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

**INTRODUCTION**

Investigating patterns of population genetic structure and levels of diversity within species can provide insights into the forces influencing genetic variation. Work on this topic is a crucial element of testing theoretical predictions on the relative importance of evolutionary, ecological, and historical processes in structuring genetic diversity. Population genetic structure is also strongly influenced by the life-history traits of organisms, with features of the reproductive biology of populations especially influential (Hamrick et al. 1979). Aspects of the genetic system that influence reproductive isolation, including variation in chromosomal structure and polyploidy, can also play an important role in affecting population genetic structure through their influence on limiting gene flow between populations (Coyne and Orr 2004). Studying the interplay between reproductive traits and the genetic system of populations in contributing to population subdivision can provide novel information on the early stages of lineage divergence and speciation.

Early studies on the patterns of enzyme polymorphism in plant populations revealed high levels of genetic diversity, especially in outcrossing species (Solbrig 1972, Brown 1979). As more taxa were surveyed, ecological geneticists began to examine the relations between patterns of genetic variation and ecological and life-history traits. Large-scale comparative surveys allowed researchers to make predictions about population genetic structure based on information regarding the species’ geographic range, population size, and diverse features of their life histories and reproductive systems (e.g. Hamrick et al. 1979). In particular, the pollination and mating system were shown to be influential in affecting population genetic structure and levels of diversity. For example, long-lived, wind-pollinated, outcrossing species, such as many temperate trees, were generally found to maintain
high levels of allozyme diversity within populations, but relative little genetic differentiation among populations. In contrast, annual, colonizing species with high rates of self-fertilization most commonly possessed low levels of diversity within populations, but a high degree of differentiation among populations (Hamrick and Godt 1996). As different life-history traits can have contrasting effects on population structure, the observed patterns will depend on the relative influence of each trait, as well as historical factors associated with the demography of species, such as the occurrence of genetic bottlenecks and rates of population expansion.

It is well established that changes in chromosome number or structure play important roles in the evolution of reproductive isolation in flowering plants (Stebbins 1971, Rieseberg 2001, Levin 2002). While there have been extensive investigations of the key role that sex chromosomes play in affecting lineage differentiation and speciation in animals (Coyne and Orr 1989, Presgraves 2002, Turelli and Moyle 2007), there have been few such studies in dioecious plants that possess sex chromosomes, presumably because plants with sex chromosomes are quite rare among flowering plants being restricted to ~12 flowering plant families (Charlesworth 2002). Despite their infrequent occurrence, plants with sex chromosomes can provide useful insights into the initial stages of sex chromosome evolution because their sex chromosomes are generally more recently evolved from autosomes than those of mammals or other animal groups, which have been the subject of extensive genetic studies e.g. Drosophila (reviewed in Westergaard 1958, Charlesworth and Charlesworth 1978). Sex-linked genes in plants show evidence of faster rates of evolution relative to autosomes (Filatov and Charlesworth 2002, Whittle and Johnston 2002), suggesting that they may play an important role in evolutionary divergence. Chromosomal regions with faster rates of evolution are expected to differentiate more rapidly than the genome-wide level of differentiation for divergent lineages (Wu and Davis 1993). This
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

process can initiate population differentiation and potentially lead to strong genetic subdivision and reproductively isolated populations.

Most studies of the evolution of reproductive isolation have focused on related species pairs, many of which have diverged for extended periods of time (reviewed in Coyne and Orr 2004). These studies while valuable are often unable to provide insights into the very early stages of lineage differentiation. This problem can be overcome by investigating species in which there is evidence of population differentiation in chromosomal factors, as the extent of divergence is likely to be less well developed. This is the approach used in this investigation.

*Rumex hastatulus* is a wind-pollinated, dioecious colonizer of fields and open, disturbed habitats usually with sandy soils. It is distributed throughout the southern United States from Texas to Massachusetts and Florida. There are two distinct geographic races that also differ in karyotype; the North Carolina sex chromosomal race, hereafter NC (females = XX, 2n = 8; males = XY₁Y₂, 2n = 9) and the Texas chromosomal race, hereafter TX (females = XX, 2n = 10; males = XY, 2n = 10; Smith 1964). The occurrence of a polymorphism for sex chromosome race is unique in *Rumex*, and to my knowledge in flowering plants. All other *Rumex* species with sex chromosome are either fixed for the XX and XY or the XX and XY₁Y₂ system. *Rumex hastatulus* is therefore a good candidate for studying the early stages of lineage differentiation arising from chromosomal variation. The chromosomal differentiation in *R. hastatulus* arose independently in this species (Navajas-Pérez et al. 2005), and while the races produce fertile, viable hybrids under laboratory conditions (Smith 1969), the extent of hybridization and gene flow between SCRs is unknown for wild populations.

In this study, I investigate patterns of DNA sequence polymorphism in the two sex chromosome races of *Rumex hastatulus*. I used Sanger sequencing of nuclear exon regions to compare levels of polymorphism at synonymous and non-synonymous sites. Additionally, I used the software
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

STRUCTURE (Pritchard et al. 2000) to investigate genetic structure within and among populations and Polymorphurama to investigate the demographic history and evolutionary origins of the SCRs. Specifically, my aim was to address three main questions:

1. What are the levels of genetic diversity within and between the chromosome races and how is the variation structured across the geographical range?

2. What is the extent of gene flow between SCRs, and are populations that are located close to the boundary between the chromosome races more similar genetically than those that are more widely separated?

3. Is there evidence based on demographic history of the timing of lineage differentiation and whether the NC SCR is derived from the TX SCR?

**MATERIALS AND METHODS**

*Population sampling*

In January 2012, I germinated 75 field-collected seeds from 15 randomly chosen maternal families sampled from each of 28 populations of *R. hastatulus*. The populations were comprised of 15 populations of the Texas Race and 13 populations of the North Carolina race. The location of sampled populations spanned the geographical ranges of both chromosomal races (Fig. 3.1). Further information on the size of populations and their sex ratio is detailed in Pickup and Barrett 2013). Seeds were placed in Petri dishes on moist filter paper in a growth cabinet maintained at 4 °C for approximately 48 hours before transferring 60 seeds per population to 3cm pots containing Sunshine Mix #1 (peat moss, coarse perlite, Gypsum, and dolomitic limestone) and Osmocote 14-14-14 in the University of Toronto glasshouse. The glasshouse was maintained at 28 °C with a 16h light 8h dark cycle. Seedlings were placed in a complete randomized block design. After 10
weeks, when plants had reached reproductive maturity, I collected leaf tissue from all individuals and immediately froze it in liquid nitrogen. From this sample I chose one male and one female individual from different maternal families from each of 24 populations for subsequent DNA analysis.

Figure 3.1 – Map of populations of *Rumex hastatulus* used in these analyses. Populations from the Texas sex chromosome race are shown with red circles and populations from the North Carolina race are shown in blue.

**Primer design, DNA extractions and Sanger sequencing**

To examine patterns of DNA polymorphism within and between populations and sex chromosome races, I performed targeted resequencing of 13 putative genes derived from transcriptome data. I extracted genomic DNA from frozen leaf material using Qiagen DNeasy plant kits following the manufacturer’s instructions. Based on a *de novo* transcriptome assembly obtained from J. Hough (University of Toronto, Canada), I performed BLASTx (Altschul et al. 1997) searches to find conserved exon regions for targeted resequencing. PCR
primers were designed to amplify 650-850 bp regions for 21 of these exon sequences using Primer Quest (Integrated DNA Technologies). All primers were 24 bp in length. I selected primers with melting temperatures within 1.6 °C of 57 °C. Based on the quality of the sequencing results I reduced the number of sequences that were analyzed to 13. A standard PCR reaction was performed with 50ng of DNA in 25 µg wells. I used DreamTaq Green PCR Master Mix (Thermo Scientific), following the directions provided by the manufacturer. The reactions for each locus ran for 35 cycles, which consisted of 30s of denaturing at 95 °C, 30s of annealing at 52 °C, and 30s of extension at 72 °C; the initial denaturing step was 3 minutes at 95 °C and the final extension step was 4 minutes at 72 °C. The primers for the amplification were used to sequence forward and reverse strands of the amplicons at the Genome Québec Innovation Centre at McGill University.

Analyses of polymorphism

I used PHASE 2.1 (Stephens and Donnelly 2003; Stephens et al. 2011), as implemented in DNAsp v4.50 (Rozas et al. 2003) to infer haplotypes for the 13 loci sequenced in the 24 R. hastatulus populations. I used a modified version of Polymorphurama (Bachtrog and Andolfatto 2006) to calculate diversity statistics, including θw (Watterson 1975), π (Tajima 1983), and Tajima’s D (Tajima 1989), for synonymous and non-synonymous sites for the 13 loci. I also calculated the number of shared and fixed polymorphisms between the sex chromosome races. I used custom Perl scripts created by S. I. Wright (University of Toronto) to calculate the number of synonymous and non-synonymous unique and shared polymorphisms, and fixed differences between the two chromosomal races.

STRUCTURE analysis

I removed singleton polymorphisms from each of the 13 loci using custom scripts before assigning haplotypes to sequence for analysis using the
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

software program STRUCTURE v2.3.4 (Pritchard et al. 2000, Falush et al. 2003, 2007, Hubisz et al. 2009). STRUCTURE uses a Bayesian approach to model-based clustering and identifies clusters based on fit to Hardy–Weinberg equilibrium and linkage equilibrium. Analyses allowed for admixture, with $K$ values ranging from 1-10, run in quintuplicate per value of $K$ (the number of populations), burn-in was set to 100,000 iterations, with each run consisting of 1,000,000 iterations. I used the method of Evanno *et al.* (2005) as implemented in STRUCTURE Harvester (Earl and vonHoldt 2012) to infer the true number of genetic clusters based on the $\Delta K$ statistic, which is calculated as the rate of change in the log probability of data between successive $K$ values. I used CLUMPP (Jakobsson and Rosenberg 2007) to align the cluster assignment probabilities for each value of $K$.

Coalescent simulations

To investigate the demographic history of *R. hastatulus* and to look for evidence of a genetic bottleneck that may be evident if the NC race was recently derived from the TX race, I used the MIMAR noanc package (Becquet and Przeworski 2007) to perform coalescent simulations. MIMAR allows for recombination between loci while fitting the observed number of shared, unique, and fixed polymorphisms to a standard Isolation-with-Migration model, where following divergence at time $T$, populations experience an instantaneous time change. After the time change, they may or may not experience gene flow subsequent to divergence in the next generation of the simulation. I estimated the number of generations since divergence and the extent and direction of gene flow. I ran MIMAR for 200,000,000 runs with a burn-in period of $10^5$ steps with the following parameter values: $\theta$ for the two chromosomal races ranging from 0.001 to 0.05; ancestral $\theta$ was constrained to 0.0005 to 0.003; migration between populations was set between $e^{-7}$ and $e^{5}$ individuals per generation; divergence time was set to between 0 and 18,000 generations; mutation rate was set to
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

7e-9, based on the average substitution rate for plants’ nuclear DNA (Wolfe et al. 1987). The parameter values were chosen by iteratively altering the prior values and examining goodness-of-fit graphs to obtain a range of parameter values that maximize the posterior likelihood in MIMAR’s output.

**RESULTS**

*Polymorphism and genetic diversity*

Once sequences were trimmed to include only sequences of high quality, there were a total of 6,155bp of aligned coding sequence for the 13 loci, with an average length of 473.5 bp per locus. In total, there were 78 segregating sites, with no fixed differences between the sex chromosomal races and approximately 1.5 times as many shared as unique polymorphisms. There was also a slightly higher number of unique polymorphisms in the TX SCR compared to the NC SCR (Table 3.1).

Table 3.1 – Unique and shared DNA polymorphisms between the Texas and North Carolina sex chromosome races of *Rumex hastatulus*, based on an analysis of 13 nuclear loci. Note there were no fixed differences between the races. N<sub>TX</sub> and N<sub>NC</sub> give the number of sequences at a given locus for the TX and NC races, respectively.

<table>
<thead>
<tr>
<th>Locus</th>
<th>N&lt;sub&gt;TX&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</th>
<th>N&lt;sub&gt;NC&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</th>
<th>UP&lt;sub&gt;TX&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>UP&lt;sub&gt;NC&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Shared polymorphisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>40</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>42</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>44</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>46</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
<td>38</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td>36</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatus*

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>30</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>48</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>46</td>
<td>42</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>48</td>
<td>40</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>46</td>
<td>42</td>
<td>8</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>46</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>48</td>
<td>23</td>
<td>20</td>
<td>35</td>
</tr>
</tbody>
</table>

\[ \text{a the sample size for each locus and sex chromosome race.} \]

\[ \text{b the number of unique polymorphisms for each locus and sex chromosome race} \]

Tajima’s \( D \) at synonymous sites was lower for the NC race, suggesting recent population expansion, and there was a higher average number of pairwise differences (\( \pi \)) at synonymous sites for the TX race (Figure 3.2). Watterson’s \( \theta \) and Tajima’s \( \pi \) did not differ significantly between the sex chromosome races (for a full overview of the summary statistics see Appendix S3.).
Figure 3.2 – Average population genetic parameters (Watterson’s $\theta$, Tajima’s $\pi$, and Tajima’s $D$) for: (A) synonymous sites and (B) non-synonymous sites at 13 nuclear loci from 24 populations of *Rumex hastatulus*. Species-wide (pooled) and sex chromosome race-specific values are shown for each parameter. Error bars are ± the standard error.

**Population Genetic Structure**

The output from STRUCTURE revealed that a value of $K=2$ (Figure 3.3) provided the best fit to the data, and evidence of clear differentiation in
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

population assignment with individuals clustering primarily by chromosomal race (Figure 3.4).

![Graph showing STRUCTURE analysis of population genetic structure in *Rumex hastatulus*. ΔK values for each value of K generated using the method of Evanno et al. (2005) as implemented in STRUCTURE Harvester. The peak value of ΔK is inferred as the true value of K for the individuals sampled.](image)
Figure 3.4 – Cluster assignment probabilities in the STRUCTURE analyses for 48 individuals of *Rumex hastatulus* sampled from 24 populations throughout the southern USA. Each line represents a separate individual. Figure is based on the CLUMPP output from 5 replicated runs for each *K*-value. Individuals are arranged by longitude.

Increasing *K*-values up to 10 consistently revealed two major clusters, separated according to sex chromosome race. Both individuals from a population of the NC race were assigned to the cluster consisting of TX individuals, whereas two individuals from different populations of the TX race were assigned to the NC cluster. None of the individuals assigned to the opposite cluster of their SCR showed any obviously distinguishing phenotypic differences.

**Demographic History**

I estimated an ancestral *θ* value of between 0.001 and 0.0025 (Figure 3.5A) and a divergence time between 5000 and 15000 generations (Figure 3.5B).
Chapter 3 – Genetic differentiation between sex chromosome races of Rumex hastatulus

The maximum posterior probabilities were for a $\theta$ of 0.001715 and a generation time of 8568. There was insufficient power in the models to infer contemporary values of $\theta$, or to determine whether individuals from populations closer to the boundary between sex chromosome races were more closely related to one another than to more distant populations.

Figure 3.5 – A) Distribution of posterior probability estimates of the ancestral $\theta$ obtained from MIMAR. B) Distribution of posterior probability estimates of the coalescence time between TX and NC SCRs obtained from MIMAR.

DISCUSSION

Theory predicts, and empirical evidence confirms, that large populations of outcrossing plants, especially those that are wind-pollinated, should have low levels of genetic differentiation among populations and high levels of within-population variation (Brown 1979, Hamrick et al. 1979). The vast majority of studies that have confirmed this prediction have investigated patterns of genetic diversity in native populations of woody, long-lived tree species occupying late successional stages, and in which large effective population sizes and extensive pollen-mediated gene flow serves to limit opportunities for genetic subdivision. In contrast, studies of annual plants of open,
disturbed, early successional environments, many of which are introduced weedy species, have most commonly reported considerable genetic differentiation among populations and low levels of genetic diversity within populations. This pattern of population genetic structure has generally been explained as resulting from interactions between the largely self-fertilizing mating system of most annual species and stochastic forces associated with frequent colonizing episodes and population bottlenecks (Barrett and Husband 1990). *Rumex hastatulus* combines strikingly different elements of these two contrasting groups of plant species. On the one hand, the species is wind-pollinated and because of its dioecious sexual system it is an obligate outcrosser; on the other hand, it is largely annual, occupies open, disturbed environments and is native to the region where populations were investigated. Because of this unusual combination of features it was not obvious at the outset of this study which patterns of population genetic structures described above would be revealed, and the extent to which the two sex chromosome races would be genetically differentiated.

My investigations of population genetic structure, based on DNA sequence polymorphism at 13 nuclear loci, revealed high levels of diversity within all populations, but limited genetic subdivision within each of the sex chromosome races. However, based on the STUCTURE analysis there was evidence of clear genetic differentiation between the sex chromosome races with $K=2$ providing the best fit to the molecular data. The inability to detect fine-scale genetic structure among populations within the sex chromosome races made an examination of isolation-by-distance uninteresting. It is possible that my failure to detect finer scale genetic structure resulted from the limited sample of nuclear loci in this study. Future investigations using next generation sequencing approaches may be better able to resolve such differences by incorporating orders of magnitude more SNPs (e.g. Hohenlohe et al. 2011). However, there is no reason to believe that undetected
differentiation within the sex chromosome races would affect the strong support for genetic differentiation between them that my study revealed.

Several studies reporting genetic differentiation among geographically separated plant populations in the southern USA have inferred that this has arisen following radiation from two distinct glacial refugia on either side of the Mississippi river at the end of the last ice age (Steyermark 1934, Delcourt et al. 1980, Watson et al. 2002). However, more recent analyses of several taxa have revealed more complex types of genetic structure (e.g. Soltis et al. 2006), indicating that patterns of genetic differentiation following glaciation are likely to be a lot more elaborate than a simple binary division centred on the Mississippi River. Moreover, not all species from this region display geographical structure and, indeed, a recent study (Victory et al. 2006) of the wind-pollinated and wind-dispersed tree *Juglans nigra* (Juglandaceae) detected very little evidence for population genetic structure, even when highly polymorphic gene regions were analyzed.

Several hypotheses could explain the particular binary structure I observed between the SCRs of *R. hastatus*. Genetic differentiation could be the result of historical processes associated with glaciation and/or the Mississippi may have served as a geographical barrier to admixture between the sex chromosome races. Alternatively the strong subdivision I detected could be the result of intrinsic barriers to gene flow as a result of sex chromosome differentiation. Although a narrow putative hybrid zone is reported between the SCRs of *R. hastatus* (see Figure 1 in Smith 1969), the existence of this zone needs to be investigated thoroughly to determine if it currently exists and what relevance it might have in explaining the patterns of subdivision I observed.

As might be expected with recent divergence and extensive pollen-mediated gene flow, I found no fixed differences between the SCRs, but there were slightly more unique polymorphisms in the TX race (Table 3-1). On average, the TX SCR has much larger contemporary population sizes than
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

the NC race (median 73732 and 8738 for TX and NC, respectively, see Pickup and Barrett 2013 Table 1), which may explain the slightly higher observed values for unique polymorphisms (Tajima 1989b). Additionally, Tajima’s $D$ was negative for the NC race, suggesting that it may have experienced recent population expansion, purifying selection, or both (Tajima 1989a). Population expansion may have been favoured in the NC race by the clearance of forest in the southeast USA, and the spread of disturbed habitats associated with agricultural expansion during historic times. The TX race, by contrast had a mean $D$ close to zero, indicative of neutrality at the loci included in this study. When the races were analyzed separately, the lack of structure in the TX race coupled with the strong clustering in the NC race and higher value of Tajima’s $D$ for the TX race is consistent with cytogenetic evidence that the NC race is derived from the TX race.

The coalescent simulation results from MIMAR suggest a recent split between the SCRs, occurring within the last 15,000 years, assuming a generation time of 1 year. This is a more recent estimate than the 600,000 years reported by Del Bosque et al. (2006) using mean distance corrected estimates from the intergenic transcribed spacer from Navajas-Pérez et al (2005). The aforementioned analysis comprised only 12 individuals from two populations and may have overestimated the extent of divergence between SCRs due to an underrepresentation of polymorphisms. The MIMAR results indicated extensive gene flow between the SCRs, but the simulations were unable to provide accurate estimates of parameter values for migration, likely due to the limited extent of differentiation between the SCRs. I was also unable to investigate whether populations in closer proximity to the zone of contact between SCRs were more closely related to one another than were populations farther away. This was due due to the absence of fixed differences between SCRs, and an insufficient number of unique polymorphisms in each race, as would be expected if the races recently diverged.
Chapter 3 – Genetic differentiation between sex chromosome races of *Rumex hastatulus*

In conclusion, a more refined understanding of the role of chromosomal differentiation in lineage divergence requires studies of a variety of organisms with diverse reproductive and life-history traits across a range of geographical contexts. Although my studies did not examine hybridization between the SCRs of *R. hastatulus*, or the nature of any post-zygotic isolation that may potentially occur, they do suggest that the chromosomal differences between the races reduce gene flow between the races. Although Smith (1969) has established that the SCRs can be crossed under glasshouse conditions, the fertility and viability of hybrids under field conditions, as well as the rate of hybridization remain unknown. The strong subdivision co-incident with the geographical distribution of the two SCRs revealed by my STRUCTURE analyses would be unexpected in a wind-pollinated plant with extensive gene flow. *Rumex hastatulus* is well suited for further investigations of the patterns and processes responsible for lineage divergence owing to chromosomal rearrangements. Future work aimed at determining the role of intrinsic genetic differences in restricting gene flow between the sex chromosome races would be worthwhile, and a thorough analysis of the putative hybrid zone would provide additional insights into the mechanisms governing the observed geographical subdivision.
CHAPTER 4 – GENERAL CONCLUSIONS

This thesis investigated patterns of phenotypic and molecular variation across a large portion of the range of the dioecious, wind-pollinated herb *Rumex hastatulus*. In Chapter 2, I examined the extent of differentiation for ecologically-relevant morphological and life-history traits between the sex chromosome races of *R hastatulus* and, within the races, the extent of sexual dimorphism for the same traits. I also investigated whether there were clines in trait values associated with either latitude or longitude. In chapter 3, I analysed DNA sequence polymorphism to evaluate contemporary patterns of genetic diversity and clustering across the range of *R. hastatulus* and used coalescent simulations to investigate the species’ demographic history and possible origins of the two principal sex chromosome races. Despite finding strong evidence of genetic differentiation between the SCRs, I found comparatively little phenotypic differentiation between SCRs. There was, however, significant sexual dimorphism for morphological traits and in flowering time. Below, I describe the major findings of each investigation and suggest possible directions for future research.

**Major findings of the thesis**

The central aim of this thesis was to investigate patterns of phenotypic and molecular genetic differentiation across the range of *Rumex hastatulus*. I detected significant differentiation between SCRs for flowering time, leaf width, and number of inflorescences, when I accounted for covariation with gender (Table 2.1, Figure 2.3). However, there were no qualitatively identifiable traits that consistently differentiated the SCRs, and the majority of the quantitative variation for phenotypic traits occurred between populations rather than between SCRs (Figure 2.4). Despite occupying an extensive area encompassing a wide range of climatic conditions across the southern USA, the relatively low degree of phenotypic differentiation
detected between SCRs made it impossible to reliably differentiate between them. I estimated a split occurring between the SCR lineages between 5000-15000 generations before present (Figure 3.5), significantly earlier than the 600,000-year estimate of Del Bosque et al. (2011). A possible cause of the disparity in estimates is that I used randomly chosen nuclear coding sequences rather than the ITS sequences used by Del Bosque et al. (2011) and that the two studies differed widely in sampling intensity. The relatively limited differentiation between SCRs is likely caused, at least partially, by gene flow between SCRs, owing to the wind dispersed pollen and seeds of *R. hastatulus*. Nevertheless, despite the absence of fixed differences I was able to detect evidence of two distinct genetic clusters, corresponding primarily to the separate SCRs (Figure 3.4). So while some gene flow occurs between SCRs, they still maintain significant genetic differentiation from one another.

In addition to phenotypic differences between SCRs, there was also evidence for differentiation between males and females for the majority of traits examined. Pickup and Barrett (2013) investigate patterns of sexual dimorphism for height throughout the lifespan of *R. hastatulus* in greater detail and provide an adaptive explanation for the patterns they detected. In addition to height dimorphism I found differences between genders for flowering time, length of the largest leaf, number of inflorescences, diameter of the largest inflorescence, and the biomass of vegetative tissues (Table 2.1). The occurrence of pervasive sexual dimorphism in *R. hastaulus* highlights the importance of using sex as a covariate in analyses of phenotypic differentiation between populations.

*Future work on Rumex hastatulus*

Since Smith’s (1963, 1964, 1969) seminal work on *R. hastatulus* there have been limited investigations of the nature and extent of the hybrid zone between SCRs. The more recent divergence time estimates in *R. hastatulus*
compared to other $XY_1Y_2$ systems (Negrutiu et al. 2001), makes it a strong candidate for examining the early stages of Y-chromosome evolution and lineage differentiation. Despite successful laboratory crosses between the SCR\(s\) of \(R.\) \textit{hastatulus}\ and anecdotal evidence of introgression, the current existence, geographical extent, and nature of the hybrid zone remains largely unknown. Agricultural and urban development throughout the range of \(R.\) \textit{hastatulus}\ populations may have led to extirpation of many of the populations that Smith investigated during the 1960s. Alternatively, because \(R.\) \textit{hastatulus}\ is a weedy annual of open, disturbed sites, it is also possible that anthropogenic influences associated with disturbance could have led to conditions fostering hybridization. Indeed, many examples of hybridization and introgression in the flowering plants are associated with altered ecological conditions and human disturbance (e.g. Lamont et al. 2003; Bleeker and Hurka 2008). Smith (1969) reported the existence of a third SCR that he called the Illinois-Missouri race, which may have arisen following hybridization between the TX and NC SCR\(s\) (Bartkowiak 1971). However, the second Y chromosomes of the Illinois-Missouri and North Carolina races are morphologically distinct, and do not correspond to the same autosomes in the Texas race, suggesting a possible second origin of the \(Y_2\) chromosome in \(R.\) \textit{hastatulus}. Inclusion of the Illinois-Missouri race in future studies of the genetic relationships and demographic histories of the SCR\(s\) of \(R.\) \textit{hastaulus}\ may provide insight into whether a second origin is likely for the \(Y_2\) chromosome.

In animals, hybrid sterility and inviability are more likely to occur in the heterogametic sex of F\(_1\) hybrid crosses, a phenomenon known as Haldane’s Rule (Haldane 1922; Orr 1997). With the recent detection of Haldane’s rule operating in some crosses between members of the plant genus \textit{Silene} (Brothers and Delph 2010), the extent to which Haldane’s rule occurs and the mechanisms underpinning unequal hybrid sterility remain to
be determined for plants. Haldane’s rule has important implications for population genetic structure, sex ratios, and individual fitness in hybrid zones. If there is differing sterility or inviability between the SCRs, it may suggest that gene flow is asymmetric in hybrid zones. Additionally, reduced hybrid male fitness may have relatively little effect on maternally inherited genes, allowing them to pass through the hybrid zone more easily (Martinsen et al. 2001). *Rumex hastatulus* may be a good candidate for future studies of Haldane’s rule because of the occurrence of distinct sex chromosome races and its amenability to crossing and studies of sex ratios in the field and glasshouse.


Bates D, Maechler M, Bolker B. 2012. lme4: linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4


Literature Cited


Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing


R Core Team. 2013. R: A language and environment for statistical computing.


Stebbins GL. 1971. Relationships between adaptive radiation, speciation and major evolutionary trends. *Taxon* 20: 3-16.


Turesson G. 1922. The species and the variety as ecological units. *Hereditas* 3: 100-113.


Literature Cited


Table S1 – A list of all populations (Pops) investigated in this thesis and their localities, latitude and longitude coordinates (Lat and Long, respectively), approximate elevation above sea level (Elev, in metres), and an estimate of their population size (from Pickup and Barrett, 2013). The SCR column indicates whether the population is of the Texas (TX) or North Carolina (NC) sex chromosome race.

<table>
<thead>
<tr>
<th>Pops</th>
<th>SCR</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elev</th>
<th>Pop size</th>
</tr>
</thead>
<tbody>
<tr>
<td>al-bre</td>
<td>NC</td>
<td>Brewton, AL</td>
<td>31°4'58&quot;</td>
<td>86°59'56&quot;</td>
<td>31</td>
<td>77403</td>
</tr>
<tr>
<td>al-bru</td>
<td>NC</td>
<td>Brundidge, AL</td>
<td>31°43'60&quot;</td>
<td>85°50'25&quot;</td>
<td>137</td>
<td>28800</td>
</tr>
<tr>
<td>fl-jas</td>
<td>NC</td>
<td>Jasper, FL</td>
<td>30°34'3&quot;</td>
<td>83°4'31&quot;</td>
<td>34</td>
<td>761</td>
</tr>
<tr>
<td>fl-mar</td>
<td>NC</td>
<td>Mariana, FL</td>
<td>30°48'43&quot;</td>
<td>85°11'28&quot;</td>
<td>39</td>
<td>348876</td>
</tr>
<tr>
<td>ga-bel</td>
<td>NC</td>
<td>Belfast, GA</td>
<td>31°50'35&quot;</td>
<td>81°17'3&quot;</td>
<td>15</td>
<td>933</td>
</tr>
<tr>
<td>ga-ell</td>
<td>NC</td>
<td>Ellaville, GA</td>
<td>32°15'6&quot;</td>
<td>84°16'47&quot;</td>
<td>168</td>
<td>2757</td>
</tr>
<tr>
<td>ga-gla</td>
<td>NC</td>
<td>Gladys, GA</td>
<td>31°28'55&quot;</td>
<td>83°14'16&quot;</td>
<td>97</td>
<td>131333</td>
</tr>
<tr>
<td>ga-sta</td>
<td>NC</td>
<td>Statesboro, GA</td>
<td>32°27'9&quot;</td>
<td>81°50'55&quot;</td>
<td>78</td>
<td>13789</td>
</tr>
<tr>
<td>nc-bat</td>
<td>NC</td>
<td>Bath, NC</td>
<td>35°31'32&quot;</td>
<td>76°52'18&quot;</td>
<td>6</td>
<td>41173</td>
</tr>
<tr>
<td>sc-bra</td>
<td>NC</td>
<td>Branchville, SC</td>
<td>33°15'3&quot;</td>
<td>80°48'27&quot;</td>
<td>34</td>
<td>8738</td>
</tr>
<tr>
<td>sc-mar</td>
<td>NC</td>
<td>Marion, SC</td>
<td>34°10'58&quot;</td>
<td>79°29'13&quot;</td>
<td>31</td>
<td>348876</td>
</tr>
<tr>
<td>sc-pro</td>
<td>NC</td>
<td>Prosperity, SC</td>
<td>34°6'29&quot;</td>
<td>81°26'14&quot;</td>
<td>126</td>
<td>827</td>
</tr>
<tr>
<td>la-ben</td>
<td>TX</td>
<td>Benson, LA</td>
<td>31°51'1&quot;</td>
<td>93°41'18&quot;</td>
<td>62</td>
<td>6027</td>
</tr>
<tr>
<td>la-man</td>
<td>TX</td>
<td>Many, LA</td>
<td>31°27'41&quot;</td>
<td>93°39'39&quot;</td>
<td>83</td>
<td>7466</td>
</tr>
<tr>
<td>ok-bac</td>
<td>TX</td>
<td>Bache, OK</td>
<td>34°53'36&quot;</td>
<td>96°45'44&quot;</td>
<td>227</td>
<td>756014</td>
</tr>
<tr>
<td>ok-rat</td>
<td>TX</td>
<td>Rattan, OK</td>
<td>34°9'27&quot;</td>
<td>95°24'48&quot;</td>
<td>138</td>
<td>89505</td>
</tr>
<tr>
<td>tx-ath</td>
<td>TX</td>
<td>Athens, TX</td>
<td>32°11'5&quot;</td>
<td>95°48'12&quot;</td>
<td>145</td>
<td>1149421</td>
</tr>
<tr>
<td>tx-buc</td>
<td>TX</td>
<td>Buckhorn, TX</td>
<td>30°45'41&quot;</td>
<td>93°40'33&quot;</td>
<td>29</td>
<td>1448221</td>
</tr>
<tr>
<td>tx-buf</td>
<td>TX</td>
<td>Buffalo, TX</td>
<td>31°26'10&quot;</td>
<td>96°2'0&quot;</td>
<td>94</td>
<td>78109</td>
</tr>
<tr>
<td>tx-gid</td>
<td>TX</td>
<td>Giddings, TX</td>
<td>30°8'19&quot;</td>
<td>96°55'9&quot;</td>
<td>141</td>
<td>32800</td>
</tr>
<tr>
<td>tx-gro</td>
<td>TX</td>
<td>Groesbeck, TX</td>
<td>31°30'49&quot;</td>
<td>96°28'50&quot;</td>
<td>134</td>
<td>75480</td>
</tr>
<tr>
<td>Code</td>
<td>State</td>
<td>Location</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Distance</td>
<td>Area</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------------------</td>
<td>------------</td>
<td>------------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>tx-liv</td>
<td>TX</td>
<td>Livingston, TX</td>
<td>30°41'58&quot;</td>
<td>94°47'59&quot;</td>
<td>83</td>
<td>73732</td>
</tr>
<tr>
<td>tx-mte</td>
<td>TX</td>
<td>Mt. Enterprise, TX</td>
<td>31°56'19&quot;</td>
<td>94°42'25&quot;</td>
<td>127</td>
<td>1188</td>
</tr>
<tr>
<td>tx-oak</td>
<td>TX</td>
<td>Oakwood, TX</td>
<td>31°33'39&quot;</td>
<td>95°54'6&quot;</td>
<td>107</td>
<td>1090534</td>
</tr>
<tr>
<td>tx-ros</td>
<td>TX</td>
<td>Rosebud, TX</td>
<td>31°7'3&quot;</td>
<td>96°51'37&quot;</td>
<td>130</td>
<td>1978464</td>
</tr>
<tr>
<td>tx-wes</td>
<td>TX</td>
<td>Wesley Chapel, TX</td>
<td>31°21'18&quot;</td>
<td>95°31'25&quot;</td>
<td>105</td>
<td>1027985</td>
</tr>
</tbody>
</table>
### Supplementary Tables

Table S2 – Means and standard errors for all phenotypic traits for all *Rumex hastatulus* populations measured in the glasshouse experiment.

<table>
<thead>
<tr>
<th>Pop</th>
<th>Race</th>
<th>FT</th>
<th>Len</th>
<th>Wid ±</th>
<th>Num ±</th>
<th>H (µg) ±</th>
<th>Diam ±</th>
<th>Veg ±</th>
<th>Rep ±</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL-BRE</td>
<td>NC</td>
<td>48±6.1</td>
<td>154.7±6.1</td>
<td>29.4±1.3</td>
<td>2.0±0.3</td>
<td>565.6±43.2</td>
<td>2.9±0.2</td>
<td>1.8±0.3</td>
<td>2.1±0.2</td>
</tr>
<tr>
<td>AL-BRU</td>
<td>NC</td>
<td>50.7±7.1</td>
<td>141.0±7.1</td>
<td>30.6±2.1</td>
<td>2.2±0.3</td>
<td>553.7±46.2</td>
<td>2.7±0.2</td>
<td>1.1±0.2</td>
<td>1.5±0.2</td>
</tr>
<tr>
<td>FL-JAS</td>
<td>NC</td>
<td>43.6±4.9</td>
<td>151.3±4.9</td>
<td>35.4±1.6</td>
<td>4.1±0.3</td>
<td>691.4±24.3</td>
<td>3.7±0.1</td>
<td>1.1±0.2</td>
<td>2.2±0.1</td>
</tr>
<tr>
<td>FL-MAR</td>
<td>NC</td>
<td>46.9±6.8</td>
<td>149.9±6.8</td>
<td>36.2±1.8</td>
<td>1.8±0.2</td>
<td>591.9±40.5</td>
<td>2.8±0.1</td>
<td>1.2±0.2</td>
<td>1.5±0.2</td>
</tr>
<tr>
<td>GA-BEL</td>
<td>NC</td>
<td>46.4±7.9</td>
<td>126.5±7.9</td>
<td>33.2±2.1</td>
<td>4.2±0.4</td>
<td>608.0±35.7</td>
<td>2.9±0.2</td>
<td>0.6±0.1</td>
<td>1.9±0.2</td>
</tr>
<tr>
<td>GA-ELL</td>
<td>NC</td>
<td>56.4±8.5</td>
<td>160.5±8.5</td>
<td>29.7±1.4</td>
<td>1.5±0.4</td>
<td>499.6±48.4</td>
<td>2.5±0.2</td>
<td>1.7±0.3</td>
<td>1.8±0.2</td>
</tr>
<tr>
<td>GA-GLA</td>
<td>NC</td>
<td>45.8±10.0</td>
<td>135.4±10.0</td>
<td>36.1±3.2</td>
<td>3.1±0.4</td>
<td>673.6±42.4</td>
<td>3.4±0.2</td>
<td>0.9±0.4</td>
<td>1.8±0.2</td>
</tr>
<tr>
<td>GA-STA</td>
<td>NC</td>
<td>44.0±4.8</td>
<td>149.4±4.8</td>
<td>28.6±1.2</td>
<td>3.3±0.3</td>
<td>667.4±29.3</td>
<td>3.1±0.1</td>
<td>1.5±0.2</td>
<td>2.0±0.0</td>
</tr>
<tr>
<td>NC-BAT</td>
<td>NC</td>
<td>50.5±6.4</td>
<td>176.6±6.4</td>
<td>37.1±1.7</td>
<td>0.5±0.2</td>
<td>473.7±41.0</td>
<td>2.5±0.2</td>
<td>3.1±0.3</td>
<td>1.4±0.2</td>
</tr>
<tr>
<td>SC-BRA</td>
<td>NC</td>
<td>48.9±7.6</td>
<td>154.6±7.6</td>
<td>32.3±2.3</td>
<td>2.4±0.4</td>
<td>668.8±48.9</td>
<td>3.2±0.2</td>
<td>1.5±0.3</td>
<td>2.0±0.2</td>
</tr>
<tr>
<td>SC-MAR</td>
<td>NC</td>
<td>47.1±5.9</td>
<td>186.9±5.9</td>
<td>40.2±1.5</td>
<td>0.9±0.2</td>
<td>449.3±46.7</td>
<td>2.7±0.2</td>
<td>3.4±0.3</td>
<td>2.3±0.2</td>
</tr>
<tr>
<td>SC-PRO</td>
<td>NC</td>
<td>44.9±6.4</td>
<td>152.8±6.4</td>
<td>36.6±1.2</td>
<td>2.5±0.3</td>
<td>648.4±35.8</td>
<td>3.3±0.1</td>
<td>1.7±0.3</td>
<td>2.2±0.1</td>
</tr>
<tr>
<td>LA-BEN</td>
<td>TX</td>
<td>39.0±4.3</td>
<td>122.4±4.3</td>
<td>42.5±1.9</td>
<td>5.2±0.2</td>
<td>737.7±18.3</td>
<td>3.3±0.1</td>
<td>0.6±0.1</td>
<td>2.1±0.1</td>
</tr>
<tr>
<td>LA-MAN</td>
<td>TX</td>
<td>38.3±3.3</td>
<td>103.9±3.3</td>
<td>31.8±2.4</td>
<td>5.3±0.2</td>
<td>649.0±19.3</td>
<td>2.9±0.1</td>
<td>0.3±0.0</td>
<td>1.6±0.1</td>
</tr>
<tr>
<td>OK-BAC</td>
<td>TX</td>
<td>48.3±3.9</td>
<td>159.3±3.9</td>
<td>46.1±2.2</td>
<td>3.8±0.2</td>
<td>799.9±23.7</td>
<td>4.5±0.1</td>
<td>1.2±0.1</td>
<td>2.8±0.2</td>
</tr>
<tr>
<td>OK-RAT</td>
<td>TX</td>
<td>46.8±5.9</td>
<td>144.3±5.9</td>
<td>48.1±2.4</td>
<td>4.6±0.3</td>
<td>788.9±25.4</td>
<td>4.2±0.2</td>
<td>0.9±0.1</td>
<td>2.6±0.2</td>
</tr>
<tr>
<td>TX-ATH</td>
<td>TX</td>
<td>44.6±4.3</td>
<td>129.4±4.3</td>
<td>43.7±2.2</td>
<td>3.9±0.2</td>
<td>738.7±22.1</td>
<td>4.0±0.1</td>
<td>0.6±0.1</td>
<td>2.0±0.1</td>
</tr>
<tr>
<td>TX-BUC</td>
<td>TX</td>
<td>40.6±4.5</td>
<td>130.0±4.5</td>
<td>46.2±2.3</td>
<td>4.8±0.2</td>
<td>708.6±21.6</td>
<td>3.6±0.1</td>
<td>0.6±0.0</td>
<td>2.0±0.1</td>
</tr>
<tr>
<td>TX-BUF</td>
<td>TX</td>
<td>40.2±4.6</td>
<td>117.6±4.6</td>
<td>43.0±1.9</td>
<td>4.8±0.2</td>
<td>677.5±18.3</td>
<td>3.7±0.1</td>
<td>0.6±0.1</td>
<td>2.1±0.1</td>
</tr>
<tr>
<td>TX-GID</td>
<td>TX</td>
<td>50.4±8.1</td>
<td>112.5±8.1</td>
<td>38.4±3.9</td>
<td>1.6±0.2</td>
<td>667.0±29.6</td>
<td>2.6±0.3</td>
<td>0.2±0.1</td>
<td>0.8±0.1</td>
</tr>
<tr>
<td>TX-GRO</td>
<td>TX</td>
<td>39.6±4.4</td>
<td>113.5±4.4</td>
<td>41.2±2.0</td>
<td>4.8±0.3</td>
<td>662.4±17.3</td>
<td>3.3±0.1</td>
<td>0.5±0.1</td>
<td>1.8±0.1</td>
</tr>
<tr>
<td>TX-LIV</td>
<td>TX</td>
<td>52.3±5.8</td>
<td>154.9±5.8</td>
<td>48.7±2.9</td>
<td>3.3±0.2</td>
<td>735.5±23.7</td>
<td>3.6±0.2</td>
<td>0.7±0.1</td>
<td>1.9±0.1</td>
</tr>
<tr>
<td>TX-MTE</td>
<td>TX</td>
<td>45.3±4.6</td>
<td>136.6±4.6</td>
<td>48.6±2.3</td>
<td>4.4±0.3</td>
<td>737.3±19.3</td>
<td>3.7±0.1</td>
<td>0.6±0.0</td>
<td>2±0.2</td>
</tr>
<tr>
<td>TX-MTP</td>
<td>TX</td>
<td>46.4±6.1</td>
<td>149.5±6.1</td>
<td>49.8±2.2</td>
<td>3.7±0.3</td>
<td>794.3±23.7</td>
<td>4.3±0.2</td>
<td>0.9±0.1</td>
<td>2.3±0.1</td>
</tr>
</tbody>
</table>
## Supplementary Tables

<table>
<thead>
<tr>
<th>TX</th>
<th>TX</th>
<th>TX</th>
<th>TX</th>
<th>TX</th>
<th>TX</th>
<th>TX</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAK</td>
<td>TX</td>
<td>45.1±6.0</td>
<td>144.6±6.0</td>
<td>47.4±3.0</td>
<td>5.4±0.3</td>
<td>731.0±19.4</td>
</tr>
<tr>
<td>TX-ROS</td>
<td>TX</td>
<td>46.5±5.5</td>
<td>119.7±5.5</td>
<td>39.2±2.8</td>
<td>3.8±0.4</td>
<td>687.2±23.8</td>
</tr>
<tr>
<td>TX-WES</td>
<td>TX</td>
<td>46.1±9.2</td>
<td>146.0±9.2</td>
<td>50.0±4.0</td>
<td>3.5±0.4</td>
<td>761.4±32.0</td>
</tr>
</tbody>
</table>

*Pop* = population, *FT* = flowering time (days), *Len* = length of the largest leaf (mm), *Wid* = width of the largest leaf (mm), *Num* = number of inflorescences, *H* = height of the tallest inflorescence (cm), *Diam* = diameter of the tallest inflorescence (mm), *Veg* = biomass of vegetative tissue (grams), *Rep* = biomass of reproductive tissues (grams).
Supplementary Tables

Table S3 – Eigenvectors from a PCA analysis with scaled means and variances of 928 individuals of *Rumex hastatulus* for seven phenotypic variables (time to first flower (FT), length and width of the largest leaf, number of inflorescences (Number), height of the tallest inflorescence (Height), diameter of the principal inflorescence’s stem (Diameter), and vegetative biomass).

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering Time</td>
<td>-0.29792</td>
<td>0.555651</td>
<td>-0.05941</td>
<td>0.56758</td>
<td>-0.02315</td>
<td>0.277977</td>
<td>-0.44611</td>
</tr>
<tr>
<td>Length</td>
<td>-0.48548</td>
<td>0.039404</td>
<td>0.010896</td>
<td>0.157543</td>
<td>0.370396</td>
<td>0.275154</td>
<td>0.724507</td>
</tr>
<tr>
<td>Width</td>
<td>-0.24893</td>
<td>-0.44629</td>
<td>0.742725</td>
<td>0.367903</td>
<td>-0.11735</td>
<td>-0.15887</td>
<td>-0.11337</td>
</tr>
<tr>
<td>Number</td>
<td>0.205298</td>
<td>-0.6215</td>
<td>-0.45552</td>
<td>0.448037</td>
<td>0.297951</td>
<td>0.2238</td>
<td>-0.15652</td>
</tr>
<tr>
<td>Height</td>
<td>-0.44548</td>
<td>-0.08819</td>
<td>0.002359</td>
<td>-0.41967</td>
<td>0.617757</td>
<td>-0.13926</td>
<td>-0.46543</td>
</tr>
<tr>
<td>Diameter</td>
<td>-0.42259</td>
<td>-0.30033</td>
<td>-0.14769</td>
<td>-0.328</td>
<td>-0.54034</td>
<td>0.542845</td>
<td>-0.12321</td>
</tr>
<tr>
<td>Biomass</td>
<td>-0.44091</td>
<td>-0.07929</td>
<td>-0.46411</td>
<td>0.182317</td>
<td>-0.29349</td>
<td>-0.67648</td>
<td>0.083154</td>
</tr>
</tbody>
</table>
Supplementary Tables

Table S4 – Population genetic data from individual populations based on the survey of 13 nuclear loci in 28 populations of *Rumex hastatulus* information. I used a modified version of Polymorphurama (Bachtrog and Andolfatto 2006) to calculate diversity statistics, including $\theta_w$ (Watterson 1975), $\pi$ (Tajima 1983), and Tajima’s $D$ (Tajima 1989), for synonymous and non-synonymous sites for the 13 loci.

<table>
<thead>
<tr>
<th>Locus</th>
<th>$\theta_w$</th>
<th>$\pi$</th>
<th>$D$</th>
<th>$\theta_w$</th>
<th>$\pi$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0118</td>
<td>0.0195</td>
<td>1.4588</td>
<td>0.0041</td>
<td>0.0006</td>
<td>-1.9964</td>
</tr>
<tr>
<td>2</td>
<td>0.0109</td>
<td>0.0096</td>
<td>-0.2915</td>
<td>0.0009</td>
<td>0.0010</td>
<td>0.1337</td>
</tr>
<tr>
<td>3</td>
<td>0.0131</td>
<td>0.0045</td>
<td>-1.5664</td>
<td>0.0012</td>
<td>0.0001</td>
<td>-1.3982</td>
</tr>
<tr>
<td>4</td>
<td>0.0172</td>
<td>0.0089</td>
<td>-1.1527</td>
<td>0.0070</td>
<td>0.0010</td>
<td>-2.2985</td>
</tr>
<tr>
<td>5</td>
<td>0.0051</td>
<td>0.0008</td>
<td>-1.5625</td>
<td>0.0023</td>
<td>0.0009</td>
<td>-1.2823</td>
</tr>
<tr>
<td>6</td>
<td>0.0042</td>
<td>0.0039</td>
<td>-0.1369</td>
<td>0.0019</td>
<td>0.0008</td>
<td>-1.2528</td>
</tr>
<tr>
<td>7</td>
<td>0.0150</td>
<td>0.0149</td>
<td>-0.0067</td>
<td>0.0040</td>
<td>0.0031</td>
<td>-0.5630</td>
</tr>
<tr>
<td>8</td>
<td>0.0046</td>
<td>0.0017</td>
<td>-1.1484</td>
<td>0.0000</td>
<td>0.0000</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>0.0033</td>
<td>0.0049</td>
<td>0.7619</td>
<td>0.0023</td>
<td>0.0010</td>
<td>-1.2204</td>
</tr>
<tr>
<td>10</td>
<td>0.0118</td>
<td>0.0117</td>
<td>-0.0231</td>
<td>0.0000</td>
<td>0.0000</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>0.0103</td>
<td>0.0049</td>
<td>-1.1428</td>
<td>0.0075</td>
<td>0.0028</td>
<td>-1.6966</td>
</tr>
<tr>
<td>12</td>
<td>0.0254</td>
<td>0.0211</td>
<td>-0.4692</td>
<td>0.0032</td>
<td>0.0045</td>
<td>0.8494</td>
</tr>
<tr>
<td>13</td>
<td>0.0130</td>
<td>0.0119</td>
<td>-0.1806</td>
<td>0.0019</td>
<td>0.0016</td>
<td>-0.2509</td>
</tr>
</tbody>
</table>
Supplementary Tables

Supplemental Data 5 – R code used for data analysis

#Change.libPaths() from containing:
"/Users/andrewsimpson/Library/R/2.15/library" and
"/Library/Frameworks/R.framework/Resources/library" to just
"/Library/Frameworks/R.framework/Resources/library"

.libPaths("/Library/Frameworks/R.framework/Resources/library")
update.packages()

#Loading some commonly used packages
require(vegan); require(car); require(lattice); require(lme4);
require(languageR); require(extrafont); loadfonts()

#Less commonly used packages
require(spdep); require(qtl); require(abc); require(gridBase);

#Citations for all packages used

################################################################
################################################################
### Importing and transforming the data
###

import1 = read.csv(file.choose(), header=T); import1$pop =

70
"ok-bac", "ok-rat", "tx-ath", "tx-buc", "tx-buf", "tx-gid", "tx-gro", "tx-liv", "tx-mte", "tx-mtp", "tx-oak", "tx-ros", "tx-wes"), ordered = is.ordered(import1$pop)); fem = which(import1$sex == 0); mal = which(import1$sex == 1); TX = which(import1[,1] == "T"); NC = which(import1[,1] == "N"); mal.tx = which(import1[506:1163,14] == 1); mal.nc = which(import1[1:505,14] == 1); fem.tx = which(import1[506:1163,14] == 0); fem.nc = which(import1[1:505,14] == 0); NR = import1[-c(fem,mal),]; repro = import1[c(fem,mal),]; repro = repro[order(repro[,1], repro[,14]),]; import2 = read.csv(file.choose(), header=T); import2$Long = import2$Long * -1; import2$Pop = factor(import2$Pop, levels = c("al-bre", "al-bru", "fl-jas", "fl-mar", "ga-bel", "ga-ell", "ga-gla", "ga-sta", "nc-bat", "sc-bra", "sc-mar", "sc-pro", "la-ben", "la-man", "ok-bac", "ok-rat", "tx-ath", "tx-buc", "tx-buf", "tx-gid", "tx-gro", "tx-liv", "tx-mte", "tx-mtp", "tx-oak", "tx-ros", "tx-wes"), ordered = is.ordered(import2$Pop)); mydat.scale = repro[,c(1:4,6:13,24,27,28)]; mydat.scale = data.frame(race=mydat.scale[,1], pop=mydat.scale[,2], fam = mydat.scale[,3], ind = mydat.scale[,4], block = mydat.scale[,5], ff.time = scale(mydat.scale[,6], center=TRUE, scale=TRUE), sex = mydat.scale[,7], len.1 = scale(mydat.scale[,8], center=TRUE, scale=TRUE), wid.1 = scale(mydat.scale[,9], center=TRUE, scale=TRUE), len.2 = scale(mydat.scale[,10], center=TRUE, scale=TRUE), wid.2 = scale(mydat.scale[,11], center=TRUE, scale=TRUE), inflo.num = scale(mydat.scale[,12], center=TRUE, scale=TRUE), inflo.H = scale(mydat.scale[,13], center=TRUE, scale=TRUE), stem.diam = scale(mydat.scale[,14], center=TRUE, scale=TRUE), bio.veg = scale(mydat.scale[,15], center=TRUE, scale=TRUE), lat=mydat.scale[,18], long=mydat.scale[,19]); mydat.scale$block = as.factor(mydat.scale$block); mydat.scale$fam = as.factor(mydat.scale$fam); mydat.scale$sex = as.factor(mydat.scale$sex)
Supplementary Tables

# Importing data and calling male, female, Texas, North Carolina, nonreproductive, and reproductive subsets of the data
# NAs in the data are going to mess things up so I will remove them
# I also log-transformed the biomasses so that they behave better.
# The last bit calls the data in a specific data frame, scales the variables to unit mean and variance and also converts mis-coded variables to factors
fix(import1)

### Adding the column necessary for searching for plants in the .csv file
### This has already been done. Keep this code for future reference
# testpaste = paste(import1$pop,import1$fam, sep=" ")
# import1$identifier = paste(testpaste,import1$ind, sep="_")
# write.csv(import1,file="/Users/andrewsimpson/Documents/Grad Work/Data/Grow out data/Grow out data with identifier column")

### Subsetting by Population
###


for (i in 1:27){
    populations[i,3] = nrow(subset(import1, pop == levels(import1$pop)[i]))
    populations[i,4] = (sum(subset(import1, pop ==
levels(import1$pop)[i])$sex, na.rm=TRUE)/mean(subset(import1, pop ==
levels(import1$pop)[i])$sex, na.rm=TRUE))/nrow(subset(import1, pop ==
levels(import1$pop)[i]))
    populations[i,5] = (length(subset(repro, pop ==
levels(import1$pop)[i])$sex)-sum(subset(repro, pop ==
levels(import1$pop)[i])$sex))/length(subset(repro, pop ==
levels(import1$pop)[i])$sex)
    populations[i,8] = sum(subset(import1, pop ==
levels(import1$pop)[i])$ff.time, na.rm= TRUE)/length(which(subset(import1, pop ==
levels(import1$pop)[i])$ff.time >"0")
    populations[i,9] = sd(subset(import1, pop ==
levels(import1$pop)[i])$ff.time, na.rm= TRUE)/sqrt(length(which(subset(import1, pop ==
levels(import1$pop)[i])$ff.time >"0"))
    populations[i,10] = length(subset(repro, pop ==
levels(import1$pop)[i])$sex)
}; populations[22,4] = 1
#Calculates summary values for the proportions of females and reproductive
dividuals per population. The 9th column is the average days to first for
each population. The 10th column is the standard error for each population's
flowering time.

reportd = populations[order(populations$prop.repro),]
repcol = rep(NA, 27)

for (i in 1:27) {
    repcol[i] = paste("gray", round(101 - 100 * populations$prop.repro[i]), sep="")
}

repord$colour = repcol
repord[, c(1, 2, 4, 5, 17, 21)]

xyplot(lat ~ long, data = repord, col = repord$colour, pch = 16, fill.color = repord$colour, las = 1, cex = 2, ylab = "Latitude", xlab = "Longitude", panel = function(...) {
    panel.xyplot(...);
    panel.text(repord$long, repord$lat, labels = repord$pop, cex = 0.75);
})

### Plotting the proportion of females in populations with averages given the weighted means for each race

populations$racenum = c(rep(0, 12), rep(1, 15))

plot(prop.fem ~ racenum, data = populations, las = 1, type = "p", pch = 16, ylab = "Proportion female individuals in population", xlab = "", xlim = c(-0.5, 1.5), xaxt = "n", family = "Century Schoolbook"); axis(1, at = c(0, 1), labels = c("NC", "TX"), family = "Century Schoolbook");
points(0.25,(sum(populations$prop.fem[1:12] * populations$weight[1:12]))/sum(populations$weight[1:12]), pch = 15);
points(0.75,(sum(populations$prop.fem[13:27] * populations$weight[13:27]))/sum(populations$weight[13:27]), pch = 15)
Supplementary Tables

(sum(repro[1:361,14], na.rm=TRUE)/mean(repro[1:361,14], na.rm=TRUE)-
sum(repro[1:361,14], na.rm=TRUE))/(sum(repro[1:361,14], na.rm=TRUE)/mean(repro[1:361,14], na.rm=TRUE));

(sum(repro[362:1019,14], na.rm=TRUE)/mean(repro[362:1019,14], na.rm=TRUE)-
sum(repro[362:1019,14], na.rm=TRUE))/(sum(repro[362:1019,14], na.rm=TRUE)/mean(repro[362:1019,14], na.rm=TRUE))

#Propotion of females for North Carolina is 0.5623.
#Proportion of females for Texas is 0.5136.

Anova(lm(populations$prop.fem~populations$race), type="2")
#The races do not differ significantly in the proportion of females, but this is likely due to a lack of statistical power.

Anova(lm(populations$prop.fem[-9]~populations$race[-9]), type="2")
#When the extreme outlier (nc-bat with 26% reproductive individuals and 33% females) is removed there is a significant difference between the races for the proportion of females.

#Trying to plot the proportion of reproductive individuals by population
barplot(prop.repro ~ pop, data=populations, las=1, family="Century Schoolbook")

########################################################################
################################################################################
### Correlations between latitude and proportions of reproductive and sexual individuals
###
########################################################################
################################################################################
Supplementary Tables

summary(aov(prop.repro~lat, data=subset(populations,race=="NC")))
#Significant correlation between latitude and proportion of reproductive
individuals (p=0.0259, F=6.824)
prmod=lm(prop.repro~lat, data=subset(populations,race=="NC"))
Anova(prmod)
#Making sure I get the same result and calling a function so that I can add
lines of best fit and confidence intervals
summary(prmod)$r.squared
summary(prmod)$coefficients[2,4]

x =
seq(min(subset(populations,race=="NC")$lat),max(subset(populations,race==
"NC")$lat),0.01)
yc = predict(prmod,list(lat=x), interval="confidence")
#Calling the vector for the 95% confidence interval

plot(prop.repro~lat, data=subset(populations,race=="NC"), las=1,
xlab="Latitude", ylab="", pch=16, family="Century Schoolbook", cex.axis=1.5,
cex.lab=1.5); matlines(x,yc,lty=c(1,2,2), lwd=c(2,1,1), col="black");
text(35,0.90, expression(italic(P)* " = 0.026"), family="Century Schoolbook",
cex=1.5); text(35, 0.95, expression(""*italic(R)^2"* = 0.4056"),
family="Century Schoolbook", cex=1.5); mtext("Proportion of reproductive
individuals", side=2, line=3.5, cex=1.5, family="Century Schoolbook")

#Plotting the relationship and adding a line of best fit and the 95%
confidence interval;

psmod=Anova(lm(prop.fem~lat+race, data=populations))
# Looking at correlation between the proportion of females in a population and the latitude. No relationship is detectable, even with the removal of an extreme outlier.

### Boxplots for Reproductive Individuals

### Setting plotting parameters

opar = par
quartz(height=10, width=8)
par(mar=c(2,4,2,2)+0.1); par(mfcol= c(4,2))

plot(prop.fem ~ prop.repro, data=populations, subset = populations$race=="NC", las=1, pch=16, xlab="Proportion of reproductive individuals", ylab="Proportion of Females")

### Important Boxplots

boxplot(prop.fem ~ race, data = populations, names= c("North Carolina", "Texas"), ylab = "Proportion of Females", las =1)

boxplot(prop.fem[-9] ~ race[-9], data = populations, names= c("North Carolina", "Texas"), ylab = "Proportion of Females", las =1)

# Plot these 8 together by race
boxplot(ff.time ~ race, ylim = c(0,80), xlab = "Days to First Flower", data=repro); boxplot(len.2 ~ race, xlab = "Length of the Largest Leaf", data=repro); boxplot(wid.2 ~ race, xlab = "Width of the Largest Leaf", data=repro); boxplot(inflo.num ~ race, xlab = "Number of Inflorescences", data=repro); boxplot(stem.diam ~ race, xlab = "Stem Diameter", data=repro); boxplot(log(as.numeric(bio.rep)) ~ race, xlab = "log-Transformed Reproductive Biomass", data=repro); boxplot(inflo.H ~ race, xlab = "Inflorescence Height", data=repro); boxplot(log(bio.veg) ~ race, xlab = "log-Transformed Vegetative Biomass", data=repro)

#Plot these 8 together by sex
boxplot(ff.time ~ sex, ylim = c(0,80), xlab = "Days to First Flower", data=repro); boxplot(len.2 ~ sex, xlab = "Length of the Largest Leaf", data=repro); boxplot(wid.2 ~ sex, xlab = "Width of the Largest Leaf", data=repro); boxplot(inflo.num ~ sex, xlab = "Number of Inflorescences", data=repro); boxplot(stem.diam ~ sex, xlab = "Stem Diameter", data=repro); boxplot(log(as.numeric(bio.rep)) ~ sex, xlab = "log-Transformed Reproductive Biomass", data=repro); boxplot(inflo.H ~ sex, xlab = "Inflorescence Height", data=repro); boxplot(log(bio.veg) ~ sex, xlab = "Sex", data=repro)
Supplementary Tables

```r	names=c("Female", "Male"), las=1, ylab = "log-Transformed Vegetative Biomass", data=repro)

#Days to first flower by population
boxplot(repro$ff.time ~ repro$pop, ylim = c(28,80), xlab = "Population", ylab = "Days to first flower", axes = FALSE); axis(2, c(0,10,20,30,40,50,60,70,80)); axis(1, at = seq(1,27,by=1), labels = c(levels(repro$pop)), cex.axis=0.70, las=1)
segments(-2,mean(repro$ff.time[1:361], na.rm=TRUE),13,mean(import1$ff.time[1:361], na.rm=TRUE), col="cadetblue4"); segments(14,mean(import1$ff.time[362:1019], na.rm=TRUE),27,mean(import1$ff.time[362:1019], na.rm=TRUE), col="coral3")

#Ordering first by race then by latitude
latord = data.frame(repro[order(repro$race,repro$lat),],ordered = is.ordered(repro$pop))
latord$pop = factor(latord$pop,
levels=levels(latord$pop)[order(populations$race, populations$lat)])
boxplot(latord$ff.time~latord$pop, cex.axis=0.7, xlab="Population", ylab="Days to first flower", las=1); segments(12.5,0,12.5,100, lwd=2)

#Ordering first by race then by longitude
longord = data.frame(repro[order(repro$race,repro$long),],ordered = is.ordered(repro$pop))
longord$pop = factor(longord$pop,
levels=levels(longord$pop)[order(populations$race, populations$long)])
boxplot(longord$ff.time~longord$pop, cex.axis=0.75, las=1)
```
Might be interesting to see the variance of quantitative traits across populations with the prediction that the NC race would have a lower amount of variance owing to its smaller population size.

```r
boxplot((repro$ff.time)~repro$pop) #Visually it's hard to tell.
var(na.omit(mydat.scale$ff.time[1:361]));
var(na.omit(mydat.scale$ff.time[362:1019])) #The TX race does have higher variance
var(mydat.scale$wid.2[1:361]); var(mydat.scale$wid.2[362:1019])
var(mydat.scale$len.2[1:361]); var(mydat.scale$len.2[362:1019])
var(mydat.scale$inflo.num[1:361]); var(mydat.scale$inflo.num[362:1019])
var(mydat.scale$inflo.H[1:361]); var(na.omit(mydat.scale$inflo.H[362:1019]))
var(mydat.scale$stem.diam[1:361]); var(mydat.scale$stem.diam[362:1019])
var(mydat.scale$bio.veg[1:361]); var(mydat.scale$bio.veg[362:1019])
```

#With scaled variance, the TX race has higher variance for time to first flower (1.0 vs 0.87), stem diameter (1.03 vs 0.92), vegetative biomass (1.02 vs 0.94) and width of the largest leaf (1.1 vs 0.6). The NC race had higher variance for inflorescence height (1.1 vs 0.9), number of inflorescences (0.96 vs 0.92), and length of the largest leaf (1.03 vs 0.97)

```r

```

### MANOVAs
```r

```

#With interaction terms.
mymod.int =
manova(cbind(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg,bio.rep)
~ race*pop*sex*fam, data=na.omit(repro))
summary(mymod.int)
#Race, population, sex, and block were all significant. Significant interaction
terms between: race:block, pop:block, race:sex, pop:sex, and pop:block:sex

summary.aov(mymod.int)
#Everything is significantly different between the races, except inflorescence
height and vegetative biomass. Race explained the greatest proportion of
variance for time to first flower, width of the largest leaf, number of
inflorescences,

#Significant differences between the sexes for: bolting time, time to first
flower, length of the largest leaf, inflorescence number, inflorescence height,
stem diameter, and vegetative and reproductive biomasses. Sex explained the
greatest proportion of variance for length of the largest leaf, inflorescence
height, stem diameter, vegetative biomass, and reproductive biomass.

#Block had an effect on log(len.2), log(wid.2), and inflorescence number
dim(repro[repro$race == "N" & repro$block == 1,])
dim(repro[repro$race == "T" & repro$block == 1,])
dim(repro[repro$race == "N" & repro$block == 2,])
dim(repro[repro$race == "T" & repro$block == 2,])
# The number of individuals of a given race did not differ to great degrees
between blocks. Block 1 had 183 NC individuals and 328 TX individuals and
block 2 had 178 NC individuals and 330 TX individuals. However, there were
a lot more females than males in the second block
Supplementary Tables

### Not sure the above is especially informative or easy to interpret. Using a simpler form, with flowering time and a sex:race term

```r
mymod.int2 = anova(lm(ff.time ~ race*pop*sex, data=na.omit(repro)));
mymod.int2
```

#All terms significant except for pop:sex

```r
mymod.int3 = anova(lm(ff.time ~ race*sex+pop, data=repro)); mymod.int3
```

#All terms significant

### Permutational multivariate analysis of variance using ADONIS

```r
myad = na.omit(repro[,c(1,2,13,14,17:20,23,24,27,28)])
adonis(cbind(as.numeric(myad$ff.time), as.numeric(myad$len.2),
as.numeric(myad$wid.2), as.numeric(myad$inflo.num),
as.numeric(myad$inflo.H), as.numeric(myad$stem.diam),
10^as.numeric(myad$bio.veg)) ~ myad$race * myad$sex * myad$pop,
permutations=100000, strata=myad$block)
```

### Df SumsOfSqs MeanSqs F.Model R2 Pr(>F)

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>race</td>
<td>1</td>
<td>0.0819</td>
<td>0.08188</td>
<td>8.074</td>
<td>0.01227</td>
<td>0.0022**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>0.6893</td>
<td>0.68931</td>
<td>67.967</td>
<td>0.10332</td>
<td>9.999e-05***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>race:sex</td>
<td>1</td>
<td>0.0077</td>
<td>0.00769</td>
<td>0.758</td>
<td>0.00115</td>
<td>0.4256</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>581</td>
<td>5.8924</td>
<td>0.01014</td>
<td></td>
<td>0.88325</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>584</td>
<td>6.6713</td>
<td>0.01014</td>
<td></td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#Trying this approach based on several comments on R-sig-eco about performing a nested permanova.
adonis(cbind(as.numeric(myad$ff.time), as.numeric(myad$len.2),
as.numeric(myad$wid.2), as.numeric(myad$inflo.num),
as.numeric(myad$inflo.H), as.numeric(myad$stem.diam),
10^as.numeric(myad$bio.veg)) ~ myad$race * myad$sex + myad$pop, strata
= myad$pop, data=myad, permutations=10000)

#Doing the same as above but with latitude and longitude
adonis(cbind(as.numeric(myad$ff.time), as.numeric(myad$len.2),
as.numeric(myad$wid.2), as.numeric(myad$inflo.num),
as.numeric(myad$inflo.H), as.numeric(myad$stem.diam),
10^as.numeric(myad$bio.veg)) ~ myad$sex + myad$race + myad$lat *
myad$long, permutations=10000)

################################################################
##################################################
### Partitioning variance with REML using lme4
###

mymod1 = lmer(ff.time ~ race * sex + (1 | race/pop/fam), data=mydat.scale,
REML=TRUE)
summary(mymod1)
Anova(mymod1)
#If individuals don’t vary consistently across treatments, the variance term will approach 0, and at the very least be less than the residual term.

mymod2 = lmer(len.2 ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE); mymod3 = lmer(wid.2 ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE); mymod4 = lmer(inflo.num ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE); mymod5 = lmer(inflo.H ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE); mymod6 = lmer(stem.diam ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE); mymod7 = lmer(bio.veg ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE)

mod1.pvals = pvals.fnc(mymod1); mod2.pvals = pvals.fnc(mymod2); mod3.pvals = pvals.fnc(mymod3);mod4.pvals = pvals.fnc(mymod4);mod5.pvals = pvals.fnc(mymod5);mod6.pvals = pvals.fnc(mymod6);mod7.pvals = pvals.fnc(mymod7)

#Generates confidence intervals and estimates of the proportion of variance attributable to independent variables based on MCMC sampling of the posterior distribution of model parameters.

#This extracts the standard deviation explained by the predictor variables for each response variable
mod1.std = mod1.pvals$random[,3]; mod2.std = mod2.pvals$random[,3];
mod3.std = mod3.pvals$random[,3]; mod4.std = mod4.pvals$random[,3];
mod5.std = mod5.pvals$random[,3]; mod6.std =
mod6.pvals$random[,3];mod7.std = mod7.pvals$random[,3]

mod.vars =
cbind(mod1.std,mod2.std,mod3.std,mod4.std,mod5.std,mod6.std,mod7.std)

#Putting the standard deviation explained by predictor variables into a table
prop.var = data.frame(ff.time=rep(NA,4), len.2=rep(NA,4), wid.2=rep(NA,4),
inflo.num=rep(NA,4), inflo.H=rep(NA,4), stem.diam=rep(NA,4),
bio.veg=rep(NA,4)); row.names(prop.var) = c("Family", "Population", "Race",
"Residual")

for (i in 1:7){
    prop.var[1,i] = mod.vars[1,i]/sum(mod.vars[,i])
    prop.var[2,i] = mod.vars[2,i]/sum(mod.vars[,i])
    prop.var[3,i] = mod.vars[3,i]/sum(mod.vars[,i])
    prop.var[4,i] = mod.vars[4,i]/sum(mod.vars[,i])
}

prop.var #This is the table of proportion of total variance for each level of
grouping for each variable

fam.av = apply(t(prop.var[,1]),2, mean); fam.sd = apply(t(prop.var[,1]),2, sd)
pop.av = apply(t(prop.var[,2]),2, mean); pop.sd = apply(t(prop.var[,2]),2, sd)
race.av = apply(t(prop.var[,3]),2, mean); race.sd = apply(t(prop.var[,3]),2, sd)
resid.av = apply(t(prop.var[,4]),2, mean); resid.sd = apply(t(prop.var[,4]),2, sd)
#Calculating average variance explained by each level of predictor variable
#and the standard deviation of that calculation

plot(c(fam.av, pop.av, race.av, resid.av), xaxt="n", xlab="", ylab="Proportion of Variance",
ylim=c(0,0.65), las=1, pch=16, family="Century Schoolbook")
#Plotting proportion of total variance explained by each term in the model.
axis(1, at=c(1,2,3,4), labels=c("Family", "Population", "Race", "Residual"),
tick=TRUE, lwd=0, lwd.ticks=1, family="Century Schoolbook")
#Adding and labelling the x-axis
Supplementary Tables

segments(1,fam.av+fam.sd,1,fam.av-fam.sd, lwd=2);
segments(2,pop.av+pop.sd,2,pop.av-pop.sd, lwd=2);
segments(3,race.av+race.sd,3,race.av-race.sd, lwd=2);
segments(4,resid.av+resid.sd,4,resid.av-resid.sd, lwd=2) #adding error bars
equal to the standard deviation

### Using lme to perform nested ANOVAs on a trait by trait basis

### (From http://www.utstat.utoronto.ca/reid/sta410/mar24.pdf). bottom line
is if you are interested in estimating the components of variance $\sigma^2$ and $\sigma^2_1$
then it is better to use method REML, but if you are interested in comparing
models, it is better to use method ML.

mymod10 = lmer(ff.time ~ race * sex + (1|race/pop/fam), data=mydat.scale,
REML=FALSE, family="gaussian")
mymod11 = lmer(ff.time ~ race + (1|race/pop/fam), data=mydat.scale,
REML=FALSE, family="gaussian")
mymod12 = lmer(ff.time ~ sex + (1|race/pop/fam), data=mydat.scale,
REML=FALSE, family="gaussian")
mymod13 = lmer(ff.time ~ 1 + (1|race/pop/fam), data=mydat.scale,
REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Does not differ significantly by race (likelihood
ratio test: 0.0597)
anova(mymod12, mymod13) # Differ significantly by sex (likelihood ratio test: 6.08e-5)

### BUT
Anova(mymod11)
# If I just run an analysis of the deviance table, race is a significant term.
Anova(mymod10)
# I also get a significant term if I run it on the full model.
# p=0.029 for race, p=4.64e-5 for sex and p=0.0005 for race:sex
Anova(mymod12)

mymod10 = lmer(len.2 ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod11 = lmer(len.2 ~ race + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(len.2 ~ sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(len.2 ~ 1 + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) # Does not differ significantly between races (likelihood ratio test: 0.101)
anova(mymod12,mymod13) # Differ significantly by sex (2.43e-11)
Anova(mymod10) # Only significant term is sex. p=0.118 for race. p=1.55-11 for sex and p=0.484 for race:sex
Anova(mymod12)

mymod10 = lmer(wid.2 ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
Supplementary Tables

mymod11 = lmer(wid.2 ~ race + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(wid.2 ~ sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(wid.2 ~ 1 + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Differs significantly between races (likelihood ratio test: 0.00505)
anova(mymod12,mymod13) #Does not differs significantly between sexes (likelihood ratio test: 0.059)
Anova(mymod10) #p=3.96e-10 for race. p=0.0564 for sex and p=0.00116 for race:sex

mymod10 = lmer(inflo.num ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod11 = lmer(inflo.num ~ race + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(inflo.num ~ sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(inflo.num ~ 1 + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Differs significantly between races (likelihood ratio test: 0.0234)
anova(mymod12,mymod13) #Differs significantly by sex (likelihood ratio test: 2.7e-5)
Anova(mymod10) #p=0.002384 for race. p=2.64e-5 for sex and p=0.731 for race:sex
mymod10 = lmer(inflo.H ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod11 = lmer(inflo.H ~ race + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(inflo.H ~ sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(inflo.H ~ 1 + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Does not differ significantly between races (likelihood ratio test: 0.99)
anova(mymod12,mymod13) #Differs significantly between the sexes (likelihood ratio test: 2.2e-16)
Anova(mymod10) #p=0.743 for race. p=2.2e-16 for sex and p=0.547 for race:sex

mymod10 = lmer(stem.diam ~ race * sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod11 = lmer(stem.diam ~ race + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(stem.diam ~ sex + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(stem.diam ~ 1 + (1 | race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Does not differ significantly (likelihood ratio test: 0.26)
anova(mymod12,mymod13) #Differs significantly between the sexes (likelihood ratio test: 2.2e-16)
Supplementary Tables

Anova(mymod10) #p=0.195 for race. p=2.2e-16 for sex and p=0.0362 for race:sex

mymod10 = lmer(bio.veg ~ race * sex + (1|race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod11 = lmer(bio.veg ~ race + (1|race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod12 = lmer(bio.veg ~ sex + (1|race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")
mymod13 = lmer(bio.veg ~ 1 + (1|race/pop/fam), data=mydat.scale, REML=FALSE, family="gaussian")

anova(mymod11,mymod13) #Does not differ significantly (0.302)
anova(mymod12,mymod13) #Differs significantly between the sexes (likelihood ratio test 2.2e-16)
Anova(mymod10) #p=0.234 for race. p=2.2e-16 for sex and p=0.0982 for race:sex

#########################################################################
#########################################################################
### Performing a PCA on grow out variables with non-reproductive individuals omitted
#########################################################################
#########################################################################

dat.pca1 = na.omit(repro[,c(1,13,14,17:20,23:24,27,28)])
dat.pca1$race = as.numeric(dat.pca1$race)
dat.pca1 = dat.pca1[order(dat.pca1[,1], dat.pca1[,4]),]
pairs(dat.pca1, gap=0, las=1, cex.axis=0.75, cex=0.6, pch=16)
#Going to omit reproductive biomass because it is highly skewed.
#The data should have a multivariate normal distribution

pca.1 = prcomp(na.omit(repro[,c(13,17:20,23:24)]), center=TRUE, scale=TRUE); summary(pca.1)
#It takes 5 (of 7) principal components to explain ~92.2% of the variance
pca.2 = prcomp(dat.pca1[-c(1,3,10,11)], center=TRUE, scale=TRUE); summary(pca.2)
#If I omit sex, race, and bolting time, it takes 5 (of 7) principal components to explain 92.21% of the variance, but the loadings on the first 2 principal components are 45.02% and 19.49%.

pca.3 = prcomp(dat.pca1[-c(1,3)], center=TRUE, scale=TRUE); summary(pca.3)
#With latitude and longitude included it takes 6 of 9 principal components to explain ~89.8% of the variance

#Extracting scores, loadings, and standard deviation from the pca
pca.scores1 = pca.2$x; pca.loadings1 = pca.2$rotation; pca.sdev1 = pca.2$sdev

###
### 1 by 2
###
###
# Plotting the results of the PCA in a Correlation biplot - First scales the variables to have unit variances then adjusts the points and the axes of the PCA biplot so that the cosines of the angles between the axes approximate the correlations between the corresponding variables (from http://www.jstatsoft.org/v30/i12/paper).
plot(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,2]/pca.sdev1[2],
xlab=expression("PC1 " *lambda* "=44.80\%")), ylab=expression("PC2 " *lambda* "=18.25\%")), type="n", las=1, family="Century Schoolbook");
segments(100,0,-100,0); segments(0,100,0,-100)

# Adding points
# Males are represented by circles, females by squares. North Carolina is
# gray and open, Texas is black.
points(pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==1),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==1),3]/pca.sdev1[2], pch=1, cex=0.75, col="gray15");
points(pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==0),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==0),2]/pca.sdev1[2], pch=2, cex=0.75, col="gray15");
points(pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==1),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==1),2]/pca.sdev1[2], pch=16, cex=0.75, col="black");
points(pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==0),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==0),2]/pca.sdev1[2], pch=17, cex=0.75, col="black")

#Adding arrows and labelling them
arrows(0,0,(pca.loadings1[,1]/pca.sdev1[1])*3,(pca.loadings1[,2]/pca.sdev1[2])*3, lwd=1.5, length=0.15, angle=25); text(pca.loadings1[,1]*3.3,
pca.loadings1[,2]*c(rep(3.3,6),-5), labels=names(pca.loadings1[,2]), cex=1.2)

#Adding a legend
legend(-3.7,3.65, legend = expression("Texas Males", "Texas Females", "North
Carolina Males", "North Carolina Females"), col = c("red", "red", "blue",
"blue"), pch = c(1,2,1,2))
### 1 by 3

plot(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,3]/pca.sdev1[3],
xlab=expression("PC1 " *lambda* "=44.80%"), ylab=expression("PC3 " *lambda* "=11.10%"), type="n", las=1, family="Century Schoolbook");
segments(100,0,-100,0); segments(0,100,0,-100)

# Adding points
# Males are represented by circles, females by squares. North Carolina is gray and open, Texas is black.
points(pca.scores1[which(dat.pca1$race=="2" & dat.pca1$sex==1),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="2" & dat.pca1$sex==1),3]/pca.sdev1[3], pch=1, cex=0.75, col="gray15");
points(pca.scores1[which(dat.pca1$race=="2" & dat.pca1$sex==0),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="2" & dat.pca1$sex==0),3]/pca.sdev1[3], pch=2, cex=0.75, col="gray15");
points(pca.scores1[which(dat.pca1$race=="1" & dat.pca1$sex==1),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="1" & dat.pca1$sex==1),3]/pca.sdev1[3], pch=16, cex=0.75, col="black");
points(pca.scores1[which(dat.pca1$race=="1" & dat.pca1$sex==0),1]/pca.sdev1[1], pca.scores1[which(dat.pca1$race=="1" & dat.pca1$sex==0),3]/pca.sdev1[3], pch=17, cex=0.75, col="black")

#Adding arrows and labelling them
arrows(0,0,(pca.loadings1[,1]/pca.sdev1[1])*3,(pca.loadings1[,3]/pca.sdev1[3])*3, lwd=1.5, length=0.15, angle=25); text(pca.loadings1[,1]*3.3, pca.loadings1[,3]*3.3, labels=names(pca.loadings1[,1]), cex=1.2)
### 2 by 3

```r
plot(pca.scores1[,2]/pca.sdev1[2], pca.scores1[,3]/pca.sdev1[3],
xlab=expression("PC2 " *lambda* "=19.49\%"), ylab=expression("PC3 " *lambda* "=11.10\%"), type="n", las=1, family="Century Schoolbook");
segments(100,0,-100,0); segments(0,100,0,-100)

# Adding points
# Males are represented by circles, females by squares. North Carolina is blue, Texas is red.
points(pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==1),2]/pca.sdev1[2], pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==1),3]/pca.sdev1[3], pch=1, cex=0.75, col = "gray15");
points(pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==0),2]/pca.sdev1[2], pca.scores1[which(dat.pca1$race=="2" &
dat.pca1$sex==0),3]/pca.sdev1[3], pch=2, cex=0.75, col = "gray15");
points(pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==1),2]/pca.sdev1[2], pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==1),3]/pca.sdev1[3], pch=16, cex=0.75, col = "black");
points(pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==0),2]/pca.sdev1[2], pca.scores1[which(dat.pca1$race=="1" &
dat.pca1$sex==0),3]/pca.sdev1[3], pch=17, cex=0.75, col = "black")

#Adding arrows and labelling them
arrows(0,0,(pca.loadings1[,2]/pca.sdev1[2])*3,(pca.loadings1[,3]/pca.sdev1[3])*3, lwd=1.5, length=0.15, angle=25); text(pca.loadings1[,2]*c(3.75), pca.loadings1[,3]*3.75, labels=names(pca.loadings1[,2]), cex=1.2)
```
### Plotting the PCA scores in 3D ###

#Using the scatterplot3d package

```r
require(scatterplot3d)
path.name = "/Users/andrewsimpson/rplots/"

for (i in 1:360){
    pdf(file = paste(path.name, "myplot_", i, ".pdf", sep=""))
    s3d = scatterplot3d(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,2]/pca.sdev1[2], pca.scores1[,3]/pca.sdev1[3], type="n", pch = ",", lwd=0.15, las=1, xlab=expression("PC1 * lambda* " = 45.02%), ylab=expression("PC2 * lambda* " = 19.49%), zlab=expression("PC3 * lambda* " = 11.10%), angle=i); s3d.coords = s3d$xyz.convert(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,2]/pca.sdev1[2], pca.scores1[,3]/pca.sdev1[3]); points(s3d.coords$x[which(dat.pca1$race=="2" & dat.pca1$sex==1)], s3d.coords$y[which(dat.pca1$race=="2" & dat.pca1$sex==1)], pch=16, cex=0.5); points(s3d.coords$x[which(dat.pca1$race=="2" & dat.pca1$sex==0)], s3d.coords$y[which(dat.pca1$race=="2" & dat.pca1$sex==0)], pch=17, cex=0.5); points(s3d.coords$x[which(dat.pca1$race=="1" & dat.pca1$sex==1)], s3d.coords$y[which(dat.pca1$race=="1" & dat.pca1$sex==1)], pch=1, cex=0.75); points(s3d.coords$x[which(dat.pca1$race=="1" & dat.pca1$sex==0)], s3d.coords$y[which(dat.pca1$race=="1" & dat.pca1$sex==0)], pch=2, cex=0.75)
    arrows(x0 = s3d$xyz.convert(x=0,y=0,z=0)$x, y0 = s3d$xyz.convert(x=0,y=0,z=0)$y, x1 = s3d$xyz.convert(x=pca.loadings1[,1]*3,y=pca.loadings1[,2]*3,z=pca.loadings1[,3]*3)$x, y1 = s3d$xyz.convert(x=pca.loadings1[,1]*3,y=pca.loadings1[,2]*3,z=pca.loadings1[,3]*3)$y, lwd=2)
}
```
# Supplementary Tables

text(x = s3d$xyz.convert(x=pca.loadings1[,1]*3.3,y=pca.loadings1[,2]*
3.3,z=pca.loadings1[,3]*3.3)$x, y = s3d$xyz.convert(x=pca.loadings1[,1]*
3.3,y=pca.loadings1[,2]*3.3,z=pca.loadings1[,3]*3.3)$y, lwd=2,
labels=names(dat.pca1[-c(1,3)]), lwd=2)
dev.off()

# Angle of 265 is good
s3d = scatterplot3d(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,2]/pca.sdev1[2],
pca.scores1[,3]/pca.sdev1[3], type="n", pch = "", lwd=0.15, las=1,
xlab=expression("PC1 " *lambda* " = 45.02%"), ylab=expression("PC2 "
*lambda* " = 19.49%"), zlab=expression("PC3 " *lambda* " = 11.10%"),
angle=265., family="Century Schoolbook"); s3d.coords =
s3d$xyz.convert(pca.scores1[,1]/pca.sdev1[1], pca.scores1[,2]/pca.sdev1[2],
pca.scores1[,3]/pca.sdev1[3]); points(s3d.coords$x[which(dat.pca1$race=="2" &
dat.pca1$sex==1)], s3d.coords$y[which(dat.pca1$race=="2" &
dat.pca1$sex==1)], pch=1, cex= 1);
points(s3d.coords$x[which(dat.pca1$race=="2" & dat.pca1$sex==0)],
points(s3d.coords$x[which(dat.pca1$race=="1" & dat.pca1$sex==1)], s3d.coords$y[which(dat.pca1$race=="1" &
dat.pca1$sex==0)], pch=16, cex= 1); points(s3d.coords$x[which(dat.pca1$race=="1" & dat.pca1$sex==0)],
s3d.coords$y[which(dat.pca1$race=="1" & dat.pca1$sex==0)], pch=17, cex= 1);
arrows(x0 = s3d$xyz.convert(x=0,y=0,z=0)$x, y0 =
s3d$xyz.convert(x=0,y=0,z=0)$y, x1 =
s3d$xyz.convert(x=pca.loadings1[,1]*5,y=pca.loadings1[,2]*5,z=pca.loadings1[
,3]*5)$x, y1 =
s3d$xyz.convert(x=pca.loadings1[,1]*5,y=pca.loadings1[,2]*5,z=pca.loadings1[
,3]*5)$y, lwd=2); text(x =
s3d$xyz.convert(x=pca.loadings1[,1]*5.5,y=pca.loadings1[,2]*
5.5,z=pca.loadings1[,3]*5.5)$x, y = s3d$xyz.convert(x=pca.loadings1[,1]*
5.5,y=pca.loadings1[,2]*5.5,z=pca.loadings1[,3]*5.5)$y, lwd=2)
Supplementary Tables

5.5, z = pca.loadings1[,3] * 5.5
5.3, y = pca.loadings1[,2] * 5.3, z = pca.loadings1[,3] * 5.3

$\text{x, y = s3d$xyz.convert(x=pca.loadings1[,1] * 5.3, y = pca.loadings1[,2] * 5.3, z = pca.loadings1[,3] * 5.3)$y, lwd=2, labels=names(dat.pca1[,1]), family="Century Schoolbook")$

nc.pca = subset(pca.2$x, dat.pca1$race==1)
tx.pca = subset(pca.2$x, dat.pca1$race==2)
t.test(nc.pca[,1], tx.pca[,1]); t.test(nc.pca[,2], tx.pca[,2]); t.test(nc.pca[,3], tx.pca[,3])

### It appears as though TX and NC differ significantly for their loadings on the second and third principal components.

head(cbind(dat.pca1$race, dat.pca1$sex, pca.2$x))
testdat = data.frame(race = dat.pca1$race, sex = dat.pca1$sex, PC1 = pca.2$x[,1], PC2 = pca.2$x[,2], PC3 = pca.2$x[,3])

# Don't need to perform a manova since the principal components are orthogonal.
Anova(lm(PC1 ~ race*sex, data=testdat)) # p for race is 0.402, 2.2e-16 for sex, 0.825 for interaction
Anova(lm(PC2 ~ race*sex, data=testdat)) # p for race is 2.2-16, 0.000441 for sex, 0.0643 for interaction
Anova(lm(PC3 ~ race*sex, data=testdat)) # p for race is 8.8e-14, 0.000377 for sex, 0.000306 for interaction

### Correlations between environmental variables and latitude and longitude
Supplementary Tables

###

#Some values omitted because of missing data. Bear with ugly, repetitive code.
corr(import2$Lat, import2$mean.annual.precip, method="pearson")
#0.06649606
corr(import2$Lat, import2$mean.annual.precip, method="kendall")
#0.02562103
corr(import2$Lat, import2$mean.annual.precip, method="spearman")
#0.07448107
corr(import2$Lat[-c(2,13,20,21,25,26)], import2$DT32[-c(2,13,20,21,25,26)])
#0.6047093
corr(import2$Lat[-c(2,13,20,21,25,26)], import2$DT90[-c(2,13,20,21,25,26)])
#-0.3074369
corr(import2$Long, import2$mean.annual.precip) #0.03077789
corr(import2$Long[-c(2,13,20,21,25,26)], import2$DT32[-c(2,13,20,21,25,26)])
#-0.1784211
corr(import2$Long[-c(2,13,20,21,25,26)], import2$DT90[-c(2,13,20,21,25,26)])
#-0.6342274
#The strongest correlations are between latitude and degree days below 0°C and longitude and degree days above 32°C.

summary(lm(mean.annual.precip~Lat, data=subset(import2,import2$Race=="NC"))) #0.7417
summary(lm(DT32 ~ Lat, data = import2[-c(2,13,20,21,25,26),])) #0.00369
summary(lm(DT90 ~ Lat, data = import2[-c(2,13,20,21,25,26),])) #0.175
summary(lm(mean.annual.precip ~ Long, data = import2)) #0.8789
summary(lm(DT32 ~ Long, data = import2[-c(2,13,20,21,25,26),])) #0.429
Supplementary Tables

summary(lm(DT90 ~ Long, data = import2[-c(2,13,20,21,25,26),])) #0.00202
#The only significant correlations are between latitude and degree days below 0°C and longitude and degree days above 32°C.

cov(cbind(scale(import2$Lat[-c(2,13,20,21,25,26)]),scale(import2$Long[-c(2,13,20,21,25,26)]),scale(import2$DT32[-c(2,13,20,21,25,26)]),scale(import2$DT90[-c(2,13,20,21,25,26)]),
  scale(import2$mean.annual.precip[-c(2,13,20,21,25,26)])))
#Note the high degree of covariance between numerous factors

cor.test(import2$mean.annual.precip,populations$prop.repro)
cor.test(import2$DT32[-c(2,13,20,21,25,26)],populations$prop.repro[-c(2,13,20,21,25,26)])
cor.test(import2$DT90[-c(2,13,20,21,25,26)],populations$prop.repro[-c(2,13,20,21,25,26)])
cor.test(import2$mean.annual.precip,populations$prop.fem)
cor.test(import2$DT32[-c(2,13,20,21,25,26)],populations$prop.fem[-c(2,13,20,21,25,26)])
cor.test(import2$DT90[-c(2,13,20,21,25,26)],populations$prop.fem[-c(2,13,20,21,25,26)])
#No significant correlations between any climatic variables and proportion of reproductive individuals or proportion of females.

###Interestingly, there is a strong correlation between annual precipitation and longitude for the TX populations but not for the NC populations.
Anova(lm(subset(import2$mean.annual.precip,import2$Race=='NC')~subset(import2$Long,import2$Race=='NC')))  
Anova(lm(subset(import2$mean.annual.precip,import2$Race=='TX')~subset(import2$Long,import2$Race=='TX')))
Anova(lm(subset(import2$DT32,import2$Race=='TX')~subset(import2$Lat,import2$Race=='TX')))  
plot(subset(import2$mean.annual.precip,import2$Race=='TX')~subset(import2$Long,import2$Race=='TX'), family="Century Schoolbook", las=1)  

pca.out = data.frame(race = na.omit(repro[,c(1:4,13,14,17:20,23:24)])[1], pop = na.omit(repro[,c(1:4,13,14,17:20,23:24)])[2], lat = na.omit(repro[,c(1:4,13,14,17:20,23:24,27)])[13], long = na.omit(repro[,c(1:4,13,14,17:20,23:24,28)])[13], PC1 = pca.2$x[1], PC2 = pca.2$x[2], PC3 = pca.2$x[3])  

summary(lm(pca.out$PC1~pca.out$lat)) #p-value : 2.2e-16  
summary(lm(pca.out$PC2~pca.out$lat)) #p-value : 0.995  
summary(lm(pca.out$PC3~pca.out$lat)) #p-value : 0.0008  
#P values change dramatically when the PCA doesn't include race or sex.  
 Ignore the above values for now  

plot(pca.out$lat, pca.out$PC1, las=1)  
PC1 = pca.out$PC1; PC2 = pca.out$PC2; PC3 = pca.out$PC3; pclat = pca.out$lat  
x1 = (seq(min(repro$lat), max(repro$lat), 0.0001))  
y1 = predict(lm(PC1~pclat), list(pclat=x1))  
yv1 = predict(lm(PC1~pclat), list(pclat=x1), interval='confidence')  
lines(x1,y1)  
matlines(x1,yv1,lty=c(1,2,2),col='black')  
#Sample plotting code.  

plot(pca.out$lat, pca.out$PC2, las=1); y2 = predict(lm(PC2~pclat), list(pclat=x1)); yv2 = predict(lm(PC2~pclat), list(pclat=x1),
Supplementary Tables

```r
interval='confidence'); lines(x1,y2); matlines(x1,yv2,lty=c(1,2,2),col='black')
#Plotting the regression
hist(resid(lm(PC2~pclat))); qqnorm(resid(lm(PC2~pclat))); qqline(resid(lm(PC2~pclat))); shapiro.test(resid(lm(PC2~pclat))) #Testing for normality

plot(pca.out$lat, pca.out$PC3, las=1); y3 = predict(lm(PC3~pclat), list(pclat=x1)); yv3 = predict(lm(PC3~pclat), list(pclat=x1), interval='confidence'); lines(x1,y3); matlines(x1,yv3,lty=c(1,2,2),col='black')
hist(resid(lm(PC3~pclat))); qqnorm(resid(lm(PC3~pclat))); qqline(resid(lm(PC3~pclat))); shapiro.test(resid(lm(PC3~pclat)))

### Correlations between latitude and traits values

fflat = lm(ff.time~lat, data=repro); summary(fflat) #p=6.827e-6 adjusted r-squared=0.0199
lenlat = lm(len.2~lat, data=repro); summary(lenlat) #p=2.001e-10 adjusted r-squared=0.0381
widlat = lm(wid.2~lat, data=repro); summary(widlat) #p=0.1468 adjusted r-squared=0.0011
numlat = lm(inflo.num~lat, data=repro); summary(numlat) #p=0.002 adjusted r-squared=0.0083
heilat = lm(inflo.H~lat, data=repro); summary(heilat) #p=4.84e-8 adjusted r-squared=0.0279
stemlat = lm(stem.diam~lat, data=repro); summary(stemlat) #p=3.8e-15 adjusted r-squared=0.0581
biolat = lm(bio.veg~lat, data=repro); summary(biolat) #p=2e-16 adjusted r-squared=0.1165
```
Supplementary Tables

# Strongest correlation between latitude and an environmental variable is stem diameter

plot(stem.diam~lat, data=repro); y3 = predict(lm(stem.diam~lat, data=repro), list(lat=x1)); yv3 = predict(lm(stem.diam~lat, data=repro), list(lat=x1), interval='confidence'); lines(x1,y3); matlines(x1,yv3,lty=c(1,2,2),col='black')

qqnorm(resid(stemlat)); qqline(resid(stemlat)); shapiro.test(resid(stemlat))

# Stem diameter is the only variable with normally distributed residuals.

plot(repro$stem.diam~repro$lat, data=repro); ystem = predict(stemlat,list(lat=x1)); yvstem = predict(stemlat, list(lat=x1), interval='confidence'); lines(x1,ystem); matlines(x1,yvstem,lty=c(1,2,2), col="black")

#########################################################################
### Performing a MANCOVA for trait variation against latitude and longitude
###
#########################################################################

Anova(lm(as.matrix(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg) ~ lat * long + race + sex, data=mydat.scale))

# All predictor variables significant except the interaction term

Anova(lm(as.matrix(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg) ~ lat + long + race + sex, data=na.omit(mydat.scale)))
#Comparing effect sizes of race and sex with and without latitude and longitude as covariates. After Colautti et al. 2009

Anova(lm(as.matrix(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg) ~ race + sex, data=na.omit(mydat.scale)))

#Effect size decreased for race (48.465 vs. 36.767) and remained about the same for sex (19.076 vs 20.073)
summary(lm(as.matrix(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg) ~ lat + long + race + sex, data=na.omit(mydat.scale)))
summary(lm(as.matrix(ff.time,len.2,wid.2,inflo.num,inflo.H,stem.diam,bio.veg) ~ race + sex, data=na.omit(mydat.scale)))

#These show that the effect size for race is larger when latitude and longitude are included (-1.88110 vs -0.40617) and about the same for sex (0.27001 vs 0.28351). However, the standard error is about 4 times higher for race (0.27021 vs 0.06699) and about the same for sex (0.06182 vs 0.06328)

testmod1 =
lmer(cbind(ff.time,len.2,wid.2,stem.diam,inflo.H,inflo.num,bio.veg) ~ race + sex + lat + long + (1 | race/pop/fam), data=mydat.scale, REML=TRUE, family="gaussian")
testmod2 =
lmer(cbind(ff.time,len.2,wid.2,stem.diam,inflo.H,inflo.num,bio.veg) ~ race + sex + (1 | race/pop/fam), data=mydat.scale, REML=TRUE, family="gaussian")

summary(testmod1); summary(testmod2)

#The results of LMEs are similar to the LMs. Effect sizes still increase for race with the inclusion of lat and long, but stay the same for sex. This is likely due to the longitudinal separation of the sex chromosome races and should become apparent when lat and long are included independently to the
Supplementary Tables

models. Interestingly, the LMEs have higher standard error than the LMs. I have no explanation for this at the moment.

```
summary(lmer(cbind(ff.time,len.2,wid.2,stem.diam,inflo.H,inflo.num,bio.veg) ~ race + sex + lat + (1 | race/pop/fam), data=mydat.scale, REML=TRUE, family="gaussian"))
summary(lmer(cbind(ff.time,len.2,wid.2,stem.diam,inflo.H,inflo.num,bio.veg) ~ race + sex + long + (1 | race/pop/fam), data=mydat.scale, REML=TRUE, family="gaussian"))
```
#As expected, adding a term for longitude increases the effect size of race to a greater degree than adding a term for latitude.

#Individual ANCOVAs. If terms are missing it's because the simplified models performed better than the more parameterized ones.
```
Anova(lm(ff.time ~ lat + long + race + sex, data=na.omit(mydat.scale)))
### One of the more important relationships. Note that, apart from residual effects, race has the largest effect size.
Anova(lm(len.2 ~ lat + long + race + sex, data=na.omit(mydat.scale)))
Anova(lm(wid.2 ~ lat + long + race + sex, data=na.omit(mydat.scale)))
Anova(lm(inflo.num ~ lat + long + race + sex, data=na.omit(mydat.scale)))
Anova(lm(inflo.H ~ lat + long + race + sex, data=na.omit(mydat.scale)))
Anova(lm(stem.diam ~ lat + long + race + sex, data=na.omit(mydat.scale)))
Anova(lm(bio.veg ~ lat + long + race + sex, data=na.omit(mydat.scale)))
```

#Linear mixed-effects models perform better
```
summary(lmer(ff.time ~ lat + long + race + sex + (1 | pop/fam), data=na.omit(mydat.scale)))
summary(lmer(len.2 ~ lat + long + race + sex + (1 | pop/fam), data=na.omit(mydat.scale)))
```
Supplementary Tables

summary(lmer(wid.2 ~ lat + long + race + sex + (1 | pop/fam), 
data=na.omit(mydat.scale)))

summary(lmer(inflo.num ~ lat + long + race + sex + (1 | pop/fam), 
data=na.omit(mydat.scale)))

summary(lmer(inflo.H ~ lat + long + race + sex + (1 | pop/fam), 
data=na.omit(mydat.scale)))

summary(lmer(stem.diam ~ lat + long + race + sex + (1 | pop/fam), 
data=na.omit(mydat.scale)))

summary(lmer(bio.rep ~ lat + long + race + sex + (1 | pop/fam), 
data=na.omit(mydat.scale)))

#Trying to plot these relationships

#Flowering Time
round(summary(ffmod)$coefficients[2,4],10)
round(summary(ffmod)$r.squared,4)
x = seq(min(repro$lat),max(repro$lat),0.01)
ffmod=lm(ff.time ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(ff.time ~ lat, data=na.omit(repro),
xlab="Latitude", ylab="Flowering time (days)", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35, 74,expression(italic(P)*" = 0.003"), cex=1.5, family="Century Schoolbook"); text(35, 78, expression("*italic(R)^2*" = 0.015"), cex=1.5, family="Century Schoolbook")

#Len.2
round(summary(ffmod)$coefficients[2,4],10)
round(summary(ffmod)$r.squared,4)
ffmod=lm(len.2 ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(len.2 ~ lat, data=na.omit(repro),

105
Supplementary Tables

xlab="Latitude", ylab="", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"; matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35,32,expression(italic(P)*" > 0.001"), cex=1.5, family="Century Schoolbook"); text(35, 48, expression(""*italic(R)^2*" = 0.040"), cex=1.5, family="Century Schoolbook"); mtext("Leaf length (mm)", side=2, line=3.5, family="Century Schoolbook", cex=1.5)

#Wid.2
round(summary(ffmod)$coefficients[2,4],10)
round(summary(ffmod)$r.squared,4)
ffmod=lm(wid.2 ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(wid.2 ~ lat, data=na.omit(repro), xlab="Latitude", ylab="Leaf width (mm)", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35,88,expression(italic(P)*" = 0.006"), cex=1.5, family="Century Schoolbook"); text(35, 94, expression(""*italic(R)^2*" = 0.013"), cex=1.5, family="Century Schoolbook")

#Stem Diameter
summary(ffmod)$coefficients[2,4]
round(summary(ffmod)$r.squared,4)
ffmod=lm(stem.diam ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(stem.diam ~ lat, data=na.omit(repro), xlab="Latitude", ylab="Stem diameter (mm)", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35,1,expression(italic(P)*" > 0.001"), cex=1.5, family="Century Schoolbook"); text(35, 1.4, expression(""*italic(R)^2*" = 0.080"), cex=1.5, family="Century Schoolbook")

#Inflorescence Height
Supplementary Tables

round(summary(ffmod)$coefficients[2,4],10)
round(summary(ffmod)$r.squared,4)
ffmod=lm(inflo.H ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(inflo.H ~ lat, data=na.omit(repro), xlab="Latitude", ylab="", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); mtext("Inflorescence height (mm)", side=2,line=4, family="Century Schoolbook", cex=1.5); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35,200,expression(italic(P)*" > 0.001"), cex=1.5, family="Century Schoolbook"); text(35, 275, expression("*italic(R)^2*" = 0.049"), cex=1.5, family="Century Schoolbook")

#Inflorescence Number
round(summary(ffmod)$coefficients[2,4],10)
round(summary(ffmod)$r.squared,4)
ffmod=lm(inflo.num ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(inflo.num ~ lat, data=na.omit(repro), xlab="Latitude", ylab="Number of inflorescences", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1), col="black"); text(35,-3.2,expression(italic(P)*" = 0.018"), cex=1.5, family="Century Schoolbook"); text(35, -2.8, expression("*italic(R)^2*" = 0.0095"), cex=1.5, family="Century Schoolbook")

#Vegetative Biomass
summary(ffmod)$coefficients[2,4]
round(summary(ffmod)$r.squared,4)
ffmod=lm(bio.veg ~ lat, data=na.omit(repro)); yc = predict(ffmod,list(lat=x), interval="c", level=0.95); plot(bio.veg ~ lat, data=na.omit(repro), xlab="Latitude", ylab="", las=1, cex.axis=1.5, cex.lab=1.5, family="Century Schoolbook"); mtext("Vegetative biomass (g)", side=2, line=3.5, cex=1.5, family="Century Schoolbook"); matlines(x,yc, lty=c(1,2,2), lwd=c(2,1,1),
Supplementary Tables

plot(ff.time ~ lat, data=na.omit(mydat.scale), xlab="Latitude", type="n", ylab="Flowering Time", las=1, family="Century Schoolbook")
x = seq(min(mydat.scale$lat), max(mydat.scale$lat), 0.0001)
y = predict(ffmod, list(lat=x))
yc = predict(ffmod, list(lat=x), interval="c", level=0.99)
lines(x,y); matlines(x,yc, col="black")

#Plotting Points with different plotting characters for each sex and race
points(subset(mydat.scale, race="N" & sex == "0")$lat, subset(mydat.scale, race="N" & sex == "0")$ff.time, pch=1);
points(subset(mydat.scale, race="N" & sex == "1")$lat, subset(mydat.scale, race="N" & sex == "1")$ff.time, pch=2);
points(subset(mydat.scale, race="T" & sex == "0")$lat, subset(mydat.scale, race="T" & sex == "0")$ff.time, pch=3);
points(subset(mydat.scale, race="T" & sex == "1")$lat, subset(mydat.scale, race="T" & sex == "1")$ff.time, pch=4)

#Generating linear models for each sex and race subset of the data
ffmod1=lm(ff.time ~ lat, data=subset(na.omit(mydat.scale),race="N" & sex == "0")); ffmod2=lm(ff.time ~ lat, data=subset(na.omit(mydat.scale),race="N" & sex == "1")); ffmod3=lm(ff.time ~ lat, data=subset(na.omit(mydat.scale),race="T" & sex == "0")); ffmod4=lm(ff.time ~ lat, data=subset(na.omit(mydat.scale),race="T" & sex == "1"))

#Generating lines of best fit and 95% confidence intervals
y1 = predict(ffmod1,list(lat=x)); yc1 = predict(ffmod1,list(lat=x), interval="c", level=0.95); y2 = predict(ffmod2,list(lat=x)); yc2 = predict(ffmod2,list(lat=x), interval="c", level=0.95); y3 = predict(ffmod3,list(lat=x)); yc3 = predict(ffmod3,list(lat=x), interval="c", level=0.95); y4 = predict(ffmod4,list(lat=x)); yc4 = predict(ffmod4,list(lat=x), interval="c", level=0.95)

#Plotting lines of best fit
lines(x,y1, lwd=2, col="darkred"); lines(x,y2, lwd=2, col="darkblue"); lines(x,y3, lwd=2, col="darkgreen"); lines(x,y4, lwd=2, col="darkgray")

#Plotting confidence intervals
matlines(x,yc1, lwd=0.75, col="darkred"); matlines(x,yc2, lwd=0.75, col="darkblue"); matlines(x,yc3, lwd=0.75, col="darkgreen"); matlines(x,yc4, lwd=0.75, col="darkgray")

#Follow-up analyses

### Follow-up analyses
Supplementary Tables

If there is a trade-off between reproductive timing and energy allocation to vegetative growth, one would expect to see a correlation between flowering time and biomass.

```
summary(lm(ff.time~bio.veg, data=subset(repro,repro$sex==0)))
summary(lm(ff.time~bio.veg, data=subset(repro,repro$sex==1)))
summary(lm(ff.time~bio.veg+sex, data=repro))
#We see just that! Plants with larger biomasses flowered later.
```

```
nc.rep = import1[which(import1$sex>=0 & import1$race=="N"),]
nr.rep = import1[-which(import1$sex>=0),]
```

```
t.test(nc.rep$len.1,nc.norep$len.1)
t.test(nc.rep$wid.1,nc.norep$wid.1)
#Significantly shorter and narrower leaves before flowering
```

```
t.test(nc.rep$bio.veg,nc.norep$bio.veg)
#Significantly greater vegetative biomass after flowering
```

### Plotting stuff Spencer requested

```
NC=subset(repro,repro$race=="N"); TX=subset(repro,repro$race=="T")

hist(TX$ff.time, las=1, family="Century Schoolbook", xlab="Days to first flower", main="", breaks=11, freq=F, ylim= c(0,0.07))
hist(NC$ff.time, las=1, family="Century Schoolbook", xlab="Days to first flower", main="", breaks=11, freq=F, ylim= c(0,0.07))
# These histograms compare the distribution of flowering times for the two races. Y-axes have been set to equal values and there are the same number of break points for both distributions. To account for unbalanced sample sizes the histograms show probability densities rather than counts.

Anova(lm(lmer(ff.time ~ sex + (1|pop/fam),
data=na.omit(subset(mydat.scale,mydat.scale$race=="T")), REML=FALSE, family="gaussian")))
Anova(lm(lmer(ff.time ~ sex + (1|pop/fam),
data=na.omit(subset(mydat.scale,mydat.scale$race=="N")), REML=FALSE, family="gaussian")))
#There is a more significant effect of sex on flowering time in the NC SCR than the TX SCR (F values of 21.995 and 5.6108, respectively)

Anova(lm(lmer(len.2 ~ sex + (1|pop/fam),
data=na.omit(subset(mydat.scale,mydat.scale$race=="T")), REML=FALSE, family="gaussian")))
Anova(lm(lmer(len.2 ~ sex + (1|pop/fam),
data=na.omit(subset(mydat.scale,mydat.scale$race=="N")), REML=FALSE, family="gaussian")))
#TX F:21.995 P: 4.074e-06
#NC F:8.7937 P:0.003253
Supplementary Tables

Anova(lm(wid.2 ~ sex + (1 | pop/fam),
data = na.omit(subset(mydat.scale, mydat.scale$race == "T"), REML = FALSE,
family = "gaussian")))

Anova(lm(wid.2 ~ sex + (1 | pop/fam), data =
na.omit(subset(mydat.scale, mydat.scale$race == "N"), REML = FALSE,
family = "gaussian")))

# No effect of sex on width for the TX race but significant effect of sex on width in the NC race (F values of 0.5074 and 20.557 and p values of 0.4765 and 8.225e-06, respectively)

Anova(lm(inflo.H ~ sex + (1 | pop/fam), data =
na.omit(subset(mydat.scale, mydat.scale$race == "T"), REML = FALSE,
family = "gaussian")))

Anova(lm(inflo.H ~ sex + (1 | pop/fam), data =
na.omit(subset(mydat.scale, mydat.scale$race == "N"), REML = FALSE,
family = "gaussian")))

# TX F: 129.62  P: <2.2e-16
# NC F: 34.898  P: 8.981e-09

Anova(lm(inflo.num ~ sex + (1 | pop/fam),
data = subset(mydat.scale, mydat.scale$race == "T"), REML = FALSE,
family = "gaussian"))

Anova(lm(inflo.num ~ sex + (1 | pop/fam),
data = subset(mydat.scale, mydat.scale$race == "N"), REML = FALSE,
family = "gaussian")))

# TX F: 7.0563 P: 0.008091
# NC F: 4.1773 P: 0.0417
Supplementary Tables

Anova(lm(lmer(stem.diam ~ sex + (1|pop/fam),
data=subset(mydat.scale,mydat.scale$race=="T"), REML=FALSE,
family="gaussian")))
Anova(lm(lmer(stem.diam ~ sex + (1|pop/fam),
data=subset(mydat.scale,mydat.scale$race=="N"), REML=FALSE,
family="gaussian")))
#TX F:80.333 P:<2.2e-16
#NC F:83.738 P:<2.2e-16

Anova(lm(lmer(bio.veg ~ sex + (1|pop/fam),
data=subset(mydat.scale,mydat.scale$race=="T"), REML=FALSE,
family="gaussian")))
Anova(lm(lmer(bio.veg ~ sex + (1|pop/fam),
data=subset(mydat.scale,mydat.scale$race=="N"), REML=FALSE,
family="gaussian")))
#TX F:88.649 P:<2.2e-16
#NC F:73.007 P:3.743e-16

#Overall, more of an effect of sex in the NC race for: flowering time, width, inflorescence number. More of an effect of sex in the TX race for: length of the largest leaf, inflorescence height, inflorescence number (this one is fairly small), and vegetative biomass. Stem diameter was approximately equal in the amount of sexual dimorphism between SCRs.

### Plotting the grand means of phenotypic traits differentiated by SCR
#There were significant differences between the races for: flowering time, width, and inflorescence number

stderr <- function(x) sqrt(var(x)/length(x))
Supplementary Tables

#Call all of this for the figure for flowering time
nc.m=mean(na.omit(subset(repro$ff.time,repro$race=="N"))); tx.m=mean(na.omit(subset(repro$ff.time,repro$race=="T"))); nc.er=stderr(na.omit(subset(repro$ff.time,repro$race=="N"))); tx.er=stderr(na.omit(subset(repro$ff.time,repro$race=="T"))); plotmeans = data.frame(fftime=c(nc.m,tx.m),fferr=c(nc.er,tx.er),SCR=c("NC", "TX")); plot(plotmeans$fftime, las=1, xlab="", xlim=c(0.75,2.25), ylab="Flowering Time (days)", type="p", pch=16, cex=1.5, xaxt="n", ylim=c(43,47.5), family="Century Schoolbook", cex.axis=0.85); axis(1, at=c(1,2), labels=c("NC", "TX"), tick=TRUE, lwd=0, lwd.ticks=1, family="Century Schoolbook"); segments(1,nc.m+nc.er,1,nc.m-nc.er, lwd=2); segments(2,tx.m+tx.er,2,tx.m-tx.er,lwd=2)

#Call all of this for the figure for leaf width
nc.m=mean(na.omit(subset(repro$wid.2,mydat.scale$race=="N"))); tx.m=mean(na.omit(subset(repro$wid.2,mydat.scale$race=="T"))); nc.er=stderr(na.omit(subset(repro$wid.2,repro$race=="N"))); tx.er=stderr(na.omit(subset(repro$wid.2,repro$race=="T"))); plotmeans = data.frame(wid2=c(nc.m,tx.m),widerr=c(nc.er,tx.er),SCR=c("NC", "TX")); plot(plotmeans$wid2, las=1, xlab="", ylab="Leaf Width (mm)", type="p", pch=16, cex=1.5, xaxt="n", ylim=c(32,45), xlim=c(0.75,2.25), family="Century Schoolbook", cex.axis=0.85); axis(1, at=c(1,2), labels=c("NC", "TX"), tick=TRUE, lwd=0, lwd.ticks=1, family="Century Schoolbook"); segments(1,nc.m+nc.er,1,nc.m-nc.er, lwd=2); segments(2,tx.m+tx.er,2,tx.m-tx.er,lwd=2)

#Call all of this for the figure for inflorescence number
nc.m=mean(na.omit(subset(repro$inflo.num,repro$race=="N"))); tx.m=mean(na.omit(subset(repro$inflo.num,repro$race=="T"))); nc.er=stderr(na.omit(subset(repro$inflo.num,repro$race=="N")));
Supplementary Tables

tx.er = stderr(na.omit(subset(repro$inflo.num, repro$race == "T"))); plotmeans =
data.frame(inflonum = c(nc.m, tx.m), fferr = c(nc.er, tx.er), SCR = c("NC", "TX"));
plot(plotmeans$inflonum, las=1, xlab="", ylab="Number of Inflorescences",
family="Century Schoolbook", cex.axis=0.85); axis(1, at=c(1,2), labels=c("NC", "TX"),
tick=TRUE, lwd=0, lwd.ticks=1, family="Century Schoolbook");
segments(1, nc.m + nc.er, 1, nc.m - nc.er, lwd=2); segments(2, tx.m + tx.er, 2, tx.m -
tx.er, lwd=2)

################################################################
# Mean and standard error for all traits for each population
################################################################

head(import1)

std <- function(x) sd(x)/sqrt(length(x))
mypops =
data.frame(pop=levels(import1$pop), ff.mean=NA, ff.err=NA, len.mean=NA, len.err=NA,
wid.mean=NA, wid.err=NA, infl.onum.mean=NA, infl.onum.err=NA, infl.oH.mean=NA,
infloH.err=NA, stem.mean=NA, stem.err=NA, veg.mean=NA, veg.err=NA, bio.mean=NA,
bio.err=NA)

for (i in 1:27) {
    for (j in 1:8) {
        mypops[i,2] =
        mean(na.omit(subset(import1$ff.time, import1$pop == levels(import1$pop)[i])))
    }
}
Supplementary Tables

mypops[i,3] =
std(na.omit(subset(import1$len.2,import1$pop==levels(import1$pop)[i])))
mypops[i,4] =
mean(na.omit(subset(import1$len.2,import1$pop==levels(import1$pop)[i])))
mypops[i,5] =
std(na.omit(subset(import1$len.2,import1$pop==levels(import1$pop)[i])))
mypops[i,6] =
mean(na.omit(subset(import1$wid.2,import1$pop==levels(import1$pop)[i])))
mypops[i,7] =
std(na.omit(subset(import1$wid.2,import1$pop==levels(import1$pop)[i])))
mypops[i,8] =
mean(na.omit(subset(import1$inflo.num,import1$pop==levels(import1$pop)[i])))
mypops[i,9] =
std(na.omit(subset(import1$inflo.num,import1$pop==levels(import1$pop)[i])))
)
mypops[i,10] =
mean(na.omit(subset(import1$inflo.H,import1$pop==levels(import1$pop)[i])))
mypops[i,11] =
std(na.omit(subset(import1$inflo.H,import1$pop==levels(import1$pop)[i])))
mypops[i,12] =
mean(na.omit(subset(import1$stem.diam,import1$pop==levels(import1$pop)[i])))
mypops[i,13] =
std(na.omit(subset(import1$stem.diam,import1$pop==levels(import1$pop)[i])))
)
mypops[i,14] =
mean(na.omit(subset(import1$bio.veg,import1$pop==levels(import1$pop)[i]))
)
Supplementary Tables

```r
mypops[i,15] =
std(na.omit(subset(import1$bio.veg,import1$pop==levels(import1$pop)[i]))
mypops[i,16] =
mean(na.omit(subset(import1$bio.rep,import1$pop==levels(import1$pop)[i]))
mypops[i,17] =
std(na.omit(subset(import1$bio.rep,import1$pop==levels(import1$pop)[i])))
}
}
round(mypops[,2:17],1)
```