DISSERTATION

POPULATION GENETICS OF *CENTAUREA STOEBE* AND *C. DIFFUSA* IN THEIR NATIVE AND INTRODUCED RANGES

Submitted by

Robin Marrs

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, CO 80523

Spring 2007

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3266359

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 3266359

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

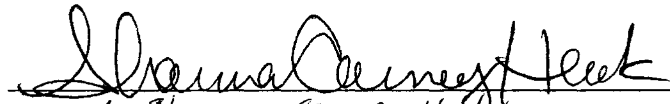
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

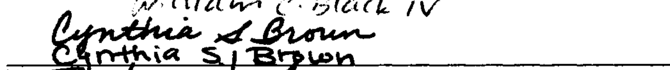
COLORADO STATE UNIVERSITY

February 1, 2007

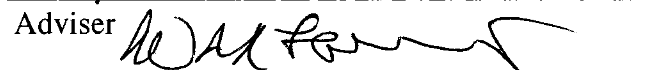
WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ROBIN MARRS ENTITLED "POPULATION GENETICS OF *CENTAUREA STOEBE* AND *C. DIFFUSA* IN THEIR NATIVE AND INTRODUCED RANGES" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work





Adviser



Department Head William Lauenroth

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT OF DISSERTATION

POPULATION GENETICS OF *CENTAUREA STOEBE* AND *C. DIFFUSA* IN THEIR NATIVE AND INTRODUCED RANGES

When species are introduced into new ranges, they can proliferate to such an extent that they dominate the new community. Research on such biological invasions has focused on ecological changes in the invasive species and the invaded communities. Recently, there has also been interest in the evolutionary changes associated with biological invasions. I explored the population genetics of two species native to Eurasia that have invaded North America (*Centaurea stoebe* and *C. diffusa*) using microsatellite loci. I compared six codominant microsatellite loci to dominant inter simple sequence repeat (ISSR) markers in tetraploid *C. stoebe*. Microsatellites and ISSRs gave similar patterns of allelic diversity and population genetic structuring, though the magnitude of structure estimates was higher in ISSRs. ISSRs were the more cost effective method, but they were also more subjective and tended to overestimate population structure and underestimate heterozygosity. I investigated genetic diversity and population structure in native and invasive *C. diffusa* using five microsatellite loci. Allelic diversity and expected heterozygosity did not differ between ranges, indicating that a bottleneck in population size is unlikely to have occurred. Populations were more strongly structured in the invasive range than in the native range of *C. diffusa,* and there was no pattern of isolation by distance in either range. This pattern is suggestive of panmixia in the native range and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of long-range dispersal within the invasive range. Assignment tests and phenetic trees indicate that *C. diffusa* may have been introduced into North America multiple times. I investigated genetic diversity and population structure in native and invasive *C. stoebe* (diploid subspecies *C. s. stoebe* and tetraploid subspecies *C. s. micranthos*) using six microsatellite loci. Genetic diversity was not reduced in the invasive range, suggesting that *C. stoebe* escaped genetic bottlenecks during colonization. Populations were genetically structured in both ranges of the species, and in the native range there was isolation by distance. A hypothesis of multiple introductions is supported by assignment tests. These results shed light on invasion pathways for these species and may form a base of knowledge for future hypotheses regarding the invasive success of spotted and diffuse knapweed.

Robin Alison Marrs
Graduate Degree Program in Ecology
Colorado State University
Fort Collins, CO 80523
Spring 2007

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGEMENTS

This work would not have been possible without the support of many people. I would like to thank Dr. Ruth Hufbauer for her endless creativity and patience over the past five and a half years, and for being a better adviser and friend than I could have ever hoped for. My labmates, Steven Rauth, Amy Blair, Dan Mackinnon, and Dr. Rebecca Kao, have cheerfully helped me through every draft revision, software question, logical fault, and crisis of confidence I have experienced. I greatly appreciate the help of Dr. René Sforza who collected my Eurasian samples, and Dr. Steven Bogdanowicz who created the library from which I designed my microsatellite primers. My committee, Dr. Shanna Carney, Dr. William Black, and Dr. Cynthia Brown, have been supportive and available when I needed their help. I would like to thank my husband, Brendan McGinn, and the Marrs family for their emotional support through all of my many academic ups and downs. Finally, this work would not have been possible without the financial support I received through grants to Dr. Ruth Hufbauer from the Colorado State University Agriculture Experiment Station and the USDA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

CHAPTER ONE: Background information on *Centaurea* 1
References for Chapter One 7

CHAPTER TWO: Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s.* stoebe and *C. s. micranthos* (S.G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae) 8
Abstract 9
Primer Note 9
References for Chapter Two.. 14
Table 2.1 16

CHAPTER THREE: Microsatellites vs. ISSRs: A comparison of two microsatellite based genetic markers in the North American invader, *Centaurea stoebe* (Asteraceae).. 18
Abstract... 19
Introduction... 20
Methods... 25
Results... 30
Discussion.. 33
References for Chapter Three.. 38
Table 3.1 46
Figure 3.1 47
Figure 3.2 48

CHAPTER FOUR: Population structuring is higher in the introduced range of *Centaurea diffusa* than in its native range, while genetic diversity is similar 49
Abstract 50
Introduction 51
Methods 55
Results 60
Discussion 64
References for Chapter Four 73
Table 4.1 82
Table 4.2 83
Table 4.3 84
Figure 4.1 85
Figure 4.2 87
Figure 4.3 88
Figure 4.4 89

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER FIVE: Evidence for multiple introductions of *Centaurea stoebe* (Asteraceae) to North America 90
Abstract 91
Introduction 92
Methods 95
Results 103
Discussion 108
References for Chapter Five 116
Table 5.1 123
Figure 5.1 125
Figure 5.2 127
Figure 5.3 128
Figure 5.4 129
Figure 5.5 131
Figure 5.6 133

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER ONE

BACKGROUND INFORMATION ON *CENTAUREA*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The genus *Centaurea* (Asteraceae) is native to Europe and comprises at least 300 species (Garcia-Jacas *et al.* 2006), many of which are split into subspecies. This genus is complex and there has been a history of disagreement among taxonomists regarding how it should be subdivided (discussed in Garcia-Jacas *et al.* 2006). Despite the confusion, the majority of species in the genus belong to a single monophyletic group, based on internal transcribed spacer (ITS) and *matK* phylogenetic trees (Garcia-Jacas *et al.* 2000, 2001). Called the *Jacea* group, these species also share a pollen type, a lateral hilum, and a tendency toward polyploidy (Garcia-Jacas *et al.* 2006). Most species in the *Jacea* group have a rather narrow distribution in either the western Mediterranean or the eastern Mediterranean and Eurasian regions, but there is also a clade of more widespread species that are distributed over much of Eurasia. These species are generally associated with agriculture, and some have significant economic impacts as weeds in their native range (Garcia-Jacas *et al.* 2006).

Many species of the genus *Centaurea* have been introduced into North America, often as seed contaminants and mixed in with soil used as ships' ballast (Watson & Renney 1974, J. Gerlach *pers. comm.*). Historical documents and shipping records indicate that seeds of Eurasian forbs, including *Centaurea*, were used to confirm the provenance of alfalfa seed shipped to North America (J. Gerlach *pers. comm.*). Thus, *Centaurea* species are likely to have been introduced to North America multiple times and founding population sizes may have been relatively large. Of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

unknown number of *Centaurea* species introduced to North America, 36 have naturalized. Thirteen of these species are problematic enough to be found on state noxious weed lists in the United States (http://plants.usda.gov). Invasive *Centaurea* species are now common throughout North America, especially in rangelands and natural areas. Of these, two species of concern in the Intermountain West region of the United States are the focus of my dissertation, and are described in greater detail below.

Spotted knapweed, *C. stoebe*, is a biennial to short-lived perennial forb that reproduces by seed (Watson & Renney 1974), and is self-incompatible (A. Blair *pers. comm.*). In Europe *C. stoebe* comprises two subspecies based on ploidy: a diploid subspecies found mainly in Western Europe (*C. stoebe stoebe* L.; 2x = 2n = 18), and a tetraploid subspecies found mainly in Eastern Europe and Western Asia (*C. stoebe micranthos* (S.G. Gmelin ex Gugler) Hayek; 2x = 4n = 36). North American invasive spotted knapweed is thought to be the tetraploid subspecies, *C. stoebe micranthos* (Ochsmann 2000, Watson & Renney 1974). *C. stoebe* was initially observed in North America in Victoria, British Columbia in 1893 (Groh 1944, Watson & Renney 1974).

Centaurea diffusa Lam. is an annual to triennial species, with the length of its life cycle dependent on resource availability (Thompson & Stout 1991). It is an obligate outcrosser that adopts a tumbling habit for seed dispersal (Watson & Renney 1974). Native range *C. diffusa* (Eurasia) comprises both diploid (2x = 2n = 18) and tetraploid (2x = 4n = 36) cytotypes. North American *C. diffusa* are diploid (A. Blair

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pers. comm.). The first *C. diffusa* specimen collected in North America was found in an alfalfa field in Bingen, Washington in 1907 (Watson & Renney 1974).

Centaurea stoebe and *C. diffusa* have been observed to form monocultures, outcompete native grasses (Kedzie-Webb *et al.* 2001), and invade natural areas in the absence of disturbance (Tyser & Key 1988) throughout North America. The competitive advantage they have over native species may be based on a lower sensitivity to soil nitrogen and phosphorous levels (Sudig *et al.* 2004), the ability to reach deeper soil water (Hill *et al.* 2006), and allelopathy (Bais *et al.* 2003, Callway & Aschehoug 2000, but see Blair *et al.* 2006). *C. stoebe* and *C. diffusa* are not generally utilized as forage by livestock, and their dense overstory can block access to more palatable grass species below (Watson & Renney 1974). Additionally, *C. diffusa* and *C. stoebe* respond to disturbance and can completely cover an overgrazed area after a single season (Watson & Renney 1974). Thirteen biological control agents have been introduced to combat the invasions of *C. diffusa* and *C. stoebe* (Müller-Scharer & Schroeder 1993, Lang *et al.* 2000). By investigating the genetic diversity and population structure of these species in their native and introduced ranges, I hope to shed some light on the invasion process and build a base of knowledge that will be useful for future research into the success of these species.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Bais, HP, R Vepachedu, S Gilroy, RM Callaway, JM Vivanco. 2003. Allelopathy and exotic plant invasion: From molecules and genes to species interactions. Science 301: 1377-1380.

Blair, AC, BD Hanson, GR Brunk, RA Marrs, P Westra, SJ Nissen, RA Hufbauer. 2005. New techniques and findings in the study of a candidate allelochemical implicated in invasion success. Ecology Letters 8: 1039-1047.

Callaway, RM, ET Aschehoug. 2000. Invasive plants versus their new and old neighbors: A mechanism for exotic invasion. Science 290: 521-523.

Garcia-Jacas, N, A Susanna, R Mozaffarian, R Ilarslan. 2000. The natural delimitation of *Centaurea* (Asteraceae: Cardueae): ITS sequence analysis of the *Centaurea jacea* group. Plant Systematics and Evolution 223: 185-199.

Garcia-Jacas, N, A Susanna, T Garnatje, R Vilatersana. 2001. Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. Annals of Botany 87: 503-515.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Garcia-Jacas, N, T Uysal, K Romaschenko, VN Suarez-Santiago, K Ertugrul, A Susanna. 2006. *Centaurea* revisited: A molecular survey of the *Jacea* group. Annals of Botany 98: 741-753.

Groh, H. 1944. Canadian Weed Survey. 2nd Annual Report of the Canadian Department of Agriculture. 74 pp.

Hill, JP, MJ Geronimo, JM Wraith, BE Olson, MB Swan. 2006. Advantages in water relations contribute to greater photosynthesis in *Centaurea maculosa* compared with established grasses. International Journal of Plant Sciences 167(2): 269-277.

Lang, RF, RD Richard, PE Parker, L Wendel. 2000. Release and establishment of diffuse and spotted knapweed biocontrol agents by USDA, APHIS, PPQ in the United States. Pan-Pacific Entomologist 76(4): 197-218.

Müller-Scharer, H, D Schroeder. 1993. The biological-control of *Centaurea* spp in North America – Do insects solve the problem. Pesticide Science 37(4): 343-353.

Ochsmann, J. 2000. Morphologische und molekularsystematische Untersuchungen an der Centaurea stoebe L.-Gruppe (Asteraceae-Cardueae) in Europa. PhD Dissertation. Ed. J. Cramer, Stuttgart, Georg-August University, Göttingen.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sudig, KN, KD LeJeune, TR Seastedt. 2004. Competitive impacts and responses of an invasive weed: dependencies on nitrogen and phosphorus availability. Oecologia 141: 526-535.

Thompson, DJ, DG Stout. 1991. Duration of the juvenile period in *C. diffusa* (*Centaurea diffusa*). Canadian Journal of Botany 69: 368-371.

Watson, AK, AJ Renney. 1974. The biology of Canadian weeds. *Centaurea diffusa* and *C. maculosa*. Canadian Journal of Plant Science 54: 687-701.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER TWO

NINE POLYMORPHIC MICROSATELLITE MARKERS IN *CENTAUREA STOEBE* L. (SUBSPECIES *C .S. STOEBE* AND *C. S. MICRANTHOS* (S.G. GMELIN EX GUGLER) HAYEK) AND *C. DIFFUSA* LAM. (ASTERACEAE)[1]

[1]This chapter has been published as: Marrs, RA, RA Hufbauer, SJ Bogdanowicz, R Sforza. 2006. Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae). Molecular Ecology Notes 6(3): 837-840.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Centaurea stoebe (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* are Eurasian plant species that have invaded much of North America. We isolated seven microsatellite loci from *C. stoebe* and two loci from *C. diffusa.* All loci described here amplify in both species, and have between six and 25 alleles each. These markers will be useful in examining population structure and addressing questions regarding these invasions.

PRIMER NOTE

Spotted knapweed, *Centaurea stoebe* L. (including subspecies *C. s. stoebe* (diploid) and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek (tetraploid), synonyms: *C. maculosa* L. and *C. biebersteinii* DC.) and diffuse knapweed, *C. diffusa* Lam. (Asteraceae) are two of the most important and pernicious rangeland weeds plaguing the Western United States (Sheley *et al.* 1998). In their invasive range these species can form monocultures (Sheley & Larson 1996), displace native grass species (Kedzie-Webb *et al.* 2001), and are able to invade natural systems in the absence of human disturbance (Tyser & Key 1988). Both species were introduced inadvertently to North America from Eurasia in the late 1800s or early 1900s,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

possibly multiple times each (Watson & Renney 1974). In the native range tetraploid and diploid forms of both species are spatially isolated; tetraploids are found in the eastern Eurasian and Asian parts of their range, while diploids occur primarily in the western European portion of the range (Ochsmann 2000). Morphological intermediates between *C. stoebe* and *C. diffusa* were described as a hybrid species, *C.* x *psammogena*, in the native range (Gáyer 1909), and similar intermediates have been observed in the introduced range (Watson & Renney 1974, Ochsmann 2000). In light of the economic and ecological importance of these species, as well as the interesting evolutionary questions raised by the potential for inter-specific crossing in these taxa, we have developed nine polymorphic microsatellite markers that amplify in both *C. stoebe* and *C. diffusa*.

Microsatellite markers were developed from mixed genomic DNA from several introduced-range sources. Genomic DNA from *C. s. micranthos* collected in Montana and Colorado was used to develop markers CM26, 21CM36, 38CM22, 42CM27, CM15, 25CM6, and CM17. Markers CD37 and CD9 were developed from mixed *C. diffusa* genomic DNA extracted from plants found in Montana and Washington.

Microsatellite cloning and sequencing was performed at the Evolutionary Genetics Core Facility of Cornell University following a modification of the protocol of Hamilton *et al.* (1999). The genomic DNA was digested with *Alu*I and *Hae*III, a SNX linker was ligated, and the DNA was enriched for repeats by hybridization to 3′ biotinylated (GT)15 and (CT)15 oligonucleotides followed by magnetic capture with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

streptavidin-coated magnetic beads. Enriched fragments were made double-stranded by polymerase chain reaction (PCR), digested with *Nhe*I, and cloned into *Xba*I-digested, dephosphorylated pUC 19. Colonies were grown on Luria-Bertani plates with ampicillin, and replicas were transferred to nylon membranes and probed with radiolabelled (GT)15 and (CT)15. Positive colonies were sequenced with a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) and a universal M13 primer that flanks the cloning site. Reactions cycled at 95°C for 50 seconds, 50°C for 20 seconds, and 60°C for 4 minutes (25 cycles total) then were analyzed on an ABI 377 Automated Sequencer.

Primer pairs were designed for regions flanking the microsatellites using the program Primer Select (DNAStar). Forward primers had universal M13(-21) tails added to their 5′ ends following the protocol of Schuelke (2000). These primer pairs were tested for amplification with genomic DNA extracted from both species, *Centaurea stoebe* (subspecies *C. s. stoebe* and *C. s. micranthos*) and *C. diffusa*. These DNAs were isolated with Plant Mini-Kits (Qiagen). PCR amplifications were performed with 10 μl total reaction volumes of 1× PCR Buffer (20 mM Tris-HCL, pH 8.4, 50 mM KCl), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 U *Taq* polymerase (Novagen), 0.1 μL *Taq*Start antibody (Clontech), 0.4 pmol of forward primer with M13(-21) tail, 1.8 pmol each of reverse primer and fluorescently labeled (6FAM, NED, or VIC) M13(-21) universal primer, and 1μL of extracted genomic DNA. Samples were amplified using a Mastercycler or Mastercycler Gradient thermal cycler (Applied Biosystems) with an initial denaturation step of 94°C for 1 minute, then 36 cycles of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

50 seconds at 94°C, 1 minute at annealing temperature (see Table 2.1), and 1 minute at 72°C for polymerase extension, followed by a final extension at 72°C for 10 minutes. PCR products were stored at -20°C until genotyping. For this initial marker development test we genotyped 229 individuals from several sample locations in their native and introduced ranges.

Prior to genotyping, PCR products were multiplexed in two batches according to the scheme given in Table 2.1. Four (primer set 2) or five (primer set 1) of these loci can be genotyped in a single sequencer run without overlap of similarly fluorescently labeled alleles. Genotyping was performed on an ABI 3100-Avant capillary sequencer (Applied Biosystems) at Colorado State University. Reference DNA samples were run with each genotyping plate. Files were analyzed with GeneMapper version 3.0 (Applied Biosystems).

The ploidy of our samples was unknown, and we used microsatellite banding patterns to infer ploidy. If a sample location contained an individual that showed three or more alleles at any of the nine loci, all individuals in that sample were analyzed as tetraploids. This assumes uniform ploidy within populations, a hypothesis that has not yet been tested. All *Centaurea diffusa* sample locations contained individuals with three or more alleles, so all *C. diffusa* were analyzed as tetraploids. Samples from two *C. stoebe* locations in Western Europe appeared to be diploid (*C. s. stoebe*) while the rest appeared to be tetraploid (*C. s. micranthos*).

The number of alleles at each locus ranged from six to 25. Observed (H_O) and expected (H_E) heterozygosities at each locus were calculated using Arlequin version

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.000 (Schneider *et al.* 2000) for diploid populations (Table 2.1). Observed heterozygosity was calculated by hand for tetraploids, and expected heterozygosity was calculated using SPAGeDi version 1.2 (Hardy & Vekemans 2002) (Table 2.1). Note that some of the assumptions of Hardy-Weinberg equilibrium may be violated in these invasive species, so H_E may not be as useful a metric for measuring allelic diversity as H_o. Observed heterozygosity was generally highest in *C.s. micranthos* (tetraploid), followed closely by tetraploid *C. diffusa*. *Centaurea stoebe stoebe* (diploid) had the lowest observed heterozygosity at each locus, which is to be expected as it has two possible alleles, while the tetraploids have four possible alleles and many more ways of being heterozygous.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Gáyer, G. 1909. Vier neue Centaureen der Flora von Ungarn. Magyar Botanikai Lapok: 58-61.

Hamilton, MB, EL Pincus, A Di Fiore, RC Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27: 500-507.

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

Kedzie-Webb, SA, RL Sheley, JJ Borkowski, JS Jacobs. 2001. Relationships between *Centaurea maculosa* and indigenous plant assemblages. Western North American Naturalist 61(1): 43-49.

Ochsmann, J. 2000. Morphologische und molekularsystematische Untersuchungen an der Centaurea stoebe L.-Gruppe (Asteraceae-Cardueae) in Europa. PhD Dissertation. Ed. J. Cramer, Stuttgart, Georg-August University, Göttingen.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Schneider, S, D Roessli, L Excoffier. 2000. Arlequin version 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sheley, RL, JS Jacobs, MF Carpinelli. 1998. Distribution, biology, and management of diffuse knapweed (*Centaurea diffusa*) and spotted knapweed (*Centaurea maculosa*). Weed Technology 12: 353-362.

Sheley, RL, LL Larson.1996. Emergence date effects on resource partitioning between diffuse knapweed seedlings. Journal of Range Management 50: 263-267.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments: A poor man's approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology18: 233-234.

Tyser, RW, CH Key. 1988. Spotted knapweed in natural area fescue grasslands: An ecological assessment. Northwest Science 62(4): 151-159.

Watson, AK, AJ Renney. 1974. The biology of Canadian weeds. 6. *Centaurea diffusa* and *C. maculosa*. Canadian Journal of Plant Science 54: 687-701.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.1 (reformatted for dissertation from publication):
Marker information for *Centaurea stoebe* and *C. diffusa*; locus names, Genbank accession numbers, forward (F) and reverse (R) primer sequences (* indicates location of M13(-21) tail for attachment of fluorescent dye to forward primer, sequence = TGTAAAACGACGGCCAGT), annealing temperatures (N_m), fluorescent dye, repeat motifs, and size ranges of PCR products.

Locus	Genbank	Primer Sequence (5′ → 3′)	N_m	Dye	Repeat Motif	Size Range
Primer set 1:						
CM26	DQ356425	F:* GAAGGGCTACGAGGGTGTTC	55°C	VIC	(TG)9T(TA)3	133-172
		R: GAAGTGTTGTGCATTTCAATCTATT				
CD37	DQ356426	F:* AGGTGCACTTTCCTGTTCAAC	58°C	NED	(CA)9	154-176
		R: CAACCCAATAAGATTACTTCCACTTC				
21CM36	DQ356419	F:* GCTATTAACAACTCCAAAATGAACAG	45°C	6FAM	(CA)6(TA)5(TG)16	187-244
		R:CCTGCTCCAACAAGTTTCCTC				
38CM22	DQ356427	F:* GGCTACATTAAGCTTATCCATTC	55°C	VIC	(GA)12AA(GA)8	289-344
		R:CTCGCATGTTATCCTCCCTC				
CD9	DQ356418	F:* GGTCCCCATACTTTCAAGCTAATAAC	55°C	NED	(CA)17	301-325
		R:ATGCTTCCCTTCTCAATGTTTTCTCT				
Primer set 2:						
42CM27	DQ356422	F:* TGGGATATTCGTTGGTTTAGTTTT	58°C	6FAM	(TG)14	155-175
		R:CCTCCCACTCCCGTTTGAC				
CM15	DQ356424	F:* GGAGGGCATGGGATTAAAGAGAT	55°C	VIC	(GT)9	185-220
		R:TGGATGCATCGGTCTGGAAATA				
25CM6	DQ356420	F:* ATGGGACATAAGATCCACAACAG	45°C	NED	(CA)9	255-272
		R:TAATTCAGCATTCAAAAATTTAGAAGAC				
CM17	DQ356421	F:*TACTTGGGCTTTTCGCTAATGAT	60°C	VIC	(AC)9	379-430
		R:ACAAACGTGTTCCAGCAGCA				

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2.1, continued:
Number of alleles observed and observed (H_O) and expected (H_E) heterozygosities for each locus, calculated by species and ploidy.

Locus	# of alleles	2n *C. s. stoebe*		4n *C. s. micranthos*		4n *C. diffusa*	
		H_O	H_E	H_O	H_E	H_O	H_E
Primer set 1:							
CM26	22	0.455	0.510	0.815	0.852	0.737	0.829
CD37	7	monomorphic		0.646	0.703	0.301	0.652
21CM36	6	0.250	0.228	0.536	0.488	0.565	0.423
38CM22	25	0.000	0.699	0.291	0.900	0.436	0.802
CD9	9	0.100	0.700	0.718	0.863	0.295	0.808
Primer set 2:							
42CM27	11	0.205	0.298	0.718	0.753	0.600	0.733
CM15	10	0.263	0.689	0.525	0.729	0.283	0.730
25CM6	7	0.190	0.251	0.200	0.824	0.458	0.864
CM17	16	0.063	0.502	0.398	0.782	0.706	0.77

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER THREE

MICROSATELLITES VS. ISSRS: A COMPARISON OF TWO MICROSATELLITE BASED GENETIC MARKERS IN THE NORTH AMERICAN INVADER, *CENTAUREA STOEBE* (ASTERACEAE)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Neutrally evolving markers have become an important tool in the study of biological invasions, and they fall into two main categories, codominant and dominant. Microsatellites are codominant, showing each separate allele present in an individual. Microsatellites are expensive to develop, but are generally thought to be worth the effort because interpretation is less subjective. This is true particularly for diploid species because heterozygotes can be identified. In dominant markers, such as inter-simple sequence repeats (ISSRs), alleles are either present or absent, and heterozygotes are indistinguishable from individuals homozygous for an allele. Dominant markers tend to be less expensive and time consuming to develop than codominant markers, but they are also less reproducible and band scoring may be more subjective. Here, I compare microsatellites and ISSRs on samples of the invasive tetraploid plant species *Centaurea stoebe* (Asteraceae) with measures of genetic diversity, population structure, and phenetic tree congruence. Because the *C. stoebe* samples were tetraploid, I was unable to perform many of the analyses standard to codominant marker studies in diploids. I found that both marker systems showed similar patterns of genetic diversity and number of private alleles between the native Eurasian and invasive North American ranges of *C. stoebe*. Expected heterozygosity was always higher from microsatellites than from ISSRs, while estimates of population structure derived from microsatellite analyses (F_{ST}) were lower than those derived from ISSR analyses (θ_P). Phenetic trees constructed from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

microsatellite and ISSR data were poorly resolved and did not share any sister groups. To compare the efficiency of each marker system I calculated the effective multiplex ratio and the marker index (Powell *et al.* 1996). Both were higher using ISSRs than they were in microsatellites, indicating that ISSRs are a more cost effective marker. Particularly for tetraploid species like *C. stoebe*, where the full utility of microsatellites is not realized, ISSRs may be the best choice unless microsatellites have been developed previously.

INTRODUCTION

Population genetic analysis has become an important tool in the study of biological invasions (e.g., Amsellam *et al.* 2000, DeWalt & Hamrick 2004, Durka *et al.* 2005, Genton *et al.* 2005, Grapputo *et al.* 2005). Many of the questions that arise when a species is transported to a new range are best addressed with the evolutionary perspective provided by molecular markers. Some of these questions can be integral to understanding and controlling an invasion. Invasion ecologists use population genetic approaches to determine whether there have been multiple introductions of invasive species into their new ranges (Durka *et al.* 2005, Genton *et al.* 2005), whether there has been a reduction in genetic diversity in the new range (Amsellam *et al.* 2000, DeWalt & Hamrick 2004), or if high diversity is retained during the invasion process (Durka *et al.* 2005, Genton *et al.* 2005, Zenger *et al.* 2003).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Molecular markers are also answering questions of how selection may affect introduced populations (Kliber *et al.* 2005, Maron *et al.* 2004), and elucidating the provenances of invasive species and histories of their spread (Amsellam *et al.* 2000, Durka *et al.* 2005, Williams *et al.* 2005).

Many different types of molecular markers have been used to examine population diversity and differentiation in invasive species in their native and introduced ranges. Allozymes and isozymes (DeWalt & Hamrick 2004, Neuffer & Hurka 1999, Squirrell *et al.* 2001) and chloroplast and mitochondrial DNA sequences (Grapputo *et al.* 2005, Squirrell *et al.* 2001, Williams *et al.* 2005) are generally slowly evolving and useful for determining relationships between species, subspecies, or highly differentiated populations of the same species. However, they do not evolve quickly enough to address many questions related to recent introductions. There are several options of more quickly evolving, and therefore more polymorphic markers that provide greater resolution for evaluating intra-specific relationships, including those resulting from biological invasions. These markers fall into two basic categories, codominant and dominant. Codominant markers show each allele present at a locus in an individual, which provides a dataset amenable to powerful statistical analyses (particularly for diploids, as discussed below). Dominant markers show alleles as either present or absent (dominant or recessive), and statistical approaches to analyzing dominant markers can be limited. Microsatellites, also called simple sequence repeat markers (Durka *et al.* 2005, Genton *et al.* 2005, Williams *et al.* 2005, Zenger *et al.* 2003), are the standard quickly-evolving codominant marker used in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

research on the population genetics of biological invasions. The dominant markers commonly used include amplified fragment length polymorphisms (AFLP: Amsellam *et al.* 2000, Grapputo *et al.* 2005), random amplified polymorphic DNA (RAPD: Kliber & Eckert 2005), and inter simple sequence repeat (ISSR: Li & Ye 2006, Meekins *et al.* 2001, Poulin *et al.* 2005, Sun *et al.* 2005). In this paper, I compare microsatellites with ISSRs in measuring genetic diversity and population structure of an invasive tetraploid plant.

Microsatellite markers are specifically amplified fragments containing a short DNA motif that is repeated five or more times. They are also known as simple sequence repeats or SSRs. Primers are designed for the flanking regions around a microsatellite, and PCR is used to amplify the fragment. Microsatellite development comprises many steps, including cloning, sequencing, primer design and testing, and finally genotyping individuals with multiple primer pairs to obtain multilocus genotypes. For all of these reasons, it tends to be an expensive and time-consuming process. Microsatellites vary in the number of times a motif is repeated, and therefore in length, due to a high rate of mutation by slipped-strand mispairing (Schlotterer & Tautz 1992). The length of a fragment reveals the number of repeat motifs present. In diploid organisms each allele is visible, allowing one to differentiate between homozygotes and heterozygotes, unless null alleles are present (null alleles occur when there are mutations in genomic DNA that prevent PCR amplification of the fragment, and lead to underestimation of heterozygotes). This increases the information provided by a genetic dataset by reducing the number of assumptions that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

must be made in subsequent analyses. Additionally, population genetic theory is best developed for data from codominant markers on diploid organisms, so there are more software packages available for ready analysis of this type of data. This makes microsatellites extremely useful markers, and often it is clearly worth the expense and time involved in development. However, in polyploid organisms, full genotypes can only be determined for individuals showing one allele (e.g. a homozygote for allele *A*) or the total possible number of alleles (e.g. a full heterozygote with alleles *A*, *B*, *C*, and *D*). Full genotypes cannot be determined for individuals showing any number of alleles between these values, because it is unknown which allele(s) are present in multiple copies. For example, when two alleles, *A* and *B*, are observed in a tetraploid, the full genotype could be *AAAB*, *AABB*, or *ABBB*. Given this disadvantage, it is less clear whether the expense of microsatellites is warranted in studies of polyploids.

ISSRs are hypervariable markers based on microsatellites, but rather than genotyping microsatellites themselves, primers are designed to amplify the regions between microsatellite loci. This produces a genetic fingerprint from single primer polymerase chain reactions (Wolfe *et al.* 1998, Mort *et al.* 2003, Hettwer & Gerowitt 2004). Primers are oligonucleotides of a di- or tri nucleotide microsatellite motif with a short anchoring sequence attached to the 5′ or 3′ end (Gupta *et al.* 1994, Zeitkiewicz *et al.* 1994). Regions of DNA between inverted repeats matching the primer sequence are amplified by PCR (Wolfe *et al.* 1998). Variation in ISSR loci is due to differences among individuals in whether a particular region amplifies or not. Variable amplification is caused either by insertions or deletions in the primer binding sites

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that prevent primer annealing, or by insertions or deletions between binding sites that make the fragment too long to be amplified or too short to be genotyped. ISSRs can be a relatively cheap and quick method of DNA fingerprinting, as no prior knowledge of the study organism's genome is needed, and many primers can be quickly tested to determine which work in the species of interest. Once primers are chosen, a single PCR reaction on an individual provides a genotype of many loci, reducing reagent use and researcher hours. However, ISSRs can be somewhat more difficult and subjective to score than microsatellites due to the presence of weak, unreproducible bands, especially when visualized on agarose gels. ISSR banding patterns are more sensitive than microsatellites to laboratory conditions, and it is therefore more difficult to obtain consistent results between laboratories. Additionally, statistical analyses are not as powerful as for codominant markers, both because they are less well developed and because the inability to distinguish heterozygotes requires that assumptions must be made to run the statistical packages that are available. This effectively increases the confidence intervals of parameter estimates.

Few studies exist that compare the performance of different quickly evolving markers in polyploid non-crop species. Plant breeders and horticulturalists occasionally compare neutral marker systems, often in polyploid crops (Goulão & Oliveira 2001, McGregor *et al.* 2000, Rakoczy-Trojanowska & Bolibok 2004, Saini *et al.* 2004). While their research goals are similar to those of invasion ecologists (elucidating crossing history, determining relationships between varieties, and calculating degree of differentiation of cultivars, for example), the organisms they

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

study are different than invasives and other wild species. As a result of domestication, crop and horticultural species are likely to be less genetically variable than non-domesticated species. Here I compare the results of two neutral markers, one codominant (microsatellites) and one dominant (ISSRs), when used on the same DNA samples from a tetraploid, non-domesticated plant species, *Centaurea stoebe* (Asteraceae), from its native Eurasian range and its invasive North American range. My objectives are twofold: 1) To compare polymorphism and diversity between the markers, and 2) To compare basic statistics of population structure between the markers.

METHODS

Study species

Centaurea stoebe (S. G. Gmelin ex Gugler) Hayek (synonyms : *C. maculosa, C. biebersteinii,* Asteraceae), commonly called spotted knapweed, is a perennial forb native to Eurasia. It was introduced to North America near the turn of the 20th century, where it has expanded its range (Watson & Renney 1974). In its invasive range *C. stoebe* can form monocultures, outcompete native grasses (Kedzie-Webb *et al.* 2001), and invade natural areas in the absence of disturbance (Tyser & Key 1988).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sample collection and preparation

Centaurea stoebe seeds were collected at eight North American sites and seven Eurasian sites (Table 3.1). Approximately ten different individuals were sampled at each site, and seeds were kept separate by maternal plant. Seeds were then sown separately and seedlings were grown in the Weed Research greenhouse at Colorado State University. Seeds from some sampling locations did not germinate well; these sites are represented by fewer than ten individuals (Table 3.1). At week four, leaf tissue was collected from each individual and total DNA was extracted using DNeasy Plant Mini Kits (Qiagen). Genomic DNA working stocks were stored at –20°C. A total of 138 DNA accessions from 15 sampling sites were included in this study.

Microsatellite analysis

I genotyped each DNA accession using six microsatellite loci (21CM36, 42CM27, 38CM22, CM15, CM26, and CM17) from Marrs *et al.* (2006). Polymerase chain reactions, allele visualization, and band scoring were performed according to Marrs *et al.* (2006). *Centaurea stoebe* comprises diploid and tetraploid subspecies, therefore a single individual could show as few as one and as many as four alleles at any given locus. Given the collection locations, I anticipated that individuals were tetraploid. In individuals showing less than four alleles, therefore, I coded a genetic

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phenotype for each individual using the following method: each allele observed was recorded once, and the unknown positions were coded as missing data. For example, if alleles *A* and *B* were observed, the recorded genotype was *AB*??, where "?" represents missing data.

I examined genetic diversity by calculating the percent polymorphic loci at each sampling site, expected heterozygosity averaged over all loci at each site, and numbers of alleles unique to Eurasia and North America using SPAGeDi 1.2, a program that computes statistics and permutation tests describing relatedness and differentiation between populations for organisms of any ploidy level (Hardy & Vekemans 2002). SPAGeDi 1.2 assumes polysomic inheritance, as seen in autopolyploids. As *C. stoebe* is a likely allopolyploid, results should be interpreted with some caution. I determined the level of population structure at several scales of interest by calculating F_{ST} over all populations, within Eurasia, within North America, and between North America and Eurasia with SPAGeDi 1.2 (Hardy & Vekemans 2002). To determine whether differences between these estimates were significant, and to simulate an AMOVA (analysis of molecular variance, which has not yet been developed for tetraploid data), I ran 20,000 permutations to obtain 95 percent confidence intervals around the estimates of F_{ST}. To visualize relationships between individuals, I used the program Populations (Langella 1999) to construct a UPGMA tree using Nei's standard genetic distance (1972).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ISSR analysis

Each DNA accession was amplified twice (for replication) with each of three different ISSR primers, $(CA)_8$-RG, $(CT)_8$-RG, and $(CAC)_5$-AG (Wolfe *et al.* 1998). I ran 10 µl PCR reactions containing 1.0 µl template DNA (ranging from one to four nanomoles DNA per microliter; or water, in negative controls), 7.02 µl HPLC-grade ultrapure water, 1.0 µl 10× PCR buffer, 0.4 µl $MgCl_2$, 0.08 µl dNTP mix, 0.5 U *Taq* polymerase (Novagen), 0.1 µl *Taq*Start (Clontech), and 0.3 pmol primer. Thermal cycling was performed in an Eppendorf Mastercycler set to the following program: 1.5 minutes at 94°C; 35 cycles of 40 seconds at 94°C, 45 seconds at 45°C, 1.5 minutes at 72°C; 40 seconds at 94°C, 45 seconds at 45°C, 5 minutes at 72°C (Wolfe *et al.* 1998). Samples were run in batches of 22 with a negative control (water) and a positive control (standard gDNA).

PCR products were separated and visualized on 1.5 percent agarose gels containing ethidium bromide. To determine fragment size, 1Kb Plus ladder (Bioline) was loaded into the first and last lanes of each gel. Gels were electrophoresed at 90 volts for approximately three hours until the bromophenol blue band of the loading dye reached 10 centimeters, then digitally photographed using BioMax ID image capture software (Kodak). Bright bands were detected using the algorithms in BioMax1D software (Kodak). Genotypes were scored by hand using the 1KB Plus size standards on each gel to determine the size of each band. Singleton loci (bands occurring only once) and weak, unreproducible bands were not included in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dataset. Primer $(CA)_8$-RG produced 22 scoreable loci, $(CT)_8$-RG gave 24 loci, and $(CAC)_5$-AG gave 25 loci. All three datasets were combined to produce a 71-locus genotype for each individual DNA accession.

I computed several estimates of genetic diversity in the ISSR dataset, including average unbiased heterozygosity (using estimated allele frequencies), percent polymorphism for each site, and the number of private alleles (loci that only showed bands in Eurasia or in North America; Neel 1973, Slatkin 1985) using Tools for Population Genetic Analysis (TFPGA, Miller 1998). I used the square root of the frequency of the recessive genotype in TFPGA (Miller 1998) to estimate allele frequencies, as ISSRs are dominant and the number of heterozygotes is unknown. This assumes populations are in Hardy-Weinberg equilibrium. Population structure was measured by calculating *F*-statistics (θ_P), pairwise genetic distances, and exact tests for population differentiation with the ISSR dataset using TFPGA. I ran an AMOVA in GenAlEx 6 (Peakall & Smouse 2006) to determine how much variation in the dataset was due to within versus among population diversity. I examined relationships between individuals by constructing a UPGMA tree using Nei's standard genetic distance (1972) in NTSYSpc (Exeter Software).

Comparison of data obtained per unit effort for ISSR and microsatellite analyses

I compared the amount of usable genetic data collected per unit effort of the two marker systems using the indices of Powell *et al.* (1996). The effective multiplex

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ratio (EMR) is the number of polymorphic products derived from a single amplification reaction, and is a rough estimate of the researcher effort involved in obtaining a multilocus genotype. The marker index (MI) is the product of the EMR and expected heterozygosity, so it takes into account both researcher effort and the quality (measured as heterozygosity) of the data.

RESULTS

Genetic diversity

Overall, I observed a total of 112 microsatellite alleles, 88 of which occurred in North America, and 85 of which occurred in Eurasia. The native range of the species, Eurasia, had 24 private alleles (alleles that did not occur in the invasive range), while North America had 27 private alleles. The six microsatellite loci I assayed were polymorphic at almost all of the sampling locations. Only the site with the lowest sample size, Worms, Germany, had any loci that were not polymorphic; Worms had three monomorphic loci. All 71 of the ISSR alleles were polymorphic; there were no ISSR bands that were fixed (shared by all individuals in our study). Percent polymorphism per population ranged from a low of 21.1 percent in the Worms, Germany individuals (N = 6), to a high of 40.8 percent in the Tecuci,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Romania site (N = 10) (Table 3.1). Eight alleles were private to Eurasian populations, while seven were private to North American populations.

Microsatellites always gave a higher estimate of heterozygosity than ISSRs (Table 2.1). Indeed, the lowest microsatellite-based expected heterozygosity (Worms, Germany, $H_E = 0.2995$) was higher than the highest ISSR-based estimate of unbiased heterozygosity (Tecuci, Romania, $H = 0.1928$) (Table 3.1).

Genetic structuring

F-statistics were calculated for the microsatellite and ISSR datasets using F_{ST} and θ_P, respectively. All were found to be significantly different from zero, indicating significant population structure (Figure 3.1). For both microsatellites and ISSRs, the continent-level *F*-statistic was the lowest, while global, within Eurasia, and within North America estimates were not significantly different from each other (Figure 3.1).

Exact tests for population differentiation (Raymond & Rousset 1995) of the ISSR dataset showed 33 of 71 loci had significant ($P \leq 0.05$) overall differentiation, when corrected for multiple comparisons with the Bonferroni method. Population differentiation was significant over all loci as well ($\chi^2 = 889.17$, d.f. = 142, $P < 0.0001$). When corrected for multiple comparisons, however, individual pairwise comparisons showed only two sampling site by sampling site comparisons were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

significantly differentiated: Michigan vs. Highway 93, MT ($P = 0.0003$), and Michigan vs. Roman, Romania ($P = 0.0001$).

UPGMA tree comparison

Individual level UPGMA trees were constructed from the multilocus genotypes derived from each marker dataset. The microsatellite-based UPGMA tree (Figure 3.2a) did not show any patterns of individuals from the same population being near to each other on the tree. The UPGMA tree constructed from the ISSR dataset (Figure 3.2b) was more well-resolved, with individuals from the same sampling site grouping in clades in many cases.

Effective marker ratio and marker index

The effective marker ratio (EMR), or the number of polymorphic products from a single PCR reaction (Powell *et al.* 1996), was 1.66 for the microsatellites, and 23.67 for the ISSRs. The marker index, the product of EMR and the expected heterozygosity was 1.20 in the microsatellites, and 3.47 in the ISSRs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DISCUSSION

Allelic diversity and polymorphism higher in microsatellites

Expected heterozygosity and percent polymorphic loci are two statistics often used to gauge levels of genetic diversity. I found that estimates of expected heterozygosity and percent polymorphism were much higher in microsatellites than in ISSRs (Table 3.1). These results are in line with expectations, since codominant markers can have many possible alleles at each locus, and H_E can be as high as 1.0, while dominant markers only have two possible alleles ("band present" or "band absent") at each locus and therefore a maximum expected heterozygosity of 0.5 (Nybom 2004). The differences are dramatic in this case since *C. stoebe* is a tetraploid species and each individual can have up to four alleles at each microsatellite locus.

Private alleles, or alleles that are unique to a certain area, can be particularly useful in comparing diversity between invasive species' native and introduced ranges. Private alleles typically occur at low frequency, and rare alleles tend to be lost quickly by genetic drift (e.g. Luikart *et al.* 1998). Thus, if an introduced species has gone through a strong bottleneck in population size, we expect fewer private alleles in the invasive range of the species than in the native range. In *Centaurea stoebe* we did not see a reduction in the number of private alleles using either marker system. Results between markers are congruent, but the number of private alleles was much

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

greater in microsatellites. In our ISSR dataset, just 15 of 71 loci (21.1 percent) were private to one range or the other, while in the microsatellite dataset, 51 of 112 alleles (45.5 percent) were private to one of the ranges. From these data, it seems one is more likely to find rare or private alleles when using a codominant method, such as microsatellites, than by using a dominant method like ISSRs.

Patterns of population structure similar across markers

Overall patterns of population genetic structuring were consistent across markers (Figure 3.1). Population structure was found to be similar at the global level (among all sites), among Eurasian sites, and among North American sites, while continent level (EU versus NA) structure was very low in both species. It would seem that this pattern is robust, as it is supported by both marker systems.

Despite the similarity in pattern, estimates of the magnitude of structuring present varied greatly between markers (Figure 3.1). F_{ST} estimates derived from microsatellite data were always much lower than θ_P estimates derived from ISSR data. The disparity seen here between microsatellite- and ISSR-derived *F*-statistics is similar to patterns observed by other researchers comparing codominant and dominant markers (Maguire *et al.* 2002, Mariette *et al.* 2001). The reason for the disparity in among-population structure estimates of microsatellites and ISSRs probably has to do with the higher heterozygosity observed in microsatellites discussed above. Since codominant markers have higher within-individual and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

within-population diversity due to the higher number of alleles at each locus, corresponding estimates of among-population structure will be lower (Hedrick 1999). Size homoplasy in microsatellites may also play a role in reducing the amount of among-population structure observed, as different alleles are not necessarily identical by descent (e.g. a six repeat allele could be produced two ways: by the loss of a repeat from a seven repeat allele, or by the addition of a repeat to a five repeat allele), thus some genetic differences between populations may be hidden. However, Qian *et al.* (2001) found that ISSRs can overestimate the amount of differentiation between closely related populations and underestimate the variation between geographically distinct populations when compared to other dominant markers such as RAPD and AFLP, which could mean that other dominant markers may give different population structure results relative to microsatellites.

Interpopulation relationships (phenetic trees) not robust across markers

The lack of consistency between individual-based UPGMA trees constructed using the microsatellite and ISSR datasets is troubling. The ISSR tree shows individuals from the same sampling sites occurring in the same clades on many occasions, while the microsatellite-based tree does not. This could mean that variation relevant to the true relationships between sites simply does not exist in the microsatellite loci surveyed (Hufbauer 2004). Larger sample sizes or greater underlying genetic structure might lead to datasets with more congruence between

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

trees built from different marker systems (see Goulãu & Oliveira 2001 for an example with cultivated apple).

ISSRs produce more data per unit researcher effort

I found that both the effective multiplex ratio and the marker index (Powell *et al.* 1996) were higher for ISSRs than they were in microsatellites. These statistics quantify the amount of usable data one obtains per unit effort for each method, and our results here indicate that ISSRs are a more efficient DNA fingerprint to use in *C. stoebe*. This result is mainly based on the EMR, or the number of polymorphic products produced per polymerase chain reaction. Single ISSR reactions in this study resulted in more than 20 alleles, while only four alleles, at most, could be observed from a single microsatellite reaction. However, these statistics do not take into account that ISSRs tend to be less repeatable than microsatellites.

Conclusions

To determine which marker is more appropriate for any given study, a careful assessment of the research goals is in order. My results indicate that if genetic diversity estimates and relative numbers of private alleles are important, as when one wants to determine whether genetic drift via bottleneck or founder effect has occurred after introduction of a species to a new range, microsatellites are probably a better

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

option. However, if one is interested in determining relative levels of population structuring, for example in a species' native and introduced ranges, both markers seem to give similar patterns. The higher estimates of *F*-statistics given by ISSRs may be useful in systems where population structure is more difficult to detect, and the tendency of this marker to emphasize differences between closely related populations (Qian *et al.* 2001) could also be useful in looking at very fine scale structure. Qualitative patterns in relative levels of population structure were robust over markers. However, in this study I show that the magnitude of *F*-statistics calculated from the same individuals using two different marker systems varied enormously. UPGMA analyses with datasets from each marker were not well-resolved enough to lead to robust comparisons.

From a practical perspective, if microsatellite markers are not readily available, ISSRs may be the more time and cost effective marker to use for polyploid species such as *C. stoebe micranthos*. Codominance, the main benefit of microsatellites, is not fully realized when working with polyploid species, since complete genotypes are not visible and genetic phenotypes must be used instead. However, even in polyploid systems, microsatellites have an advantage over ISSRs in that they are more repeatable from one laboratory to another. If microsatellite markers are available, then they may be the best choice despite the added expense incurred by the need to run more PCR reactions to obtain a similar number of variable markers.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Amsellam, L, JL Noyer, T Le Bourgeois, M Hossaert-McKey. 2000. Comparison of genetic diversity in the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. Molecular Ecology 9: 443-455.

DeWalt, SJ, JL Hamrick. 2004. Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, *Clidemia hirta* (Melastomataceae). American Journal of Botany 91(8): 1155-1162.

Durka, W, O Bossdorf, D Prati, H Auge. 2005. Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. Molecular Ecology 14: 1697-1706.

Genton, BJ, A Shykoff, T Giraud. 2005. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. Molecular Ecology 14: 4275-4285.

Goulãu, L, CM Oliveira. 2001. Molecular characterization of cultivars of apple (*Malus* x *domestica* Borkh.) using microsatellite (SSR and ISSR) markers. Euphytica 122: 81-89.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Grapputo, A, S Boman, L Lindstrom, A Lyytinen, J Mappes. 2005. The voyage of an invasive species across continents: Genetic diversity of North American and European Colorado potato beetle populations. Molecular Ecology 14: 4207-4219.

Gupta, M, Y-S Chyi, J Romero-Severson, JL Owen. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theoretical and Applied Genetics 89: 998-1006.

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

Hedrick, PW. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313-318.

Hettwer, U, B Gerowitt. 2004. An investigation of genetic variation in *Circium arvense* field patches. Weed Research 44: 289-297.

Hufbauer, RA. 2004. Population genetics of invasions: Can we link neutral markers to management? Weed Technology 18: 1522-1527.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hufbauer, RA, SM Bogdanowicz, RG Harrison. 2004. The population genetics of a biological control introduction: mitochondrial DNA and microsatellite variation in native and introduced populations of *Aphidius ervi*, a parasitoid wasp. Molecular Ecology 13: 337-348.

Kedzie-Webb, SA, RL Sheley, JJ Borkowski, JS Jacobs. 2001. Relationships between *Centaurea maculosa* and indigenous plant assemblages. Western North American Naturalist 61(1): 43-49.

Kliber, A, CG Eckert. 2005. Interaction between founder effect and selection during biological invasion in an aquatic plant. Evolution 59(9): 1900-1913.

Langella, O. 1999. Populations 1.2.28 (12/5/2002): a population genetic software. CNRS UPR9034. Available at http://pge.cnrs-gif.fr/bioinfo/populations/index.php.

Li, J, WH Ye. 2006. Genetic diversity of alligator weed ecotypes is not the reason for their different responses to biological control. Aquatic Botany 85(2): 155-158.

Luikart, G, FW Allendorf, J-M Cornuet, WB Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. The Journal of Heredity 89(3): 238-247.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Maguire, TL, R Peakall, P Saenger. 2002. Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. Theoretical and Applied Genetics 104: 388-398.

Mariette, S, D Chagné, C Lézier. 2001. Genetic diversity within and among *Pinus pinaster* populations: comparisons between AFLP and microsatellite markers. Heredity 86: 469-479.

Maron, JL, M Vilà, R Bommarco, S Elmendorf, P Beardsley. 2004. Rapid evolution of an invasive plant. Ecological Monographs 74(2): 261-280.

Marrs, RA, RA Hufbauer, SJ Bogdanowicz, R Sforza. 2006. Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae). Molecular Ecology Notes 6: 897-899.

McGregor, CE, CA Lambert, MM Greyling, JH Louw, L Warnich. 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica 113: 135-144.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Meekins, JF, HE Ballard, BC McCarthy. 2001. Genetic variation and molecular biogeography of a North American invasive plant species (*Alliaria petiolata*, Brassicaceae). International Journal of Plant Sciences 162(1): 161-169.

Miller, MP. 1998. Tools for Population Genetic Analysis (TFPGA). University of Utah, USA. (http://bioweb.usu.edu/mpmbio/).

Mort, ME, DJ Crawford, A Santos-Guerra, J Francisco-Ortega, EJ Esselman, AD Wolfe. 2003. Relationships among the Macaronesian members of Tolpis (Asteraceae: Latuceae) based upon analyses of inter simple sequence repeat (ISSR) markers. Taxon 52: 511-518.

Neel, JV. 1973. Private genetic variants and frequency of mutation among South-American Indians. Proceedings of the National Academy of Sciences of the United States of America 70(12): 3311-3315.

Nei, M. 1972. Genetic distance between populations. American Naturalist 106(949): 283.

Neuffer, B, H Hurka. 1999. Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. Molecular Ecology 8: 1667-1681.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143-1155.

Peakall, R, PE Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Poulin, J, SG Weller, AK Sakai. 2005. Genetic diversity does not affect the invasiveness of fountain grass (*Pennisetum setaceum*) in Arizona, California, and Hawaii. Diversity and Distributions 11(3): 241-247.

Powell, W, M Morgante, C Andre, M Hanafey, J Vogel, S Tingey, A Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2(3): 225-238.

Qian, W, S Ge, D-Y Hong. 2001. Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. Theoretical and Applied Genetics 102: 440-449.

Rakoczy-Trojanowska, M, H Bolibok. 2004. Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. Cellular and Molecular Biology Letters 9: 221-238.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raymond, M, F Rousset.1995. An exact test for population differentiation. Evolution 49: 1280-1283

Saini, N, N Jain, S Jain, RK Jain. 2004. Assessment of genetic diversity within and among Basmati and non-Basmati rice varieties using AFLP, ISSR, and SSR markers. Euphytica 140: 133-146.

Schlotterer, C, D Tautz. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Research 20(2): 211-215.

Squirrell, J, PM Hollingsworth, RM Bateman, JH Dickson, MHS Light, M MacConaill, MC Tebbitt. 2001. Partitioning and diversity of nuclear and organelle markers in native and introduced populations of *Epipactis helleborine* (Orchidaceae). American Journal of Botany 88(8): 1409-1418.

Sun, JH, Z-C Li, DK Jewett, KO Britton, WH Ye, X-J Ge. 2005. Genetic diversity of *Pueraria lobata* (kudzu) and closely related taxa as revealed by inter-simple sequence repeat analysis. Weed Research 45(4): 255-260.

Tyser, RW, CH Key. 1988. Spotted knapweed in natural area fescue grasslands: An ecological assessment. Northwest Science 62(4): 151-159.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Watson, AK, AJ Renney. 1974. The biology of Canadian weeds. 6. *Centaurea diffusa* and *C. maculosa*. Canadian Journal of Plant Science 54: 687-701.

Williams, DA, WA Overholt, JP Cuda, CR Hughes. 2005. Chloroplast and microsatellite DNA diversities reveal the introduction history of Brazilian peppertree (*Schinus terebinthifolius*) in Florida. Molecular Ecology 14: 3643-3656.

Wolfe, AD, Q Xiang, SR Kephart. 1998. Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. Molecular Ecology 7: 1107-1125.

Zeitkiewicz, E, A Rafalski, D Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176-183.

Zenger, KR, BJ Richardson, AM Vachot-Griffin. 2003. A rapid population expansion retains genetic diversity within European rabbits in Australia. Molecular Ecology 12: 789-794.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3.1:
Sampling locations, number of individuals sampled (N), percent polymorphic loci for the ISSR dataset (P_I), percent polymorphic loci for the microsatellite dataset (P_M), average unbiased heterozygosity from the ISSR dataset (H), and average expected heterozygosity from the microsatellite dataset (H_E).

Location	N	P_I	P_M	H	H_E
Michigan, USA	9	39.4	100	0.1765	0.7191
Nevada, USA	9	36.6	100	0.1675	0.6856
Glacier, MT, USA	10	26.8	100	0.1194	0.7559
Idaho, USA	10	40.0	100	0.1420	0.8082
Washington, USA	10	33.8	100	0.1388	0.7776
Missoula, MT, USA	10	36.6	100	0.1731	0.8127
Highway 93, MT, USA	10	31.0	100	0.1347	0.7694
Butler Creek, MT, USA	10	26.8	100	0.1265	0.7675
Falticeni, Romania	10	32.4	100	0.1495	0.7722
Iasi, Romania	10	33.8	100	0.1527	0.6407
Tecuci, Romania	10	40.8	100	0.1908	0.8225
Roman, Romania	10	38.0	100	0.1742	0.7793
Worms, Germany	6	21.1	50	0.0914	0.2995
Wiener, Austria	7	31.0	100	0.1498	0.7820
Deutschlereuz, Austria	7	23.9	100	0.1129	0.6689

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

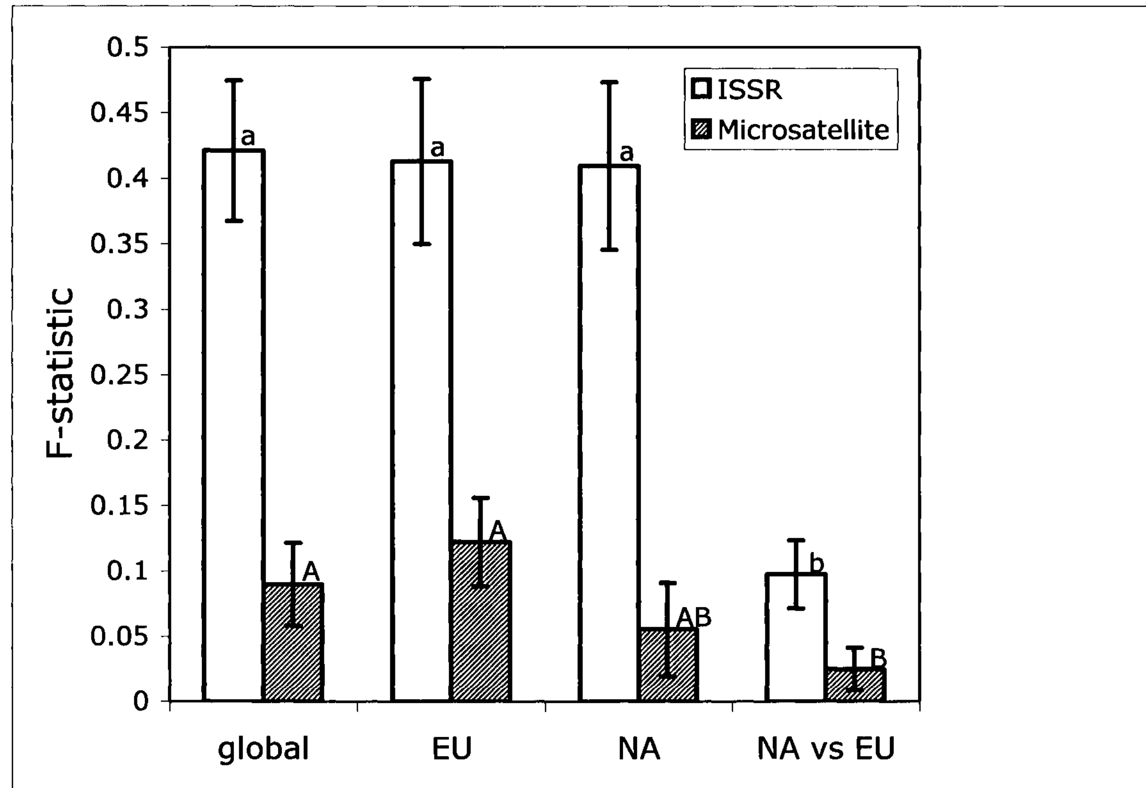


Figure 3.1:
F-statistics derived from ISSRs (open bars = θ_P) and microsatellites (hatched bars = F_{ST}). Estimates of F-statistics are over all sampling locations (global), over Eurasian sampling locations (EU), over North American sampling locations (NA), and between Eurasian and North American samples (NA vs EU). Error bars are 95% confidence intervals, and different letters indicate significantly different values.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

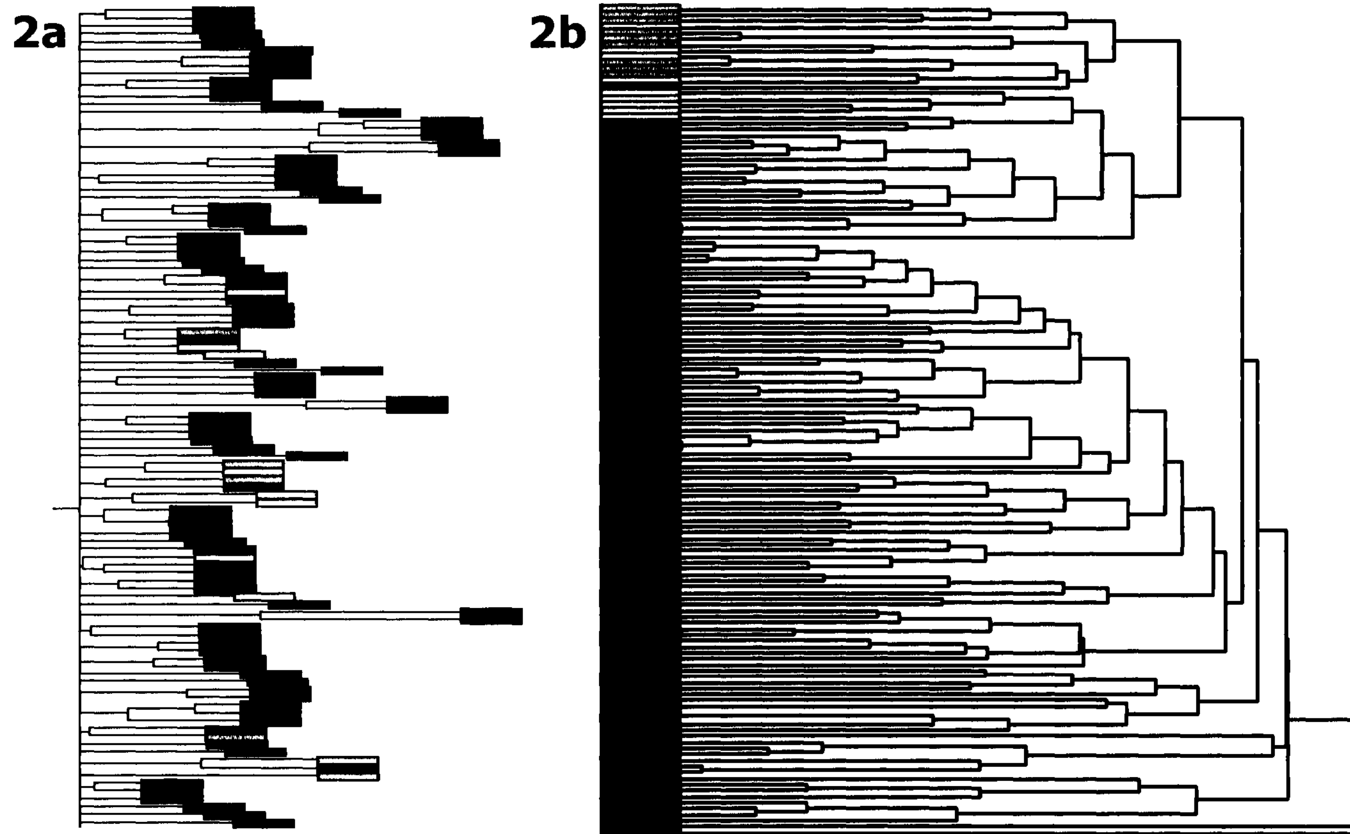


Figure 3.2:
Individual-based UPGMA trees from multilocus microsatellite (Figure 3.2a) and ISSR genotypes (Figure 3.2b). Different colors represent different sampling locations; reds, oranges, yellows, pinks, and purples are North American (invasive range) sites, while different shades of greens and blues represent European (native range) sites.

CHAPTER FOUR

POPULATION STRUCTURING IS HIGHER IN THE INTRODUCED RANGE OF *CENTAUREA DIFFUSA* THAN IN ITS NATIVE RANGE, WHILE GENETIC DIVERSITY IS SIMILAR

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Biological invasions offer excellent systems to study the evolutionary processes involved in introductions of species to new ranges. Molecular markers can reveal invasion histories and the effects of introductions on amounts and structuring of genetic variation. I used five polymorphic microsatellites to elucidate genetic diversity and population structure differences between native range and introduced range populations of a prominent North American rangeland weed, *Centaurea diffusa* (Asteraceae). I found that the total number of alleles and the number of private alleles was slightly higher in the native Eurasian range, and that allelic richness did not differ between the ranges, indicating overall levels of diversity were similar in Eurasia and North America. It therefore seems unlikely that this invasion has been affected by genetic bottlenecks or founder effects. Indeed, results of assignment tests suggest that multiple introductions have contributed to North America's *C. diffusa* invasion. Additionally, assignment tests show that both Eurasian and North American sites had a strong pattern of mixed genetic ancestry. This mixed assignment corresponded to a lack of geographic population structure among Eurasian samples. The lack of population structure in the native range conflicts with general expectations and findings to date for invasion genetics, and cautions that even species' native ranges may show signs of recent ecological upheaval. Despite the mixed assignments, North American samples showed strong population structure, suggesting that the invasion

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

has been characterized by long-range dispersal of genetically distinct propagules across the introduced range.

INTRODUCTION

Human-mediated transport of species into novel ecosystems can substantially alter the genetic variation of introduced populations relative to native populations, both in total amount and in how variation is partitioned within and between populations. It is becoming widely recognized that adaptive evolution of introduced species to novel ecosystems (e.g. Blossey & Notzold 1995, Bossdorf *et al.* 2005) plays an important role in biological invasions. Because genetic variation provides the raw materials for adaptive evolution, it is critical to understand how introductions affect the amount and structure of genetic variation. Genetic variation can be measured most easily by using neutral molecular markers, though it is quantitative genetic variation in phenotypic traits that is required for adaptive evolution. A pair of recent meta-analyses explored the relationship between neutral molecular marker diversity and quantitative trait variation; Merila and Crnokrak (2001) found a strong positive correlation, while Reed and Frankham (2001) found a weak positive correlation. Thus, variation in molecular markers is an adequate proxy for overall genetic variation. Molecular genetic data are often used to draw conclusions about how genetic diversity has changed following an introduction (see Durka *et al.* 2005,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Genton *et al.* 2005, Grapputo *et al.* 2005 for current examples). Additionally, the insights gleaned by studying the population genetics of invasive species can provide evidence for the invasion patterns of introduced populations (e.g. of multiple introductions, Durka *et al.* 2005, or of single founding events, Grapputo *et al.* 2005). Such information can guide research into control efforts. For example, Goolsby *et al.* (2004) inferred the origin of Old World climbing fern invasive to Florida, USA and used the information to find a potential biological control agent that is highly host specific to the invasive genotypes of the fern from that area of the native range.

When a species is introduced to a new range, the genetic variation within the newly founded populations can be reduced, increased, or unchanged relative to that in the native range. Reduction of genetic diversity is a frequent outcome of species introductions (Amsellam *et al.* 2000, DeWalt & Hamrick 2004, Grapputo *et al.* 2005). Founder effects and bottlenecks in population size have long been thought to be the dominant processes influencing genetic diversity in species' new ranges (Husband & Barrett 1991). These processes reduce genetic variability in the invaded range as a result of sampling effects or reduced population size (Nei *et al.* 1975). Selection can also lead to a pattern of reduced genetic diversity in the introduced range of a species. When one introduced genotype is more fit in the new range, its numbers will tend to increase relative to other genotypes, leading to a reduction in genetic variability (Kliber & Eckert 2005). If these types of processes predominate in an introduction, the introduced populations should be genetically depauperate when compared to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

native populations (Amsellam *et al.* 2000, DeWalt & Hamrick 2004, Grapputo *et al.* 2005).

Alternatively, genetic diversity can be higher within introduced populations than within native populations if introduced populations are founded from multiple, genetically differentiated regions within the native range. Crossing of genetically distinct lineages can lead to increased heterozygosity or novel combinations of alleles (Ellstrand & Scheirenbeck 2000). Several recent studies have shown patterns consistent with multiple introductions and high genetic diversity in invaders (Neuffer & Hurka 1999, Kolbe *et al.* 2004, Williams *et al.* 2005, Durka *et al.* 2005, Genton *et al.* 2005). Because increased genetic diversity also increases the raw materials necessary for adaptive evolution, an invader with high diversity may be quicker to evolve in its new range. A final alternate evolutionary outcome of a species introduction is that of no appreciable change in genetic diversity between the native and introduced ranges. This is a possibility if large population sizes are maintained during (Brown & Marshall 1981) or regained shortly after introductions (Zenger *et al.* 2003).

The way genetic variation is partitioned within and among populations, or the population structure, can differ between species' original and new ranges. In general, population structure develops when populations differentiate genetically either by adapting to local conditions or via random genetic drift. While local adaptation has been shown in some cases to be quite rapid (e.g. metal tolerance in plants, Pauwels *et al.* 2005), in general the differentiation process takes time, particularly in neutral

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

markers that evolve by genetic drift. Population structure can be apparent at small geographic scales if a species has a relatively short dispersal distance and individuals are likely to mate with others that are spatially close to them.

As pointed out by Roderick and Navajas (2003), it is often assumed that population structure will be lower in an organisms' introduced range than in its native range. This assumption has been borne out by multiple sets of data on plant introductions (reviewed in Bossdorf *et al.* 2005). Three processes may underlie this pattern. First, for recent introductions, there may not have been enough time for local adaptation or neutral mutations to lead to differentiation of populations in the new range. Second, the introduction and subsequent spread in a novel range represent major dispersal events and are likely linked to high gene flow, which minimizes population structure. Finally, if an introduction has imposed a bottleneck in population size that resulted in reduced genetic diversity, it will take time for new variation to arise and population structure to develop, even with a slow rate of spread. It is conceivable, however, that a pattern of lower population structure in the native range than in the introduced range could be observed if the native populations experienced high rates of gene flow, perhaps mediated by humans, and/or the invasive populations were founded by multiple introductions of distinct genotypes into geographically separate areas of the new range. This pattern could be propagated throughout the introduced range by subsequent discrete founding events, leading to a pattern of population structure without isolation by distance.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Here, I examine genetic diversity and population structure in the native and introduced ranges of a prominent rangeland weed, *Centaurea diffusa* Lam. (Asteraceae) using microsatellite loci. My questions are: 1) Is genetic diversity in the introduced range of *C. diffusa* reduced or increased within populations relative to the native range? 2) Is the population genetic structure of variation comparable between the two ranges? If so, is that structure organized in accordance with the expectations of isolation by distance? 3) Are multiple introductions likely to have occurred in this system? Do the Eurasian samples represent possible areas of origin of our invasive populations?

METHODS

Study Species

Centaurea diffusa Lam. (Asteraceae) is an annual to triennial species, with the length of its life cycle dependent on resource availability (Thompson & Stout 1991). It is an obligate outcrosser that adopts a tumbling habit for seed dispersal (Watson & Renney 1974). Native range *C. diffusa* (Eurasia) comprises both diploid (2x = 2n = 18) and tetraploid (2x = 4n = 36) cytotypes. North American *C. diffusa* appear to be diploid (A. Blair, *pers. comm.*). The first *C. diffusa* specimen collected in North America was found in an alfalfa field in Bingen, Washington in 1907 (Watson and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Renney 1974). One hundred years later, the species is distributed widely across North America, especially in xeric rangelands of the western USA (LeJeune & Seastedt 2001). The ecological destruction caused by this plant in its invaded range has spurred intensive efforts at biological control; thirteen insect species have been introduced to North America as natural enemies of *C. diffusa* (Müller-Scharer & Schroeder 1993, Lang et al. 2000).

Sample Collection and Preparation

My colleagues and I sampled *C. diffusa* from six Eurasian and eight North American locations (Table 4.1). Eurasian sampling sites were clustered around the Black Sea (Figure 4.1a), within the suspected areas of origin of the invasive populations, but I was not able to include areas further east (e.g. Turkmenistan), also suspected as a source of the introduction via contaminated alfalfa seed (Roché & Roché 1999). North American sites were spread over much of the invasive range of the species (Figure 4.1b). Sampling locations with less than 30 individuals were sampled exhaustively, while 30 or more plants were sampled in locations containing many individuals. When a site was not sampled exhaustively, samples were taken at least one meter apart to reduce the chance of sampling siblings and to capture the genetic variation present at the site. Depending upon the season of collection, either leaf tissue or mature seedheads were collected from individual plants. Leaf tissue was stored on desiccant for transport to the laboratory, while seeds were kept separate by

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

maternal plant, then germinated to provide fresh leaf tissue for DNA extraction. Genomic DNA was extracted from desiccated and fresh leaves using Qiagen Plant-Mini kits (Qiagen, USA) and stored at -20°C until genotyping.

Microsatellite analysis

I genotyped the DNA samples using five microsatellite loci (21CM36, 42CM27, 38CM22, CM15, and CM26 from Marrs *et al.* 2006). Polymerase chain reaction, allele visualization, and band scoring were performed according to Marrs *et al.* (2006). Multilocus microsatellite genotype data were analyzed by several different methods. To evaluate assumptions of neutrality by determining if any sampling sites deviated from the expectations of Hardy-Weinberg equilibrium and to test for linkage disequilibrium, I used the program GENEPOP on the web (http://wbiomed.curtin.edu.au/genepop/index.html, Raymond & Rousset 1995). I calculated several measures of genetic diversity, including the number of alleles found in each sampling location, the number of private alleles (Neel 1973, Slatkin 1985) in each sampling location, and allelic richness (a statistic corrected for sample size) in each range of the species using FSTAT version 2.9.3 (Goudet 2001). I calculated gene diversity as expected heterozygosity with standard deviation for each sampling location using Arlequin version 3.01 (Excoffier *et al.* 2005). I used analysis of covariance to compare private alleles per sampling site, total alleles per sampling site, and expected heterozygosity between European and North American *C. diffusa*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with continent as the predictor variable and sample size of each population as the covariate (JMP version 5.0, SAS Institute). Using the numbers of private alleles found per population, I calculated estimates of gene flow (migration) over all sampling sites, in North America, and in Eurasia (Slatkin 1985). Population genetic structuring in the native and introduced ranges were compared with an analysis of molecular variance (AMOVA) computed in Arlequin (Excoffier *et al.* 2005). Three sets of groups were compared with AMOVA: among continents (North America and Eurasia), among populations within Eurasia, and among populations within North America. In each case permutations to determine significance were set at 10,000 and the distance method used to determine variation in the dataset was the number of different alleles present in each group. As another way to look at population differentiation, I computed pairwise F_{ST} (Weir & Cockerham 1984) with P-values for each pair of sampling locations, excluding Cankiri, Turkey with N = 2, using Arlequin (Excoffier *et al.* 2005). I used a conservative Bonferroni correction to adjust α to equal $P \leq 0.05$ over all comparisons when determining significance. To test whether genetic distance was correlated to geographical distance (isolation by distance) I plotted pairwise F_{ST} against the geographic distance separating sampling locations. I then calculated Slatkin's (1993) similarity measure ($M = ((1/F_{ST}) - 1)/4$) for each pairwise comparison and used the program IBD (Bohonak 2002) to determine if a significant relationship existed between genetic and geographic distances. Finally, I used the assignment test software Structure version 2.0 (Pritchard *et al.* 2000, Falush *et al.* 2003) to infer the number of genetic clusters present in the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dataset and to assign individuals to these clusters. The model uses a Bayesian approach: it starts with prior distributions of model parameters and updates these with observed data by the Markov-chain Monte Carlo method, then it finds population groupings that are not in disequilibrium and assigns individuals to these populations. Model-based methods like Bayesian inference fundamentally assume that observations from each cluster are randomly drawn and that all potential source populations are predefined (Manel *et al.* 2005). Structure also assumes that that Hardy-Weinberg equilibrium is present within populations and that there is complete linkage equilibrium between loci within populations (Pritchard *et al.* 2000). I used Pritchard *et al.*'s (2000) *ad hoc* method for determining the approximate number of genetic clusters (K = number of clusters of individuals characterized by a set of allele frequencies at each locus) by giving the program a range of values for K as priors and determining which one gave the highest estimated log probability of the data, and which was consistent between runs. I computed three independent runs for each possible K from two to 18. The three runs were consistent for all values of K and assured that a burn-in of 250,000 followed by 750,000 data collection repetitions was sufficient to reach a stable α and estimated log probability of the data. I selected the admixture model, which assumes that individuals may have mixed ancestry (each individual is modeled as having inherited some fraction of its genome from ancestors) because of the likelihood of interpopulation and interspecific crossing in the *C. diffusa* system. I also selected to model allele frequencies as independent between

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

populations, a prior that expects allele frequencies in different populations to be somewhat different from one another (Falush *et al.* 2003).

Slatkin's private alleles migration model (1985) was used to estimate the amount of gene flow between sites. This algorithm uses the number of populations and the number of private alleles per population to estimate the degree of migration or gene flow between populations. Estimates of gene flow (N_M) were calculated among sites within Europe, among sites within North America, and among all sites on both continents.

To visualize genetic similarity between sampling sites, I constructed a phenetic tree based on populations using Nei's standard genetic distance (1972) with the unweighted pair group method with arithmetic mean (UPGMA) implemented in the software program Populations (Langella 1999). Nodes were bootstrapped with 10,000 permutations.

RESULTS

Hardy-Weinberg and linkage equilibria

Probability tests showed that the majority of the locus by sampling location comparisons did not deviate significantly from Hardy-Weinberg equilibrium. When corrected for multiple comparisons using the Bonferroni method, thirteen of 70

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

comparisons deviated from expectations and showed a deficit of heterozygotes. Of these thirteen deviations, seven occurred at locus 38CM22. The remainder of the heterozygote deficiencies occurred without pattern at the other loci. Only three sampling locations had more than one locus out of Hardy-Weinberg equilibrium: in Eurasia, the Pervomajsk, Ukraine site had three loci significantly deficient in heterozygotes, and Taiman Bay, Russia had two loci that deviated in the same manner, while in North America Helena, MT had two loci that showed heterozygote deficiency. Linkage disequilibrium was not significant at any locus by locus comparison.

Genetic diversity

All loci were polymorphic in all sampling locations, except for locus 21CM36 in Ürgüp, Turkey (Table 4.1). Overall, 76 of the 92 total alleles I recorded appeared in Eurasia, and 64 were observed in North America. The average number of alleles observed per locus ranged from 3.8 to 8.0 for Eurasian sites, and from 3.8 to 6.8 for North American sites (Table 4.1). There was no significant effect of region on the number of alleles found per sampling site (EU mean = 6.24, NA mean = 5.58, $F_{1,12} = 0.89$, $P = 0.368$), though sample size significantly affected the number of alleles observed at a site ($F_{1,12} = 17.13$, $P = 0.002$). I found 28 alleles private (exclusive) to Eurasia; nine of these were found in more than one Eurasian sampling location, and 19 were unique to individual sites. North America contained 16 total

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

private alleles, nine of which were unique to a single sampling location and seven of which were found at more than one North American site. The number of private alleles observed per sampling site was significantly higher in Eurasia than in North America (EU mean = 3.60, NA mean = 1.13, $F_{1,12} = 16.00$, $P = 0.003$). Sample size did not have a significant effect on the number of private alleles per site ($F_{1,12} = 0.08$, $P = 0.390$). Each sampling location was found to have alleles unique to it except two of the North American sites, Afton, WY and Kittitas, WA. In spite of the greater number of private alleles present in the native range of *C. diffusa*, allelic richness in the Eurasian ($R_S = 1.649$) and North American ($R_S = 1.638$) regions did not differ significantly ($P = 0.883$). Expected heterozygosity was slightly higher in North America than in Eurasia, but the difference was not significant (EU mean = 0.482, NA mean = 0.534, $F_{1,12} = 2.76$, $P = 0.131$). Population sample size had a significant effect on expected heterozygosity ($P = 0.014$), and there was a significant interaction between population sample size and region ($P = 0.049$).

Population structuring

Analysis of molecular variation (AMOVA, Table 4.2) showed a high degree of within-population variation. In each AMOVA I ran, the within-population variance component accounted for at least 90 percent of the variation. In the continental comparison (native + introduced ranges), none of the variation was due to continent-level differences, while among-population variation within continents was significant.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Eurasian comparison showed a barely significant among-population variance component explaining 2.93 percent of the variance, while the within-population component was highly significant and explained the remainder of the variance. North American sampling locations showed a stronger pattern of population structuring, with 10.38 percent of the variance explained by the among-population component. Even here, most of the variance was still explained by within-population variation.

Pairwise F_{ST} comparisons (Table 4.3) also demonstrated an overall trend of more population-level structuring in North America than in Eurasia. Only three of the 10 possible Eurasian to Eurasian comparisons had a significant F_{ST} ($P \leq 0.05$, adjusted for multiple comparisons with Bonferroni correction). Meanwhile, 23 of 28 different North American to North American comparisons resulted in significant F_{ST} values at $P \leq 0.05$, and when we corrected these for multiple comparisons, 18 comparisons remained significant. When pairwise comparisons from North American sites equal to or less than 1200 km apart (the maximum distance separating Eurasian sites) were included, 13 of 21 comparisons were significant. Of the 40 North American to Eurasian comparisons, 31 were significant at $P \leq 0.05$, and 17 of these remained significant after the correction for multiple comparisons. No relationship existed between geographical distance and pairwise genetic similarity measure (linearized F_{ST}, Slatkin 1993) in either range of the species (Figure 4.3, North America: $Z = 65614.38$, $r = -0.23$, $P = 0.184$; Eurasia: $Z = 17382.73$, $r = 0.14$, $P = 0.699$).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The Structure analyses determined that the dataset was consistent with eight genetic clusters ($K = 8$ with 94.8% probability). Proportion membership of each individual to each of the eight clusters is shown in Figure 4.3. Overall, most individuals did not assign strongly to a single genetic cluster, indicating a lack of genetic differentiation in the dataset. Most individuals from the Helena, MT site, however, did assign strongly to the cluster indicated by red. The UPGMA analysis was poorly resolved, and most branches had low bootstrap support, except for a long branch separating Ürgüp Turkey from the rest of the samples (Figure 4.4).

Gene flow, inferred from the numbers of private alleles per site (Slatkin 1985), was highest among Eurasian sites ($N_M = 0.078$) and lowest among North American sites ($N_M = 0.010$), while gene flow between continents was at an intermediate level ($N_M = 0.031$).

DISCUSSION

Departures from equilibrium

I observed several deviations from Hardy-Weinberg equilibrium in the dataset, all in the direction of heterozygote deficit. Seven of the deviant population by locus comparisons occurred at the same locus, 38CM22. This could be indicative of the presence of null alleles at this locus, or it could be due to selection. Null alleles

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

result from mutations (usually insertions, deletions, or point mutations in a primer binding site) that prevent PCR amplification of the DNA fragments containing the microsatellites. These mutations mask the presence of heterozygotes, leading to a miscount of the actual number of heterozygotes present. However, if the locus is near a gene under selection a similar pattern could be seen. Three sampling sites deviated from Hardy-Weinberg equilibrium at several loci, indicating that one or more of the assumptions underlying Hardy-Weinberg equilibrium (no selection, no mutation, no migration, no drift) are violated at these sites. One site was in the invasive range of the species and two sites were in the native range, an indication that native range *C. diffusa* are not necessarily stable or at equilibrium.

Genetic diversity similar in introduced and native ranges

Overall, allelic richness was similar between samples from the introduced range and samples from the native range of *C. diffusa*. Allelic diversity was slightly higher in the Eurasian samples, but not significantly so. The number of private alleles was higher in the sampled native range sites than in the introduced range sites. Even so, I detected many North American private alleles, suggesting that the European samples do not encompass all sources of the *C. diffusa* invasion into North America. There was no difference detected between the ranges in expected heterozygosity. Heterozygosity was higher, though not significantly so, in the invasive range, indicating that there has been no reduction in genetic diversity as a result of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

introduction. Sample size had a significant effect on heterozygosity and there was also a significant interaction between sample size and region, such that heterozygosity appeared to be more sensitive to sample size in Eurasia than in North America. The average number of alleles per locus was dependent on sample size, but did not differ between the native range sampling locations and the invasive range sites, indicating no per-population reduction in genetic diversity. Other studies of invasive species also have found comparable levels of genetic diversity across both ranges, and generally attribute this pattern to a history of multiple introductions to the new range (Genton *et al.* 2005, Durka *et al.* 2005, Therriault *et al.* 2005). If multiple introductions occur, slight bottlenecks in each introduction will not necessarily reduce overall levels of genetic diversity, consistent with the pattern seen here.

Population structure stronger in introduced range, no pattern of isolation by distance

As discussed earlier, samples representing populations of organisms are likely to be genetically structured by geographic location if dispersal distances are smaller than the distance between sampling locations. Analysis of molecular variance showed that the majority of genetic variation in *C. diffusa* occurs within the sampling locations, which is not uncommon for an obligate outcrosser (Nybom 2004). There was no significant continent level variance component, which is unsurprising given that Eurasia is the source for the North American invasion. Additionally, size homoplasy (different types of mutations leading to microsatellites of the same length)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

could cause differences between continents to be missed. When I compared variance partitioning between native and introduced ranges, however, I saw a surprising pattern. The invasive North American range had a higher degree of among-population variation than the native Eurasian range. Indeed, the sampled native Eurasian sites showed a small and only marginally significant level of population structure, indicating that they are probably close to panmictic. Meanwhile, over 10 percent of variation in North America was found between sampling locations. Pairwise F_{ST} calculations also showed increased population genetic structuring in the introduced range of *C. diffusa*. I found that only about one third (3 of 10) of the Eurasian to Eurasian sampling site comparisons had a pairwise F_{ST} significantly different from zero. At the same time, nearly two thirds (18 of 28) of the North American to North American comparisons had significant pairwise F_{ST}, indicating a degree of population structure that was not observed in the native range. To ensure that this was not merely the effect of the larger spatial scale of North American sampling, I also examined the pairwise comparisons between North American sites less than 1200 km apart (the maximum distance between Eurasian sampling sites). Here, 13 of 21 (over 60 percent) of comparisons were significant, indicating that the pattern of increased structure in the invasive range is not due to the greater distance between samples in North America. When I compared North American to Eurasian sampling sites, only 17 of the 40 pairwise F_{ST} values were significant, indicating less difference overall among continents than among sampling sites within North America. The results of the clustering software Structure are in agreement with AMOVA, pairwise F_{ST}, and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

isolation by distance results. The only site with most individuals strongly assigned to a single cluster was from Helena, MT (Figure 4.3). None of the Eurasian sites showed consistent assignment of individuals to a single cluster, suggesting a high degree of gene flow. Indeed, gene flow estimates based on the numbers of private alleles (Slatkin 1985) were higher in Eurasia than in North America. This is an unexpected result; if population genetic structuring is not present in the native range, why is it present in the introduced range?

Isolation by distance is a pattern we might expect if populations are likely to be founded by or occasionally share genetic material with their closest neighbors. It is a more specific case of population structuring where geographic and genetic distances are positively correlated, and closely related individuals tend to be clustered spatially. Since there was negligible population structure in the native range of the species, I would not expect to see any significant isolation by distance, and I did not. I did observe population structure in the introduced range of the species, however, so the results of isolation by distance tests here are more interesting. There was no relationship between geographic and genetic distance between sampling sites in North America (Figure 4.2).

Centaurea diffusa, even in its native range, tends to be a roadside weed that flourishes in disturbed areas (Watson & Renney 1974). As a result, there may have never been much structuring in the native range, and the structure we see in the introduced range of the species may simply be a result of sampling effects and establishment of different genotypes in different areas after multiple introductions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Alternately, the native range *C. diffusa* might historically have been more genetically structured, but that structure may have been disrupted by human-mediated transport or interspecific hybridization with *C. stoebe*. The region sampled has been intensively disturbed via both agriculture and wars. Invasion biologists tend to think about native ranges of species as being more 'natural', stable, and closer to equilibrium, but this may not always be the case. Few areas of the world remain unaffected by human development, and clearly, native range populations may be just as genetically disrupted as introduced range populations. Some hypotheses of invasion success rely on the assumption that native range populations are structured, and our dataset demonstrates that these assumptions are not always appropriate.

Multiple introductions are likely

The AMOVA and isolation by distance analyses showed that sampling sites within North America are genetically distinct, but sites that are close to one another spatially are not necessarily closer genetically. This pattern is consistent with multiple introductions of different genotypes from the native range into different areas of North America. In such a situation, population structure is the result of the chance establishment of different genotypes in different areas. This pattern could be reinforced with long-range 'leapfrog' dispersal, likely mediated by human transport of propagules. When populations are not founded by their nearest neighbors, a correlation between geographic and genetic distance will not develop.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Several relationships can be inferred from the Structure results. The cluster assignments of individuals from Bingen, WA, Afton, WY, and Fort Collins, CO are similar, suggesting that those populations are related. *Centaurea diffusa* was first recorded in North America from Bingen, WA, and the sampling site there shows mixed assignment. The two prominent clusters are indicated with blue and orange (Figure 4.3). I cannot discern the probable native range origin for these individuals, as the genetic clusters are poorly represented in my samples from Eurasia, but the same clusters appear in the Afton, WY and Fort Collins, CO sampling sites, indicating that they may have been founded from propagules originating at this initial region of introduction. Interestingly, the Red Feather, CO site, which is less than 40 miles from Fort Collins, does not share assignment to either of these clusters. This is additional evidence for leapfrog dispersal. The Helena, MT site assigned most strongly to a single cluster, indicating a different introduction event for this population. None of the Eurasian locations sampled correspond to the Helena, MT cluster.

An unrooted UPGMA tree built from this dataset (Figure 4.4) shows many of the North American sites together and most closely related to the Taiman Bay, Russia and Panitsovo, Bulgaria native range sites, but bootstrap support on all of these nodes is very low and results should be cautiously interpreted. The Helena, MT and Trinity, CA invasive range sites were not included in this group, possibly indicating novel introduction events for each of these sites, but again, only the node placing Trinity, CA furthest from the rest of the introduced range sites has bootstrap support greater than 50 percent. The Ürgüp, Turkey site from the native range appears to be unique.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Given the long branch separating this population from all the others and the relatively strong (81 percent) bootstrap support for it being the farthest terminal, it seems unlikely to be the source region for the introduced populations. Additionally, this Turkish site seems to be distinct from all other native range sites. It is possible that interspecific hybridization may be affecting this result. *Centaurea diffusa* is known to hybridize with its congener, *C. stoebe* (Gáyer 1909), which is found throughout most of *C. diffusa*'s native and introduced ranges (Watson & Renney 1974). It has been suspected that most or all of our North American *C. diffusa* may be the result of hybridization or introgression with *C. stoebe* (Ochsmann 2001). If this is true, the Ürgüp Turkey site might be unique from the rest of the samples simply in that it represents the only example of 'pure' *C. diffusa* included in this study.

Given the extremely low degree of population structuring in the native range of this species and the evidence from private alleles that I did not sample some regions that contributed to this invasion, it is unsurprising that assignment tests failed to positively assign our North American sampling sites to probable Eurasian progenitors. If the native range were sampled more extensively, we might eventually be able to track down the origins of our North American invasion(s), but we might not. If near-panmixia is found across the Eurasian range then it may not be possible to pinpoint origins of invasive populations even with more samples.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Conclusion

My results suggest that *Centaurea diffusa* has been introduced multiple times into North America. This scenario is supported by the lack of a significant reduction in genetic diversity in the introduced range, and by assignment tests and *F*-statistics revealing strong population structure within North America. Population structure was found to be stronger in the species' introduced range than in its native range. This higher invasive range structure may be a result of founder effects associated with separate introductions. The results are consistent with two or more introductions, however given the lack of population structure within the native range, I am unable to pinpoint origins of those introduction. Additional samples from further east in Asia than I was able to include here might help us to draw more definitive conclusions regarding the specific origins of North American *C. diffusa*, as well as the evolutionary forces likely to be at work in this system.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Amsellam, L, JL Noyer, T Le Bourgeois, M Hossaert-McKey. 2000. Comparison of genetic diversity in the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. Molecular Ecology 9: 443-455.

Blossey, B, R Notzold. 1995. Evolution of increased competitive ability in invasive nonindiginous plants: a hypothesis. Journal of Ecology 83: 887-889.

Bohonak, AJ. 2002. IBD (Isolation By Distance): A program for analyses of isolation by distance. Journal of Heredity 93:153-154.

Bossdorf, O, H Auge, L Lafuma, WE Rogers, E Siemann, D Prati. 2005. Phenotypic and genetic differentiation between native and introduced plant populations. Oecologia 144(1): 1-11.

Bossdorf O, D Prati, H Auge, B Schmid. 2004. Reduced competitive ability in an invasive plant. Ecology Letters 7: 346-353.

Brown, AHD, DR Marshall. 1981. Evolutionary changes accompanying colonization in plants. In: Eds Scudder GEC and JL Reveal. Evolution Today, Proceedings of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Second International Congress of Systematic and Evolutionary Biology, pp. 351-363. Hunt Institute for Botanical Documentation, Carnegie-Mellon University, Pittsburgh.

Carroll, SP, JE Loye, H Dingle, M Mathieson, TR Famula, MP Zalucki. 2005. And the beak shall inherit – evolution in response to invasion. Ecology Letters 8(9): 944-951.

DeWalt, SJ, JL Hamrick. 2004. Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, *Clidemia hirta* (Melastomataceae). American Journal of Botany 91(8): 1155-1162.

Durka, W, O Bossdorf, D Prati, H Auge. 2005. Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. Molecular Ecology 14: 1697-1706.

Ellstrand, NC, KA Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Academy of Sciences, 97(13): 7043-7050.

Excoffier, L, G Laval, S Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Falush D, M Stephens, JK Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164(4): 1567-1587.

Gáyer, G. 1909. Vier neue Centaureen der Flora von Ungarn. Magyar Botanikai Lapok 8: 58-61.

Genton, BJ, A Shykoff, T Giraud. 2005. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. Molecular Ecology 14: 4275-4285.

Goolsby, JA, JR Makinson, DM Hartley, R Zonneveld, AD Wright. 2004. Pre-release evaluation and host range testing of *Floracarus perrepae* (Eriophyidae) genotypes for biological control of Old World climbing fern. *In* JM Cullen, DT Briese, DJ Kriticos, WM Lonsdale, L Morin and JK Scott, eds. Proceedings of the XI International Symposium on Biological Control of Weeds. Canberra, Australia: CSIRO Entomology. Pp. 113-116.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www.unil.ch/izea/softwares/fstat.html.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Grapputo, A, S Boman, L Lindstrom, A Lyytinen, J Mappes. 2005. The voyage of an invasive species across continents: Genetic diversity of North American and European Colorado potato beetle populations. Molecular Ecology 14: 4207-4219.

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

Hufbauer, RA. 2004. Population genetics of invasions: Can we link neutral markers to management? Weed Technology 18: 1522-1527.

Husband, BC, SCH Barrett. 1991. Colonization history and population genetic structure of *Eichornia paniculata* in Jamaica. Heredity 66: 287-296.

Kliber, A, CG Eckert. 2005. Interaction between founder effect and selection during biological invasion in an aquatic plant. Evolution 59(9): 1900-1913.

Kolbe, JJ, RE Glor, LRG Schettino, AC Lara, A Larson, JB Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431: 177-181.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lang, RF, RD Richard, PE Parker, L Wendel. 2000. Release and establishment of diffuse and spotted knapweed biocontrol agents by USDA, APHIS, PPQ in the United States. Pan-Pacific Entomologist 76(4): 197-218.

Langella, O. 1999. Populations 1.2.28 (12/5/2002): a population genetic software. CNRS UPR9034. Available at http://pge.cnrs-gif.fr/bioinfo/populations/index.php.

Lau, JA. 2006. Evolutionary responses of native plants to novel community members. Evolution 60(1): 56-63.

LeJeune, KD, TR Seastedt. 2001. *Centaurea* species: The forb that won the West. Conservation Biology 15(6): 1568-1574.

Manel, S, OE Gaggiotti, RS Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20(3): 136-142.

Marrs, RA, RA Hufbauer, SJ Bogdanowicz, R Sforza. 2006. Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae). Molecular Ecology Notes 6(3): 837-840.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Merila, J, P Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology 14: 892-903.

Mooney, HA, EE Cleland. 2001. The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences 98(10): 5446-5451.

Müller-Scharer H, D Schroeder. 1993. The biological-control of *Centaurea* spp in North America – Do insects solve the problem. Pesticide Science 37(4): 343-353.

Neel, JV. 1973. Private genetic variants and frequency of mutation among South-American Indians. Proceedings of the National Academy of Sciences of the United States of America 70(12): 3311-3315

Nei, M. 1972. Genetic distance between populations. American Naturalist 106(949): 283.

Nei, M, T Maruyama, R Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution 29 (1): 1-10.

Neuffer, B, H Hurka. 1999. Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. Molecular Ecology 8: 1667-1681.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular Ecology 13: 1143-1155.

Ochsmann, J. 2000. Morphologische und molekularsystematische Untersuchungen an der Centaurea stoebe L.-Gruppe (Asteraceae-Cardueae) in Europa. PhD Dissertation. Ed. J. Cramer, Stuttgart, Georg-August University, Göttingen.

Ochsmann, J. 2001. An overlooked knapweed hybrid in North America: *Centaurea* x *psammogena* Gayer (diffuse knapweed x spotted knapweed). In: L. Smith, ed. The First International Knapweed Symposium of the Twenty-First Century 15-16 March 2001, Couer d' Alene, Idaho.

Pauwels, M, P Saumitou-Laprade, AC Holl, D Petit, I Bonnin. 2005. Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. Molecular Ecology 14(14): 4403-4414.

Pritchard, JK, M Stephens, P Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Raymond, M, F Rousset. 1995. GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248-249.

Reed, DH, R Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55(6): 1095-1103.

Roché, BF Jr., CT Roché. 1999. Diffuse knapweed. In: RL Sheley and JK Petroff, eds. Biology and Management of Noxious Rangeland Weeds. Oregon State University Press, Corvalis, USA.

Roderick, GK, M Navajas. 2003. Genotypes in novel environments: Genetics and evolution in biological control. Nature Reviews Genetics 4: 889-899

Slatkin, M.1985. Rare alleles as indicators of gene flow. Evolution 39(1): 53-65.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264-279.

Therriault, TW, MI Orlova, MF Docker, HJ MacIsaac, DD Heath. 2005. Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in eastern Europe: high gene flow and multiple introductions. Heredity 95(1): 16-23.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Thompson, DJ, DG Stout. 1991. Duration of the juvenile period in *C. diffusa* (*Centaurea diffusa*). Canadian Journal of Botany 69: 368-371.

Watson, AK, AJ Renney. 1974. The biology of Canadian weeds: 6. *Centaurea diffusa* and *C. maculosa*. Canadian Jounal of Plant Science 54: 687-701.

Weir, BS, CC Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Williams, DA, WA Overholt, JP Cuda, CR Hughes. 2005. Chloroplast and microsatellite DNA diversities reveal the introduction history of Brazilian peppertree (*Schinus terebinthifolius*) in Florida. Molecular Ecology 14: 3643-3656.

Zenger, KR, BJ Richardson, AM Vachot-Griffin. 2003. A rapid population expansion retains genetic diversity within European rabbits in Australia. Molecular Ecology 12: 789-794.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4.1:
Sampling locations of *C. diffusa*, site codes used in Table 2 and Figures 3 and 4, sample size (N), approximate GPS coordinates, percent polymorphic loci (P), average number of alleles per locus (A_A), number of private alleles (A_P: sampling location A_P values are the number of alleles unique to a single sampling location; North America and Eurasia A_P values are the number of alleles unique to the continent, including alleles that appeared in more than one sampling location), and gene diversity (H) with standard deviation (SD).

Sampling location	Site code	N	GPS coordinates	P	A_A	A_P	H	SD
North America						**16**		
Afton, WY	USWY1	9	42.12N 110.15W	100	3.8	0	0.4917	0.3141
Helena, MT	USMT1	25	46.63N 112.00W	100	6.4	2	0.5913	0.3493
Bingen, WA	USWA1	34	45.72N 121.50W	100	6.8	1	0.5081	0.3066
Kittitas, WA	USWA2	14	46.62N 120.37W	100	6.2	0	0.6917	0.4063
Trinity, CA	USCA2	10	41.01N 121.95W	100	4.8	2	0.4450	0.2846
Fort Collins, CO	USCO1	29	40.57N 105.22W	100	5.4	1	0.4865	0.2976
Red Feather, CO	USCO2	27	39.60N 105.00W	100	6.6	2	0.6197	0.3627
Sedona, AZ	USAZ1	14	34.83N 111.15W	100	4.6	1	0.4365	0.2773
Eurasia						**28**		
Panitsovo, Bulgaria	BG34	20	42.36N 27.71E	100	7.4	5	0.5559	0.3336
Taiman Bay, Russia	RUS5	28	45.33N 37.22E	100	8.0	3	0.5904	0.3481
Anapa, Russia	RUS15	14	44.94N 37.32E	100	5.2	2	0.3594	0.2432
Ürgüp, Turkey	TR6	15	38.69N 34.87E	80	3.8	5	0.3107	0.2271
Cankiri, Turkey	TR10	2	40.76N 33.53E	n/a	n/a	1	n/a	n/a
Pervomajsk, Ukraine	UA3	23	48.09N 30.75E	100	6.8	3	0.5924	0.3508

Table 4.2:
Analysis of molecular variace (AMOVA) results. Distance method used was R_{ST}-like (sum of squared differences). 10,000 permutations were run to determine significance. "Populations" refers to sampling locations.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Native + introduced ranges				
Between continents	1	11.099	0.02061	1.41(n/s)
Among populations, within continents	12	63.353	0.11503	7.86****
Within populations	486	645.13	1.32743	90.73****
Native range (Europe)				
Among populations	5	12.056	0.03876	2.93*
Within populations	180	231.38	1.28544	97.07****
Introduced range (North America)				
Among populations	7	51.298	0.15664	10.38****
Within populations	306	413.75	1.35212	89.62****

* $P < 0.05$, **** $P < 0.00001$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4.3:
Pairwise F_{ST} values, by sampling location (see Table 1 for site abbreviations). The upper left corner of the table shows NA × NA comparisons, the lower left block shows NA × EU comparisons, and the lower right shows EU × EU comparisons. Bold values are significant at $P \leq 0.05$; starred values (*) are significant at $P \leq 0.05$ after Bonferroni correction for multiple comparisons.

	USWY1	USWA1	USCO1	USMT1	USWA2	USCA2	USCO2	USAZ1	TR6	UA3	BG34	RUS15
USWA1	0.0131											
USCO1	**0.0455**	**0.0959***										
USMT1	**0.1171**	**0.1132***	**0.2134***									
USWA2	0.0086	0.0076	**0.1035***	**0.0939***								
USCA2	**0.1675***	**0.1381***	**0.2139***	**0.1117***	0.0437							
USCO2	**0.0682**	**0.0852***	**0.1211***	**0.1410***	**0.0726***	**0.1250***						
USAZ1	**0.0736**	**0.0676***	**0.0788***	**0.0850***	0.0268	**0.0909**	**0.0867***					
TR6	**0.1397***	**0.0678**	**0.0918***	0.0219	-0.0397	**0.1954***	0.0274	0.0258				
UA3	**0.0861**	**0.1064***	**0.1833***	**0.0920***	**0.0579**	**0.0513**	**0.0829***	**0.0633**	-0.0253			
BG34	**0.0927**	**0.0942***	**0.1558***	**0.1106***	**0.0710**	**0.1206***	**0.0947***	**0.0727**	0.0406	**0.0820***		
RUS15	**0.1006**	**0.0659**	**0.0872**	0.0192	-0.0153	**0.0890**	0.0160	0.0258	-0.0830	0.0374	0.0172	
RUS5	**0.0828**	**0.0889***	**0.1400***	**0.0782***	0.0306	**0.0779***	**0.0635***	**0.0418**	-0.0537	**0.0591***	**0.0662***	-0.0227

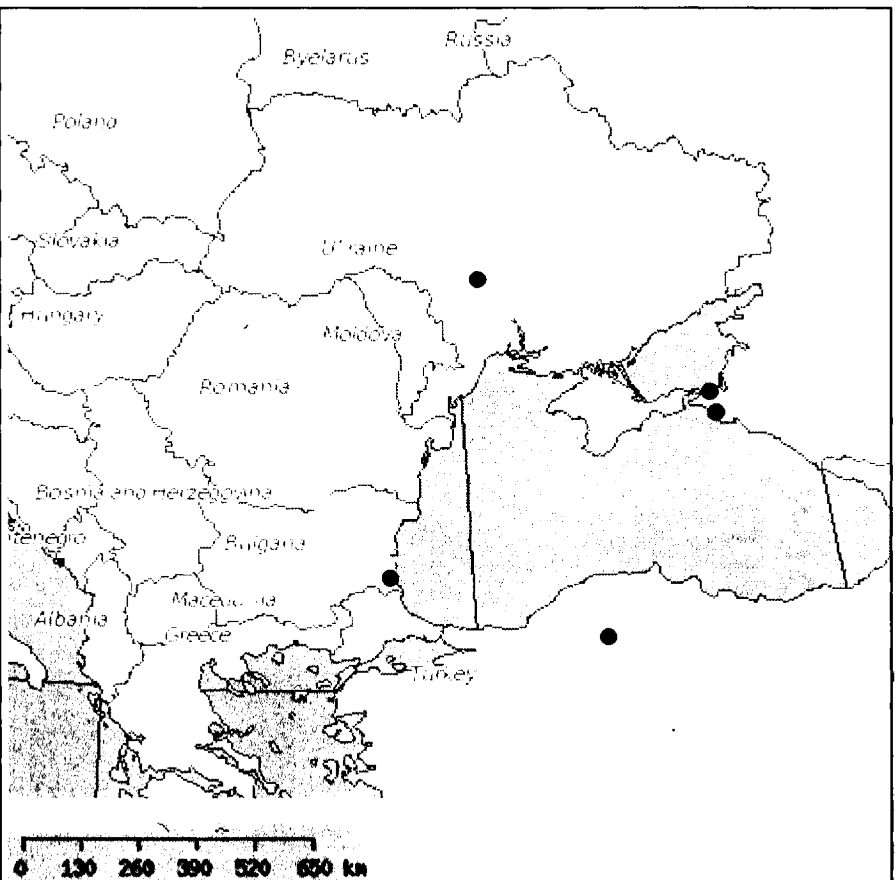


Figure 4.1a:
Eurasian *C. diffusa* sampling site locations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

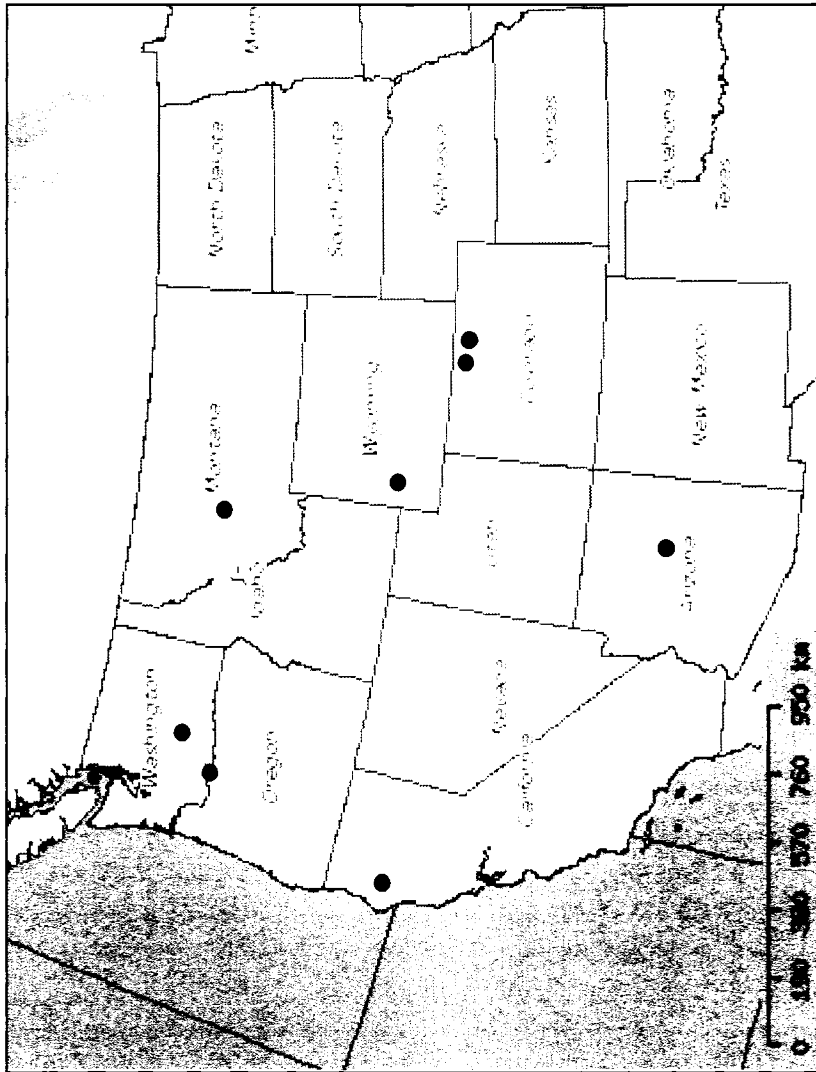


Figure 4.1b:
North American *C. diffusa* sampling site locations.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

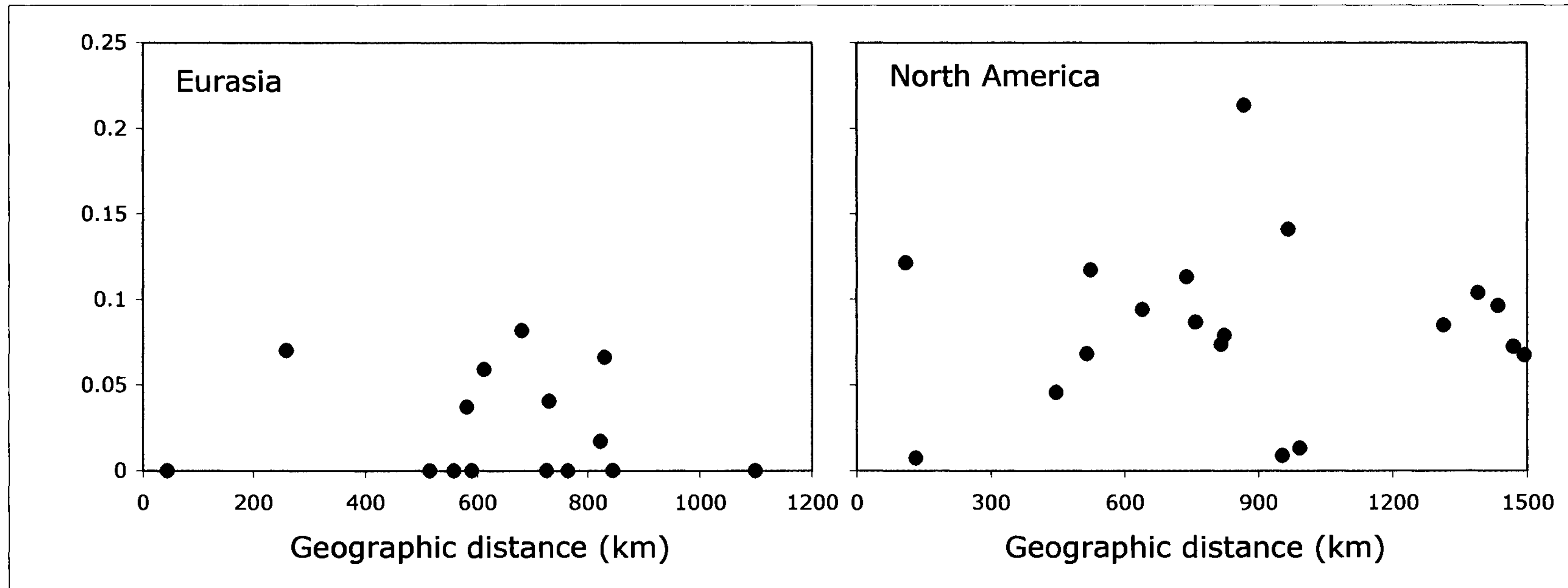


Figure 4.2:
Relationship between genetic and geographic distances in *Centaurea diffusa*. Pairwise F_{ST} is plotted against spatial distance in kilometers for all within-North America and within-Eurasia sampling location comparisons. All F_{ST} values < 0 were changed to zero for plotting. Isolation by distance was not significant in either range (see text for details). Note the difference in scale on the *x*-axis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

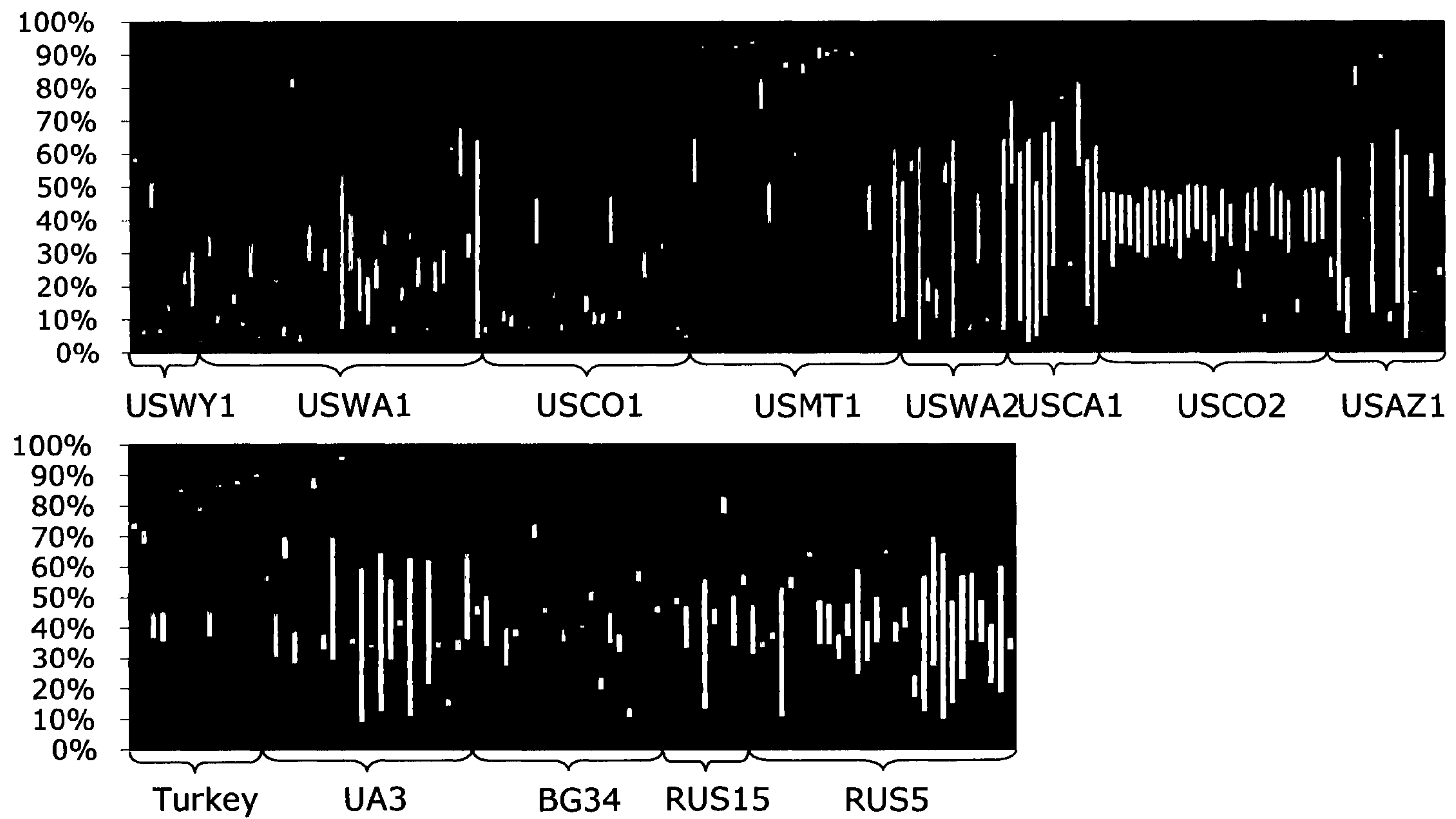


Figure 4.3:
Bar graph shows percent assignment of each individual to each genetic cluster (represented by different colors) when K (number of clusters) = 8. Each vertical bar represents one individual, and sampling location codes (see Table 4.1) are indicated along the x-axis. North American individuals are shown at the top, and Eurasian individuals are shown at the bottom part of the figure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

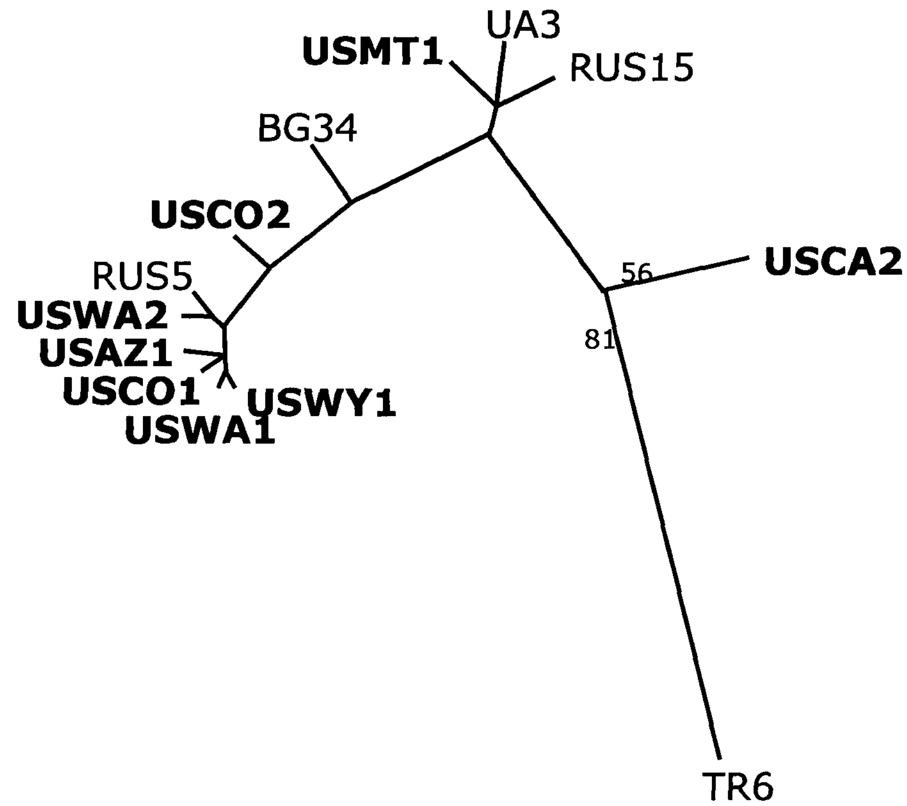


Figure 4.4:
Unrooted UPGMA tree showing relationships between sampling sites. Native range Eurasian sites are given in regular text, while introduced range North American sites are bolded; see Table 4.1 for site codes. Nodes with bootstrap support >50% are indicated with a number.

CHAPTER FIVE

EVIDENCE FOR MULTIPLE INTRODUCTIONS OF *CENTAUREA STOEBE* (ASTERACEAE) TO NORTH AMERICA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

The success of invasive species may depend strongly on the genetic resources they maintain through the invasion process. Here, I ask how many introductions have occurred in the North American weed *Centaurea stoebe* (Asteraceae, native to Eurasia), and I explore whether genetic diversity and population structure have changed as a result of introduction. I surveyed individuals from 15 European native range sites and 11 North American introduced range sites at six polymorphic microsatellite loci. I found more alleles overall and more private alleles in the native European range of the species, but the differences were not significant. Shannon-Weaver genetic phenotype diversity indices were also not significantly different between the ranges, while expected heterozygosity was marginally significantly higher in the invasive range. Thus, invasive *C. stoebe* in North America has not undergone a significant reduction in genetic diversity. Population structure (F_{ST}) was similar between the native range and the invasive range. Isolation by distance was significant in the native range, but not in the invasive range. Assignment tests grouped individuals into six genetic clusters. North American invasive individuals tended to have genetically mixed profiles, while European natives tended to assign more strongly to a single cluster. Several possible relationships were elucidated between North American and European sites in the assignment tests. Many North American individuals seem to share ancestry with Romania and Bulgaria, suggesting two separate invasions that have undergone gene flow in the invasive range. Samples

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from a site in New York clustered with samples from Hungary. Finally, a Colorado site was genetically quite distinct from the other invasive sites, and may represent another separate introduction, though its provenance could not be linked to one of the sampled European sites. Multiple introductions and the maintenance of high genetic diversity through the introduction process may be at least partially responsible for the success of *C. stoebe* in North America.

INTRODUCTION

Biological invasions are initiated by one or more introductions of a novel species into a new range. Introduction events can consist of many or only a few individuals (Gaskin *et al.* 2005), and both the number of introductions and the number of propagules introduced during each event can have a large effect on the genetic outcome of an invasion. Because genetic diversity provides the raw materials necessary for adaptive evolution, it has been hypothesized that multiple introductions may be one of the characteristics leading to especially problematic invaders capable of swift evolutionary response to selection pressure (Kolar & Lodge 2001).

When a small number of individuals are introduced, evolution via genetic drift is likely to occur (Husband & Barrett 1991). Genetic drift by bottlenecks or founder effects tends to cause marked decreases in genetic diversity in the invasive range of a species relative to its native range (Nei *et al.* 1975). For example, genetically

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

depauperate invasive populations have been observed in the Colorado potato beetle *Leptinotarsa decemlineata* (Grapputo *et al.* 2005) and the soap bush *Clidemia hirta* (DeWalt & Hamrick 2004). The lack of genetic variation is thought to be caused by few introductions and small founding population sizes. Additionally, successive, nested founder events can result in cumulative losses of genetic diversity, as in the giant bramble *Rubus alceifolius* (Amsellam *et al.* 2000).

Alternatively, invasive populations can be founded multiple times, and founding population sizes can be relatively large (Brown & Marshall 1981). In such cases, genetic diversity tends to be slightly reduced in the invasive range, but not significantly so. This pattern has been observed in the mustards *Capsella bursa-pastoris* (Neuffer & Hurka 1999) and *Alliaria petiolata* (Durka *et al.* 2005), in the Brazilian peppertree *Schinus terebinthifolius* (Williams *et al.* 2005), and in the freshwater mussel *Dreissena rostriformis bugensis* (Theirralt *et al.* 2005). Even if initial founding population sizes are small, genetic diversity can be maintained if population size rebounds quickly after introduction, as observed in the European rabbit *Oryctolagus cuniculus* (Zenger *et al.* 2003). If multiple introductions occur from genetically differentiated lineages in the native range, intraspecific crossing can lead to increased heterozygosity or novel combinations of alleles in the new range (Ellstrand & Scheirenbeck 2000). This can lead to diversity being higher within populations in the invasive range than within populations in the native range, a pattern seen in the brown anole *Anolis sagrei* (Kolbe *et al.* 2004) and in the common ragweed *Ambrosia artemisiifolia* (Genton *et al.* 2005).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Spotted knapweed, *Centaurea stoebe* L. (Asteraceae, synonyms: *C. maculosa* Lam. and *C. biebersteinii* DC.), is an especially problematic invader in North America for which little information on genetic diversity and population structure exists. After over 110 years of invasion, *C. stoebe* has spread over much of North America, and is considered noxious by fifteen U.S. states (USDA Plants Database 2006). Spotted knapweed is a species of great concern in the USA and Canada because of its ability to displace native grass species (Kedzie-Webb *et al.* 2001) and invade undisturbed natural systems (Tyser & Key 1988), reducing the forage quality of rangelands and changing the ecosystems of wildlands. The ecological and economic impact of spotted knapweed has encouraged intensive study into possible management options, including biological control. Studying the genetic diversity and population structure of *C. stoebe* will help us understand the introduction history of the species. Additionally, information about probable native range origins of *C. stoebe* may help focus future explorations for potential biological control agents.

Polymorphic molecular markers from presumably neutrally evolving areas of the genome, such as microsatellites, can be used to elucidate genetic diversity and population structure differences between native and introduced populations. Two recent meta-analyses examined whether a relationship exists between neutral molecular marker diversity and quantitative trait variation; Merila and Crnokrak (2001) found a strong correlation, while Reed and Frankham (2001) found a weak correlation. Though the correlation is imperfect and care should be exercised in interpretation (Hufbauer 2004), neutral marker variation is a good, relatively easy to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

measure proxy for overall quantitative genetic variation, and is often used to draw conclusions about how genetic diversity has changed following an introduction (see Durka *et al.* 2005, Genton *et al.* 2005, Grapputo *et al.* 2005 for current examples). Here, I use multilocus microsatellite genotypes of individuals from the native and introduced ranges of *C. stoebe* to investigate three questions regarding the invasion of this species: 1) Is genetic diversity higher in the native range of *C. stoebe* than in its introduced range? 2) Is genetic variation structured among sampling sites in the European and North American ranges of *C. stoebe*? If so, is isolation by distance significant in either range? 3) Are multiple introductions likely to have occurred, and can we infer native range origins of the invasive populations we sampled?

METHODS

Study species

Spotted knapweed, *C. stoebe*, is a biennial to short-lived perennial forb that has invaded North America (Watson & Renney 1974). *C. stoebe* is native to Europe, where it comprises two subspecies based on ploidy: a diploid subspecies found mainly in Western Europe (*C. stoebe stoebe* L.; 2x = 2n = 18), and a tetraploid subspecies found mainly in Eastern Europe and Western Asia (*C. stoebe micranthos* (S.G. Gmelin ex Gugler) Hayek; 2x = 4n = 36). North American invasive spotted

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

knapweed is thought to be the tetraploid subspecies, *C. stoebe micranthos* (Ochsmann 2000, Watson & Renney 1974). *C. stoebe* was first recorded in North America in Victoria, British Columbia in 1893 and was probably imported as a contaminant of alfalfa seed or ships' ballast (Groh 1944, Watson & Renney 1974).

Sample Collection and Preparation

My collaborators and I sampled from 17 European and 11 North American locations (Table 5.1, Figures 5.1a and 5.1b). European sampling sites were focused in Eastern Europe where the tetraploid subspecies *C. stoebe micranthos* is indigenous, since North American plants are thought to be entirely tetraploid (Ochsmann 2000). Several Western European (likely diploid) locations were sampled as well. North American sites were spread across the invasive range of *C. stoebe* within the U.S., including Western, Midwestern, and East Coast locations.

There are over 300 species of *Centaurea* in the native range, many of which are difficult to identify, and there are likely to be cryptic species not yet named (Ochsmann 2000). Examination of herbarium specimens from this study by experts indicated that samples from two of the locations in the native range that were thought to be *C. stoebe micranthos* instead represented two additional species. Samples from a location in Greece were identified as *C. grisebachii confusa*, a closely related endemic Greek species. Samples from a location in Bulgaria could not be identified (*C. spp.* unknown). Thus, the final sample included 15 *C. stoebe* sites in the native

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

range and 11 in the introduced range. Data from the two non-*C. stoebe* species are not included in the analyses below unless explicitly noted. Sites with fewer than 30 individuals were sampled exhaustively, while 30 or more plants were sampled at locations containing many individuals. At these larger sites, plants were sampled at least one meter apart to reduce the chance of sampling siblings, and to sample the range of genetic variation present at each site. Leaf tissue or mature seedheads were collected from individual plants, depending on the season of collection. Leaf tissue was dried and stored on desiccant for transport to the laboratory, while seeds were kept separate by maternal plant, then germinated to provide fresh leaf tissue for DNA extraction. Genomic DNA was extracted from desiccated and fresh leaves using Qiagen Plant-Mini kits and stored at -20°C until genotyping.

Microsatellite analysis

I genotyped *C. stoebe* DNA samples using six microsatellite loci (21CM36, 42CM27, 38CM22, CM15, CM26, and CM17) from Marrs *et al.* (2006). Polymerase chain reaction conditions, allele visualization techniques, and band scoring followed Marrs et al. (2006). Once all individuals from a sampling site were genotyped, I examined the genotypes to determine whether each site was likely to contain diploid or tetraploid individuals. If a site had no individuals with more than two alleles at any of the six loci, I inferred that the site contained the diploid subspecies, *C. stoebe stoebe*. If three or more alleles were observed at any locus for any individual, the site

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was assumed to contain the tetraploid subspecies, *C. stoebe micranthos*. I describe the analysis methods in detail below. When microsatellite loci are surveyed in polyploid species, it is not possible to know the copy number of each allele when an individual is not homozygous or fully heterozygous (for example, when one observes the allelic phenotype *A, B* in a tetraploid individual, the genotype could be *ABBB*, *AABB*, or *AAAB*), therefore commonly computed allelic richness statistics and many other common approaches to analyzing microsatellite data are not appropriate. Because the full genotypes for individuals that were not fully homozygous (e.g. showed only allele *A*) or fully heterozygous (e.g. showed four different alleles, *A*, *B*, *C*, and *D*) could not be determined, each allele observed in an individual was recorded once, and the remainder of the genotype was coded as missing data (e.g. when alleles *A* and *B* were observed, the genotype recorded was *AB*??, where '?' represents missing data). All individuals were coded and analyzed as tetraploids, even those from the two likely diploid Western European sites.

Comparing genetic diversity in native and introduced ranges

To examine genetic diversity in the European and North American ranges of *C. stoebe*, I computed expected heterozygosity, the average number of alleles and the number of private alleles (Neel 1973, Slatkin 1985) in each *C. stoebe* sampling location using the software program SPAGeDi (Hardy & Vekemans 2002), software that computes statistics and permutation tests of relatedness and differentiation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between populations for organisms of any ploidy level (Hardy & Vekemans 2002). SPAGeDi 1.2 assumes polysomic inheritance, as seen in autopolyploids. As *C. stoebe* is a likely allopolyploid, results should be interpreted with some caution. I used analysis of covariance to compare expected heterozygosity, the average number of alleles per locus, and the number of private alleles per site between European tetraploid *C. stoebe* and North American invasive *C. stoebe* with continent as the predictor variable and sample size of each population as the covariate (JMP version 5.0, SAS Institute). To determine if allelic diversity varied significantly between the two ranges, I computed Shannon-Weaver phenotype diversity indices for each range and compared them using permutation tests (10,000 permutations) with the program F-DASH (Obbard *et al.* 2006). A genetic phenotype is simply the observed alleles seen in an individual without any assumptions made about the copy number of each allele, and is used here because we do not have full genotype information for the majority of individuals in the study. The analyses with F-DASH were performed two different ways. First, I compared all European *C. stoebe* to the North American samples. Then, I compared only the European samples thought to be tetraploid (*C. stoebe micranthos*) to the North American samples.

Comparing population structure in the native and introduced ranges

I compared *F*-statistics (F_{ST}), which estimate the amount of among-population variation in a sample, within European *C. stoebe*, within North America, over all *C.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stoebe samples, and between European and North American *C. stoebe* using SPAGeDi (Hardy & Vekemans 2002). Permutation tests (20,000 permutations) were implemented to provide 95 percent confidence intervals around estimates of F_{ST} for each of these comparisons. Confidence intervals allowed me to determine whether population structure was greater in either the native or introduced range, and to determine whether continent-level structuring was significantly different from zero. I calculated pairwise F_{ST} values (Weir & Cockerham 1984) to determine genetic distance between sample locations using SPAGeDi 1.2 (Hardy & Vekemans 2002). Pairwise F_{ST} was plotted against the spatial distance between sampling locations to visualize the correlation between genetic and geographic distances. I calculated Slatkin's similarity measure (or linearized F_{ST}, $M = ((1/F_{ST}) - 1) / 4$, Slatkin 1993) for each population comparison and used Mantel tests and RMA regression implemented in the program IBD (Bohonak 2002) to determine if isolation by distance trends seen in the pairwise F_{ST} plots were significant.

The number of sites sampled does not necessarily represent the true number of genetic populations (K). The assignment test software Structure version 2.0 (Falush *et al.* 2003, Pritchard *et al.* 2000) was used to infer the number of genetic clusters present in the dataset (K), and to assign individuals to these clusters. The model uses a Bayesian approach, beginning with prior distributions of model parameters. It updates these with observed data by the Markov-chain Monte Carlo method, then finds population groupings that are in equilibrium and assigns individuals to these populations. Bayesian methods assume that observations are randomly drawn from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

each cluster and that all potential source populations are predefined (Manel *et al.* 2005). Structure also assumes that that Hardy-Weinberg equilibrium is present within populations and that there is complete linkage equilibrium between loci within populations (Pritchard *et al.* 2000). Individuals are therefore grouped with other individuals to form clusters that are in Hardy-Weinberg and linkage equilibrium, and K is simply the number of those clusters. In addition to determining the number of genetic clusters, Structure also determines percentage assignment of each individual to each genetic cluster. Though the program has been recompiled for polyploids (Falush *et al.* 2003), it is important to realize that Hardy-Weinberg equilibrium applies to diploids and the Structure results should be interpreted conservatively. I used Pritchard *et al.*'s (2000) *ad hoc* method for determining the approximate number of genetic clusters present in the complete dataset by giving the program a range of values for K as priors and determining which one gave the highest estimated log probability of the data. Because of their morphological similarity to *C. stoebe*, Structure analyses were run both with and without the two other species sampled (*C. grisebachii confusa* and *C. spp.* unknown, Table 1). These species were included to determine if they share a relationship with invasive *C. stoebe*, a result that would indicate a cryptic invasion of previously undocumented *Centaurea* species into North America. Structure analyses may also be useful in distinguishing diploid *C. stoebe* from tetraploids. I computed three independent runs for each possible K from two to 30 using a burn-in of 500,000 followed by 750,000 data collection repetitions, sufficient to reach a stable α and estimated log probability of the data. Results

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between runs were consistent. All iterations were run with the admixture model, which assumes that individuals may have mixed ancestry, because of the likelihood of interpopulation and interspecific crossing in the *C. stoebe* system. I also selected to model allele frequencies as independent between populations, a prior that expects allele frequencies in different populations to be somewhat different from one another (Falush *et al.* 2003).

Slatkin's private alleles migration model (1985) was implemented to estimate the amount of gene flow occurring between sites. This algorithm uses the number of populations and the number of private alleles per population to estimate the degree of migration or gene flow between populations. Estimates of gene flow (N_M) were calculated among sites within Europe (both overall, and just among the putatively tetraploid sites), among sites within North America, and among all sites on both continents.

To visualize genetic similarity between sampling sites, I constructed a phenetic tree using Nei's standard genetic distance (1972) with the unweighted pair group method with arithmetic mean (UPGMA) using the program Populations (Langella 1999) with 10,000 bootstrap replications. We included the *C. grisebachii confusa* individuals from Greece and the *C. spp.* unknown individuals from Bulgaria in our phenetic tree, for the reasons explained above in the Structure analysis description.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RESULTS

Ploidy

All North American sites contained individuals whose genotypes were consistent with tetraploidy (i.e. three or more alleles in at least one locus). Most European sites also appeared to be tetraploid, except the two Western European sites, Kembs, France and Basel, Switzerland. Most individuals in the Batmonostor, Hungary site had genotypes with two or fewer alleles at each locus (consistent with diploidy), but a few individuals had more than two alleles at a locus, indicating they were tetraploid.

Genetic diversity

The microsatellite loci examined in this study were highly polymorphic. All six loci were polymorphic in every sampling location except one; Kembs, France was monomorphic at locus CM17. Over all six loci and 28 sampling locations, I recorded a total of 178 alleles and, excluding the non-*C. stoebe* sites from Greece and Bulgaria, I found a total of 176 alleles. Of the *C. stoebe* alleles, 148 were found in the native European range of the species, while 125 were found in the invasive North American range. Overall, European *C. stoebe* sites averaged 23.7 alleles per locus, while North American sites had about 20.8 alleles per locus on average. Average numbers of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alleles per locus for individual sampling sites ranged from 3.0 to 12.5 in Europe, and from 4.7 to 10.0 in North America (Table 5.1), and did not differ significantly between the ranges (EU mean = 5.62, NA mean = 7.13, $F_{1,20} = 1.36$, $P = 0.257$), though sample size had a significant effect on the average number of alleles ($F_{1,20} = 5.01$, $P = 0.037$). Private alleles were found in both ranges; European *C. stoebe* had more private alleles (51) than North America (28) overall, but there was not a significant difference in the number of private alleles per site between the ranges (EU mean = 2.42, NA mean = 1.91, $F_{1,20} = 1.84$, $P = 0.190$). There was, however, a significant effect of sample size on the number of private alleles per site ($F_{1,20} = 18.85$, $P = 0.020$). Of the alleles private to European *C. stoebe*, 30 (58.8 percent) were private to individual sites (Table 5.1), while the remainder were shared among two or more European sites. Twenty-one of the 28 (75 percent) total private alleles in North America were unique to individual sites, and seven alleles were shared between two or more North American sites. Expected heterozygosity varied from a low of 0.373 at the likely diploid Kembs, France site to a high of 0.856 at the Jundola, Bulgaria site. Mean expected heterozygosity within sample locations was marginally significantly higher in the invasive range than in native range tetraploid *C. stoebe* (EU mean = 0.690, NA mean = 0.759, $F_{1,20} = 4.03$, $P = 0.058$), while location sample size had no effect on expected heterozygosity ($F_{1,20} = 0.20$, $P = 0.659$). Shannon-Weaver phenotype diversity was higher in North American samples (3.40) than in European *C. stoebe* samples (3.08), but the difference was not significant ($P = 0.123$). When diploid Western European samples were excluded from the dataset, North America

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

still had a slightly, but not significantly, higher phenotypic diversity (EU = 3.13, NA = 3.40, $P = 0.134$). The average number of different alleles carried by each individual was significantly higher in North America than in Europe (EU = 1.72, NA = 1.92, $P = 0.023$) when the likely diploid European samples were included. When the diploids were removed, the invasive range still had more different alleles per individual, though the difference was no longer significant (EU = 1.79, NA = 1.92, $P = 0.089$).

Population structure

I compared F_{ST} over all *C. stoebe* sites, among European *C. stoebe* sites, among North American sites, and between European and North American *C. stoebe* sites to determine the scale of population structure in our dataset (Figure 5.2). All four F_{ST} values were significantly greater than zero, indicating some level of population structure. The global average F_{ST} (among-site variation over all sites) was 0.117, indicating a highly significant amount of population structure overall. Among-site variation was lower in North America ($F_{ST} = 0.081$) than in Europe ($F_{ST} = 0.151$), but this difference was not significant. Differences between regions (EU versus NA) explained a small but significant amount of the variation ($F_{ST} = 0.023$). When population pairwise F_{ST} values were plotted against the geographical distance separating the *C. stoebe* sampling sites (Figure 5.3) a positive correlation between genetic and geographic distance in the native European range of the species was observed. No relationship appeared to exist between genetic and geographic distance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in North America. The program IBD (Bohonak 2002) confirmed these patterns. Slatkin's similarity measure (1993) was negatively correlated to geographic distance in Europe (Z = 139564.22, r = -0.35, P = 0.001), indicating a significant positive correlation between genetic and geographic distance. In North America, there was no significant relationship between the similarity measure and geographic distance (Z = 357828.70, r = -0.0942, P = 0.277). The private alleles migration model (Slatkin 1985) showed similar levels of gene flow between sites within Europe (N_M = 0.031) and between sites within North America (N_M = 0.029), while gene flow was slightly lower between the continents (N_M = 0.013), and slightly higher between putative tetraploids in the native European range (N_M = 0.039).

Multiple introductions and origins

Using Pritchard *et al.*'s (2000) method for estimating the number of genetic clusters in a dataset, I found the dataset including the non-*C. stoebe* individuals was consistent with K = 6 with over 99% probability. I then plotted each individual's percentage assignment to each of these six genetic clusters (Figure 5.4). While individuals from many sampling sites assigned to multiple clusters, at some sites most individuals assigned strongly to the same genetic cluster, indicating a higher degree of within-population similarity. Most European individuals assigned strongly to the same genetic cluster as the other individuals from their sampling site, a visual indication of population structure, while the North American individuals and sites

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tended to be more genetically mixed. Within Europe, I found that three sampling sites (Basel, Switzerland; Kembs, France; and Batmonoster, Hungary) assigned strongly to the cluster represented by green in Figure 5.4. Most individuals from these three sites also shared genotype profiles consistent with the diploid subspecies *C. stoebe stoebe* ($\leq$ 2 alleles per individual). Few individuals outside of these three sites assigned strongly to this diploid group. Sozopol, Bulgaria (*C. spp.* unknown) and Kastraki, Greece (*C. grisebachii confusa*) both showed strong assignment to the red cluster. The Jundola, Bulgaria and Monastery Ridge, Bulgaria sites were largely assigned to the orange cluster. Two European sites showed consistent assignment to genetic clusters not found elsewhere in the native range. Bohonye, Hungary assigned strongly to the cluster represented by grey, while Baia Mare, Romania assigned to the blue group. Finally, the European sites from Slovenia and the Ukraine showed much genetic admixture, including some partial assignment to the green group that included what I infer to be the diploid subspecies. Many of the North American admixed sites showed a split assignment, sharing possible ancestry from the blue cluster (strongly represented in the native range in Baia Mare, Romania) and the orange cluster (represented in the Jundola and Monastery Ridge, Bulgaria sites in Europe). The Keene Valley, NY site showed a large proportion assignment to the grey cluster, also found in Bohonye, Hungary. Vail, CO showed the strongest assignments in the invasive range of the species, to the cluster shown in yellow; this cluster was not well represented in the native range of *C. stoebe*. When the non-*C. stoebe* individuals were excluded from the analysis, the outcome was similar with one exception: an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

additional genetic cluster appeared (resulting in $K = 6$ again), and most individuals from Rüse, Bulgaria (BG27) and Grimes Creek, ID (USID1) shared about 50 percent assignment to this cluster (Figure 5.5).

Analysis of the dataset using UPGMA produced an unrooted tree with three main branches (Figure 5.6). Although there was low bootstrap support for the branches, most North American sites were grouped on a single branch with the Romanian and Hungarian sites. The other two branches of the tree largely consisted of European sites, each including only one North American site. The Couer D' Alene, ID site clustered with the Greek and Bulgarian sites, as well as with the presumably diploid Western European sites in Switzerland and France. In the final branch, the Vail, CO invasive site groups with the Ukrainian and Slovenian sites from the native range.

DISCUSSION

Ploidy

The microsatellite loci used in this study differentiated between diploid and tetraploid *C. stoebe*. Individuals from the range of the diploid subspecies *C. stoebe stoebe* assigned strongly to a distinct genetic cluster (Figures 5.4 and 5. 5). In the area of overlap of the subspecies' ranges, one site included individuals strongly assigned to that diploid cluster. These individuals also had allelic phenotypes consistent with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

diploidy (no more than two alleles present at any locus). Future researchers interested in answering questions related to the population genetics of the two subspecies may find this useful, as determination of ploidy level by cytology or flow cytometry prior to genotyping may be unnecessary.

Genetic diversity is comparable between ranges

Invasive North American sampling locations had fewer total and private alleles than native European sampling locations, but those reductions in numbers of alleles were not significant. Sample size strongly influences allele counts, and therefore the smaller sample sizes from North American sites may have exaggerated reductions in allelic diversity of the introduced populations. That those reductions are nonetheless not statistically significant suggests that invasive *C. stoebe* was not subjected to a severe demographic bottleneck during its introduction. Indeed, several other lines of evidence also suggest there was no founder effect or bottleneck during the invasion process. Expected heterozygosities of sample locations were significantly higher in the invasive populations than in the native populations. Shannon-Weaver phenotype diversity, which explicitly corrects for sample size, also did not differ significantly between ranges.

While the native range had more total and private alleles, the invasive range had more combinations of alleles. This corresponded with individuals on average carrying more different alleles in North American than in Europe. This outcome may

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

be related to the allele frequencies. European samples had some alleles that were dominant, and other rare alleles that occurred only at low frequency, while North American samples had a more even distribution of allele frequencies, and fewer rare alleles. The high degree of invasive range diversity seen in this study is similar to results seen by others and is consistent with an invasion pattern of multiple introductions (Durka *et al.* 2005, Genton *et al.* 2005, Gaskin *et al.* 2005), or of a quick return to large population sizes, allowing retention of genetic diversity after a demographic bottleneck (Zenger *et al.* 2003).

Population structure is similar between ranges

Population genetic structure (measured as F_{ST}) was similar between Europe and North America, with a trend toward higher structuring in the native range (Figure 5.2). Though most variation occurred within populations in both ranges, there was a significant among-population component to genetic variation in each range. In both *C. stoebe*'s native European and its introduced North American range, population structure is apparent. Additionally, we saw a small but significant amount of population differentiation at the continent level. This continent level effect could be simply because I have not captured all of the source populations in my sampling, but it could also be the result of genetic drift that occurred during the introduction process, population differentiation due to drift since introduction, or the result of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selection on different traits in the introduced range (Mooney & Cleland 2001, Blossey & Notzold 1995).

I saw significant isolation by distance in the native range of the species, but not in the invasive range (Figure 5.3). This pattern has also been found in other systems (Genton *et al.* 2005) and is thought to be due to the long history of undisturbed European populations.

Though population genetic structure is significant in *C. stoebe*, it seems to be operating on a larger scale than the one at which I sampled. If all of the collection locations truly corresponded to genetically distinct clusters, I would expect to find a K (the number of clusters inferred using Structure, Pritchard *et al.* 2000) close to the number of sites in the study (28). Instead, I estimated the number of genetic clusters in the dataset at just six.

Introductions and relationships between European and North American sites

When I looked at overall genetic similarity between sampling sites by constructing a UPGMA tree (Figure 5.6), I found three major branches. Nine of the eleven invasive North American *C. stoebe* sites cluster together on a single branch of the tree, along with the Romanian and Hungarian sites. The other branches of the tree are dominated by European sites. All of the Bulgarian sites (including the unknown *Centaurea* species from Sozopol) cluster together on one branch, along with several others: the invasive range Coeur D' Alene, ID site, the *C. grisebachii confusa* site,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and two Western European sites (likely of the diploid subspecies *C. stoebe stoebe*). The final branch consists of most of the Ukrainian sites, the Slovenian site, and the Vail, CO site from the invasive range. Overall, most native range sites tended to group on the tree with sites from the same country. This is consistent with the significant isolation by distance seen in Europe and discussed earlier; sites that are spatially close to one another tend to be genetically similar, whether the metric used is pairwise F_{ST} (as in isolation by distance analysis) or overall genetic similarity (as in UPGMA analysis). In the invasive range, there was no apparent pattern in the tree relating spatial to genetic distances. Though bootstrap support for the three branches is very weak, overall evidence of the phenetic tree seems to support at least three separate introductions of *C. stoebe* to North America, since each of the three branches of the tree include one or more North American sites. Because nine widely geographically separated North American sites (East Coast, Midwest, and West) are grouped into one clade, it seems that there is a predominant type of *C. stoebe* in North America, with other, perhaps minor, introductions also contributing to invasive range diversity.

I saw similar, but not identical, results using Bayesian assignment test methods (Figures 5.4 and 5.5). Rather than seeing three groups, Structure inferred six genetic clusters. The assignment test method is more sensitive, allowing me to look at individuals and see how well they each fit into the predefined clusters, while individual differences are collapsed to the population level in the UPGMA analysis. Some individuals assign very strongly to a single cluster, while others have a mixed

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

genetic profile and seem to share ancestry from two or more different clusters. Consistent with the previous analyses, I was able to visualize the higher degree of population genetic structure present in the native range of the species. Overall, most individuals within a sampling site in Europe assigned fairly strongly to a single genetic cluster, resulting in blocks of color that corresponded to sample site in the bar plot (Figures 5.4 and 5.5). In the invasive range, however, most individuals were not strongly assigned to a single cluster.

Genetically mixed *C. stoebe* is spread across the invasive range, and most of the individuals of mixed heritage seem to share greatest assignments to the clusters represented by blue and orange in Figure 5.4. The cluster represented by blue is also predominant in the Baia Mare, Romania site, which would indicate a Romanian origin for an introduction event, while the orange cluster is found in several Bulgarian sites. This result is also supported by the UPGMA analysis; most North American sites grouped with this Romanian site and the Hungarian sites. Another congruence between the UPGMA and Structure analyses is the uniqueness of the Vail, CO site. In both analyses this site is distinct from all other North American sites. The yellow cluster that dominates the genetic assignment of this site is not well represented in the Structure plot of the European samples, except in several Slovenian individuals. In the UPGMA analysis, Vail, CO is also most closely related to Slovenian sites, as well as most of the Ukrainian sites. The lack of broad assignment of European individuals to this yellow cluster suggests we missed at least some genetic sources of invasive populations when we sampled the native range. The fact that I found 28 private

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

alleles in North America also supports this conclusion. Despite the agreement between methods, the assignment test results also diverge in some ways from the patterns seen on the phenetic tree. The North American site from Keene Valley, NY assigns strongly to the grey Structure cluster, indicating a possible relationship to the Bohonye, Hungary site, and another probable introduction event (Figure 5.4), though in the UPGMA analysis, this site clusters with many other North American sampling locations (Figure 5.6).

Conclusions

Diversity reducing processes such as selection, genetic drift by bottleneck, or founder effect do not seem to have played an important role in the invasion of this weed. Indeed, if anything, invasive *C. stoebe* are more heterozygous than their native counterparts. Population genetic structuring was greatest among European sites, though it was also significant among North American sites and at the continent level, between Europe and North America. Isolation by distance was significant in the native range of *C.* stoebe, but not in its invasive range. I found evidence for at least four separate introductions of this species to North America. It seems likely that we have had at least one introduction from Romania, which contributed significantly to nearly all of the genetically admixed North American individuals. The other area that may have contributed to the mixed invasive populations is Bulgaria, though the evidence from the UPGMA tree and the assignment tests are in conflict over this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

point. An introduction from Hungary seems to be more highly localized, contributing mainly to the Keene Valley, NY samples. Finally, the Vail, CO site is unique among our North American sites, but none of the European sites assigned strongly to this cluster. It is likely to represent an introduction from an area of the native range that we did not sample.

Centaurea stoebe has been able to maintain its genetic diversity through the invasion process as a result of multiple introductions. Most North American sites appear to contain individuals with a similar genetically admixed assignment profile, likely the result of gene flow between separately introduced genotypes of *C. stoebe*. A minority of sites contain individuals that share a strong co-assignment and are genetically unique from other sites. From a management perspective, these results suggest that *C. stoebe* will continue to be difficult to control. North American *C. stoebe* has several native range origins and some of these genotypes appear to have undergone gene flow in the new range. If biological control agents are specialized at the level of host genotypes, then samples from any single one of the areas of origin within the native range are unlikely to be effective on all invasive genotypes of *C. stoebe*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Amsellam, L, JL Noyer, T Le Bourgeois, M Hossaert-McKey. 2000. Comparison of genetic diversity in the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. Molecular Ecology 9: 443-455.

Blossey, B, R Notzold. 1995. Evolution of increased competitive ability in invasive nonindiginous plants: a hypothesis. Journal of Ecology 83: 887-889.

Bohonak, AJ. 2002. IBD (Isolation By Distance): A program for analyses of isolation by distance. Journal of Heredity 93:153-154.

Burdon, JJ, RH Groves, PE Kaye, SS Speer. 1984. Competition in mixtures of susceptible and resistant genotypes of *Chondrilla juncea* differentially infected with rust. Oecologia 64:199-203.

DeWalt, SJ, JL Hamrick. 2004. Genetic variation of introduced Hawaiian and native Costa Rican populations of an invasive tropical shrub, *Clidemia hirta* (Melastomataceae). American Journal of Botany 91(8): 1155-1162.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Durka, W, O Bossdorf, D Prati, H Auge. 2005. Molecular evidence for multiple introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America. Molecular Ecology 14: 1697-1706.

Ellstrand, NC, KA Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Academy of Sciences 97(13): 7043-7050.

Falush, D, M Stephens, JK Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164(4): 1567-1587.

Gaskin, JF, D-Y Zhang, M-C Bon. 2005. Invasion of *Lepidium draba* (Brassicaceae) in the Western United States: distributions and origins of chloroplast haplotypes. Molecular Ecology 14: 2331-2341.

Genton, BJ, A Shykoff, T Giraud. 2005. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. Molecular ecology 14: 4275-4285.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Grapputo, A, S Boman, L Lindstrom, A Lyytinen, J Mappes. 2005. The voyage of an invasive species across continents: Genetic diversity of North American and European Colorado potato beetle populations. Molecular Ecology 14: 4207-4219.

Groh, H. 1944. Canadian Weed Survey. 2nd Annual Report of the Canadian Department of Agriculture. 74 pp.

Hardy, OJ, X Vekemans. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618-620.

Hufbauer, RA. 2004. Population genetics of invasions: Can we link neutral markers to management? Weed Technology 18: 1522-1527.

Husband, BC, SCH Barrett. 1991. Colonization history and population genetic structure of *Eichornia paniculata* in Jamaica. Heredity 66: 287-296.

Kedzie-Webb, SA, RL Sheley, JJ Borkowski, JS Jacobs. 2001. Relationships between *Centaurea maculosa* and indigenous plant assemblages. Western North American Naturalist 61(1): 43-49.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Kolar, CS, DM Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution. 16(4): 199-204.

Kolbe, JJ, RE Glor, LRG Schettino, AC Lara, A Larson, JB Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431: 177-181.

Langella, O. 1999. Populations 1.2.28 (12/5/2002): a population genetic software. CNRS UPR9034. Available at http://pge.cnrs-gif.fr/bioinfo/populations/index.php.

Manel, S, OE Gaggiotti, RS Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20(3): 136-142.

Maron, JL, M Vilà, R Bommarco, S Elmendorf, P Beardsley. 2004. Rapid evolution of an invasive plant. Ecological Monographs 74(2): 261-280.

Marrs, RA, RA Hufbauer, SJ Bogdanowicz, R Sforza. 2006. Nine polymorphic microsatellite markers in *Centaurea stoebe* L. (subspecies *C. s. stoebe* and *C. s. micranthos* (S. G. Gmelin ex Gugler) Hayek) and *C. diffusa* Lam. (Asteraceae). Molecular Ecology Notes 6(3): 837-840.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Merila, J, P Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology 14: 892-903.

Mooney, HA, EE Cleland. 2001. The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences 98(10): 5446-5451.

Neel, JV. 1973. Private genetic variants and frequency of mutation among South-American Indians. Proceedings of the National Academy of Sciences of the United States of America 70(12): 3311-3315

Nei, M. 1972. Genetic distance between populations. American Naturalist 106(949): 283.

Nei, M, T Maruyama, R Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution 29 (1): 1-10.

Neuffer, B, H Hurka. 1999. Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. Molecular Ecology 8: 1667-1681.

Obbard, DJ, SA Harris, JR Pannell. 2006. Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. Heredity 97: 296-303.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ochsmann, J. 2000. Morphologische und molekularsystematische Untersuchungen an der Centaurea stoebe L.-Gruppe (Asteraceae-Cardueae) in Europa. PhD Dissertation. Ed. J. Cramer, Stuttgart, Georg-August University, Göttingen.

Pritchard, JK, M Stephens and P Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Reed, DH, R Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55(6): 1095-1103.

Slatkin, M.1985. Rare alleles as indicators of gene flow. Evolution 39(1): 53-65.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264-279.

Therriault, TW, MI Orlova, MF Docker, HJ MacIsaac, DD Heath. 2005. Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in eastern Europe: high gene flow and multiple introductions. Heredity 95(1): 16-23.

Tyser, RW, CH Key. 1988. Spotted knapweed in natural area fescue grasslands: An ecological assessment. Northwest Science 62(4): 151-159.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

USDA Natural Resources Conservation Service Plants Database. 2006.
http://plants.usda.gov/java/nameSearch?mode=Scientific+Name&keyword
query=centaurea+maculosa.

Watson, AK, AJ Renney. 1974. The biology of Canadian weeds: 6. *Centaurea diffusa* and *C. maculosa*. Canadian Jounal of Plant Science 54: 687-701.

Weir, BS, CC Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Williams, DA, WA Overholt, JP Cuda, CR Hughes. 2005. Chloroplast and microsatellite DNA diversities reveal the introduction history of Brazilian peppertree (*Schinus terebinthifolius*) in Florida. Molecular Ecology 14: 3643-3656.

Zenger, KR, BJ Richardson, AM Vachot-Griffin. 2003. A rapid population expansion retains genetic diversity within European rabbits in Australia. Molecular Ecology 12: 789-794.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5.1:
Centaurea stoebe sampling locations, site codes used in figures 5.4, 5.5, and 5.6, number of individuals sampled per site (N), approximate GPS coordinates of sampling sites, average gene diversity over all loci (H_E, expected heterozygosity), average number of alleles per locus (A_A), and number of private alleles (A_P, the numbers of alleles private to each sampling site do not total to the number of alleles private to the range, because some alleles private to Europe or North America are found in more than one site within the range).

Location	Site code	N	GPS	H_E	A_A	A_P
Europe					23.7	51
Centaurea stoebe stoebe (2n)						
Basel, Switzerland	CH1	21	47.550N 7.583E	0.559	3.0	0
Kembs, France	F3	20	48.139N 7.083E	0.373	3.2	1
Centaurea stoebe micranthos (4n)						
Rüse, Bulgaria	BG27	28	43.716N 25.918E	0.705	7.8	2
Jundola, Bulgaria	BG4	20	42.056N 23.834E	0.856	12.5	7
Monastery Route, Bulgaria	BG8	10	42.123N 23.268E	0.740	5.7	3
Bohonye, Hungary	HU12	29	46.403N 17.468E	0.668	6.8	4
Batmonostor, Hungary	HU17	27	46.114N 18.931E	0.642	4.7	4
Baia Mare, Romania	RO20	25	47.407N 23.500E	0.650	5.8	0
Valea Argovei, Romania	RO25	3	44.368N 26.810E	0.756	4.3	1
Maribor, Slovenia	SLO6	21	46.447N 15.794E	0.686	4.7	5
Kamjanac Podilsky, Ukraine	UA19	4	48.658N 26.577E	0.763	3.0	1
Ostrag, Ukraine	UA31	8	50.201N 26.319E	0.521	5.2	0
Berdiciv, Ukraine	UA35	6	49.826N 28.630E	0.563	2.7	1
Kholodne Jar, Ukraine	UA7	9	49.016N 32.210E	0.700	4.5	1
Bila Tserkva, Ukraine	UA15	16	49.794N 30.050E	0.703	5.5	0

Table 5.1, continued:

Location	Site code	N	GPS	H_E	A_A	A_P
Europe, non-*C.stoebe*						
Centaurea grisebachii confusa						
Kastraki, Greece	GR50	22	39.724N 21.636E	0.696	8.0	2
Centaurea spp. unknown						
Sozopol, Bulgaria	BG37	13	42.412N 27.701E	0.753	7.2	0
North America					20.8	28
Centaurea stoebe micranthos (4n)						
Bend, OR	USOR1	30	44.050N 121.300W	0.791	7.0	3
Grimes Creek, ID	USID1	28	43.850N 115.750W	0.732	6.7	0
Couer D'Alene, ID	USID2	18	47.700N 116.800W	0.805	8.8	3
Florence, MT	USMT1	13	46.633N 114.200W	0.782	8.3	3
Hamilton, MT	USMT2	21	46.231N 114.150W	0.798	8.3	0
Seeley, MT	USMT3	10	47.217N 113.500W	0.809	6.2	1
Big Bend, CA	USCA1	9	41.007N 121.958W	0.762	4.7	0
Vail, CO	USCO3	30	39.650N 106.450W	0.669	6.2	2
Bayfield, WI	USWI1	10	46.800N 90.817W	0.795	6.5	2
Keene Valley, NY	USNY1	21	44.206N 73.767W	0.616	5.7	2
Middletown, VA	USVA1	29	39.000N 78.250W	0.761	10.0	5

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

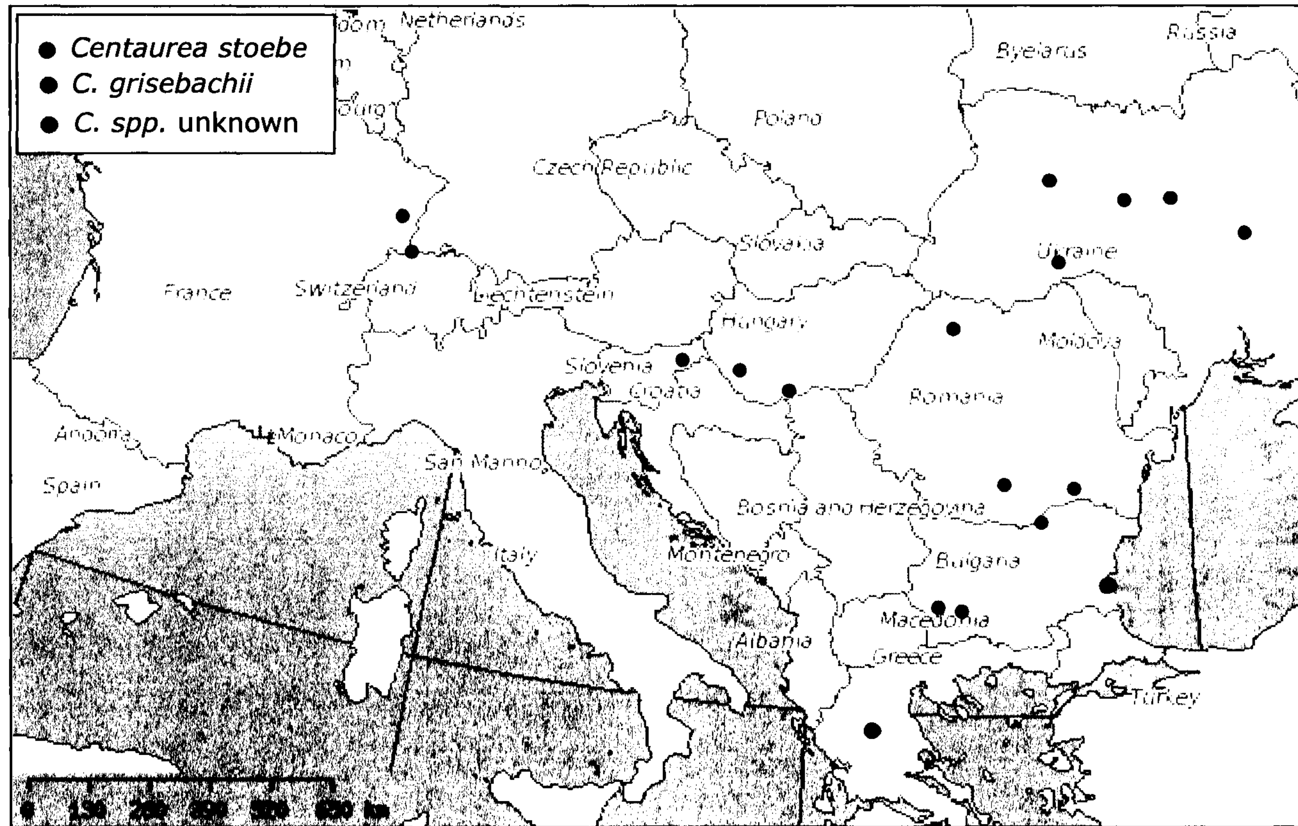


Figure 5.1a:
European *Centaurea stoebe* sampling sites are shown as filled circles. Non-*C. stoebe* sites are indicated in red (*C. grisebachii confusus*) and blue (*C.* spp. unknown).

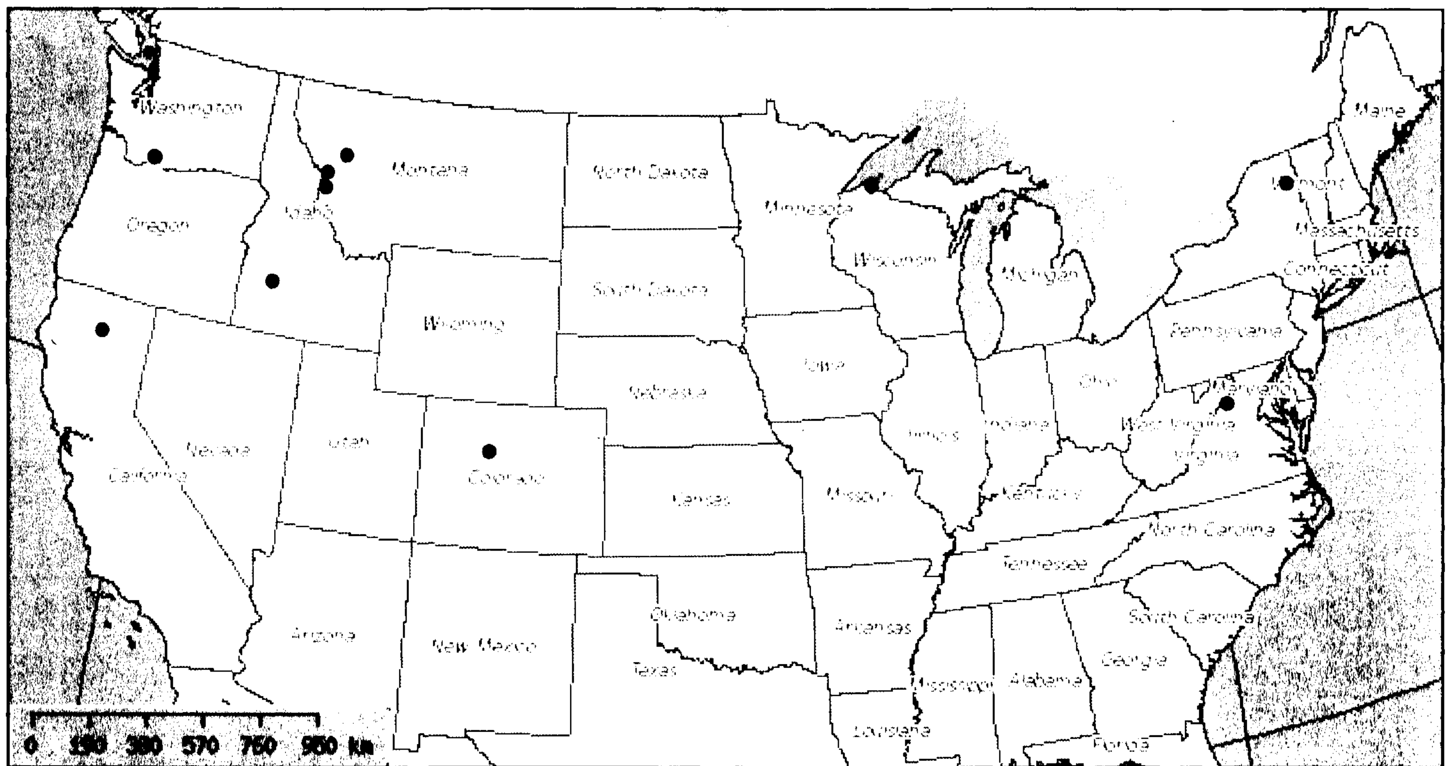


Figure 5.1b:
North American *Centaurea stoebe* sampling sites are shown as filled circles.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

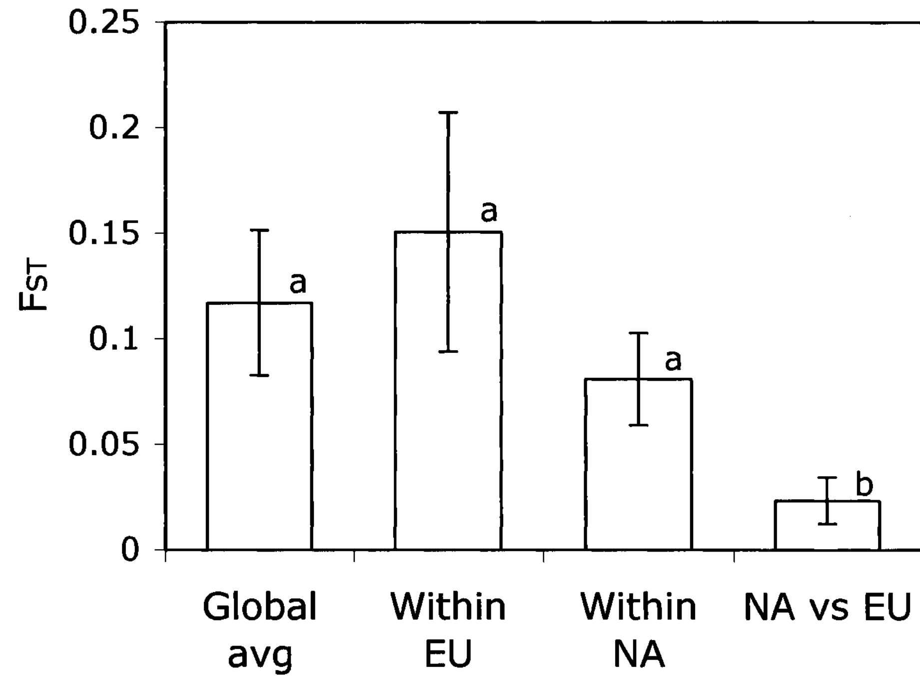


Figure 5.2:
Estimates of F_{ST} over all *C. stoebe* sites (global avg), among European *C. stoebe* sites (within EU), among North American sites (within NA), and between *C. stoebe* on the two continents (NA vs EU). Error bars are 95% confidence intervals and different letters indicate significantly different F_{ST} values.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

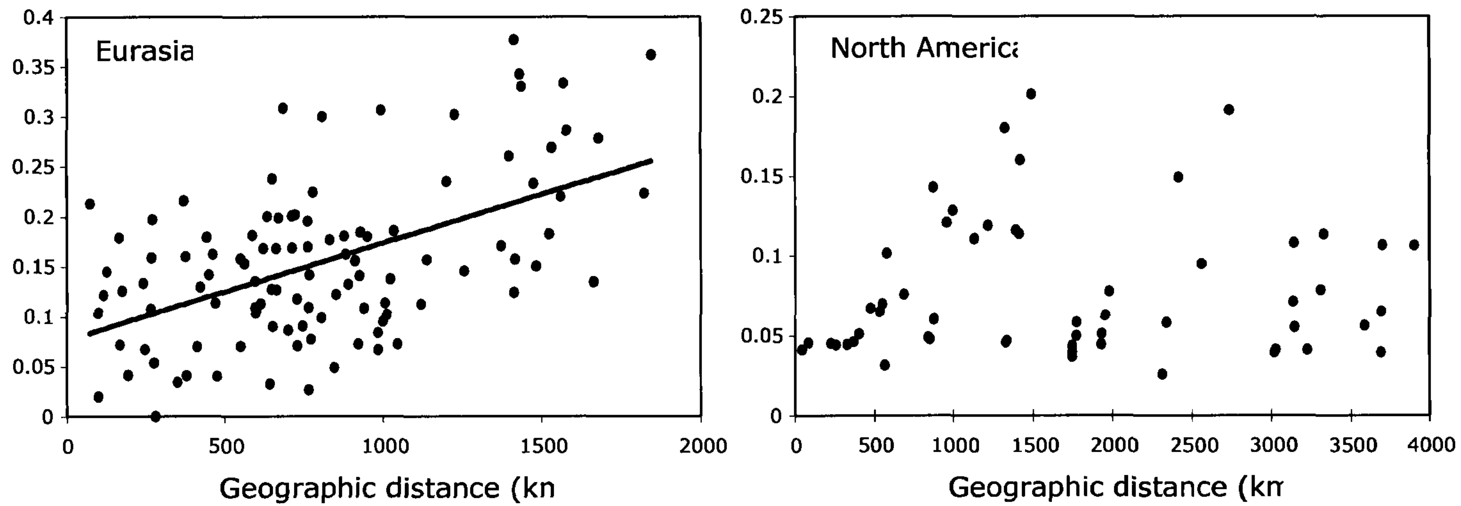


Figure 5.3:
Relationship between geographical distance (in kilometers) and genetic distance (given as pairwise F_{ST}) in *C. stoebe*. A significant positive correlation exists between geographic distance and linearized F_{ST} in Europe, while no relationship exists in North America. Note the difference in scale on both axes between the graphs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

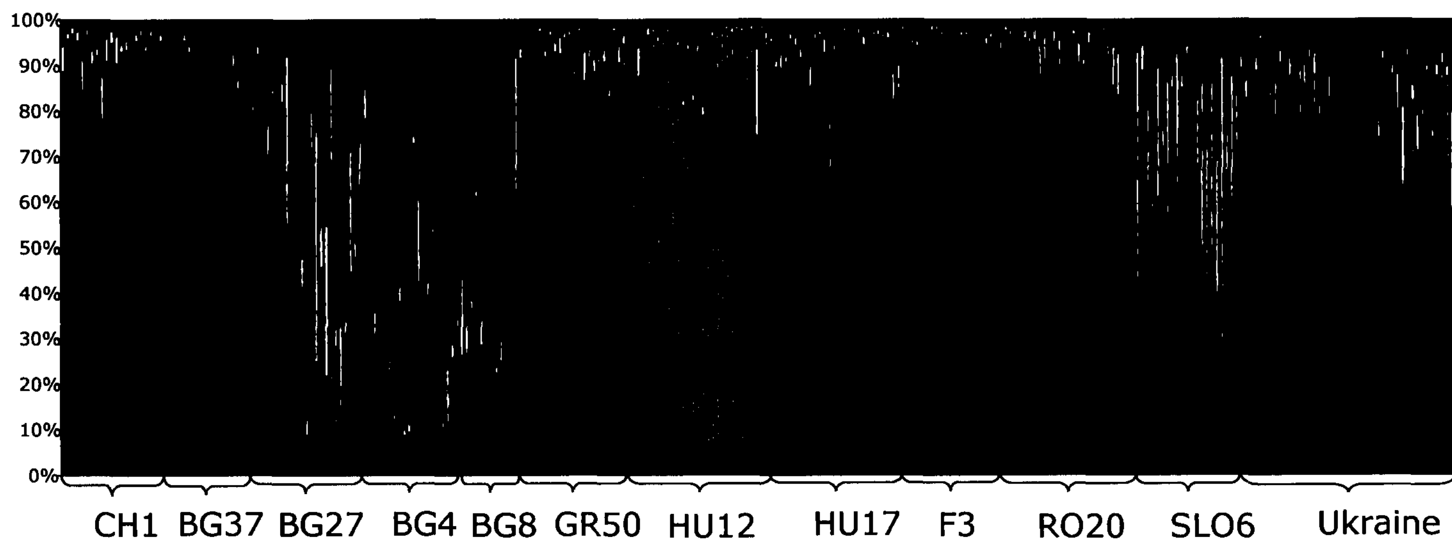


Figure 5.4:
Assignment test results: percentage assignment of each European individual (represented by vertical bars) to each of the 6 genetic clusters (represented by different colors) inferred by the program Structure (Pritchard *et al.* 2000). Site codes (Table 5.1) indicate the geographic location of individuals along the x-axis. Non-*C. stoebe* individuals included.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

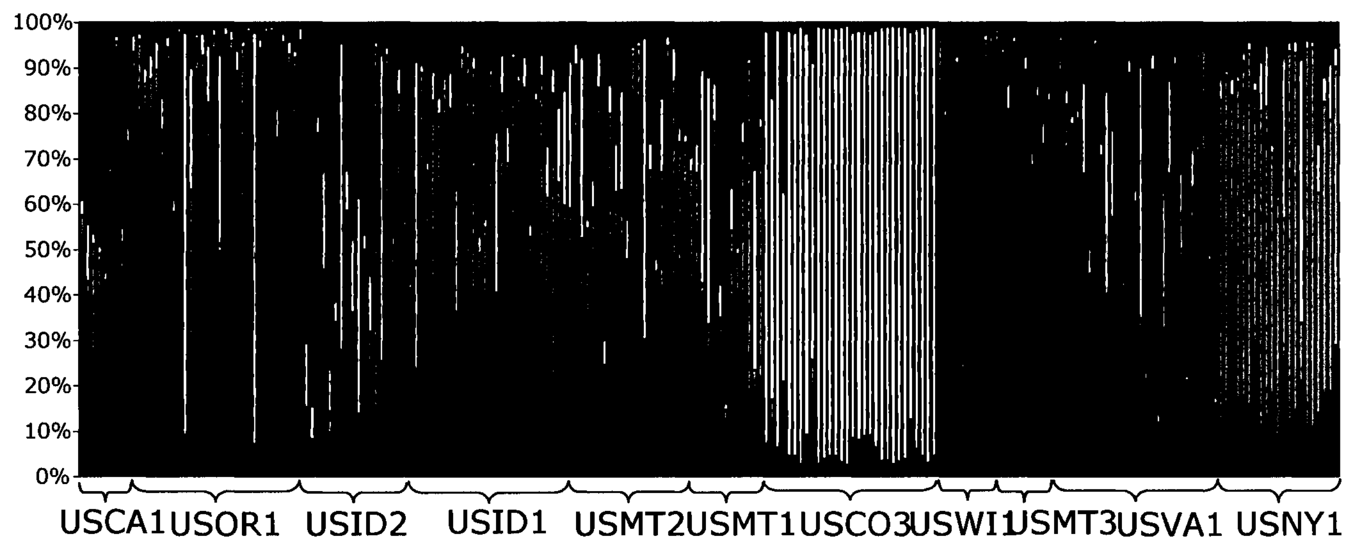


Figure 5.4, continued:
Assignment test results: percentage assignment of each North American individual (represented by vertical bars) to each of the 6 genetic clusters (represented by different colors) inferred by the program Structure (Pritchard *et al.* 2000). Site codes (Table 5.1) indicate the geographic location of individuals along the x-axis.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

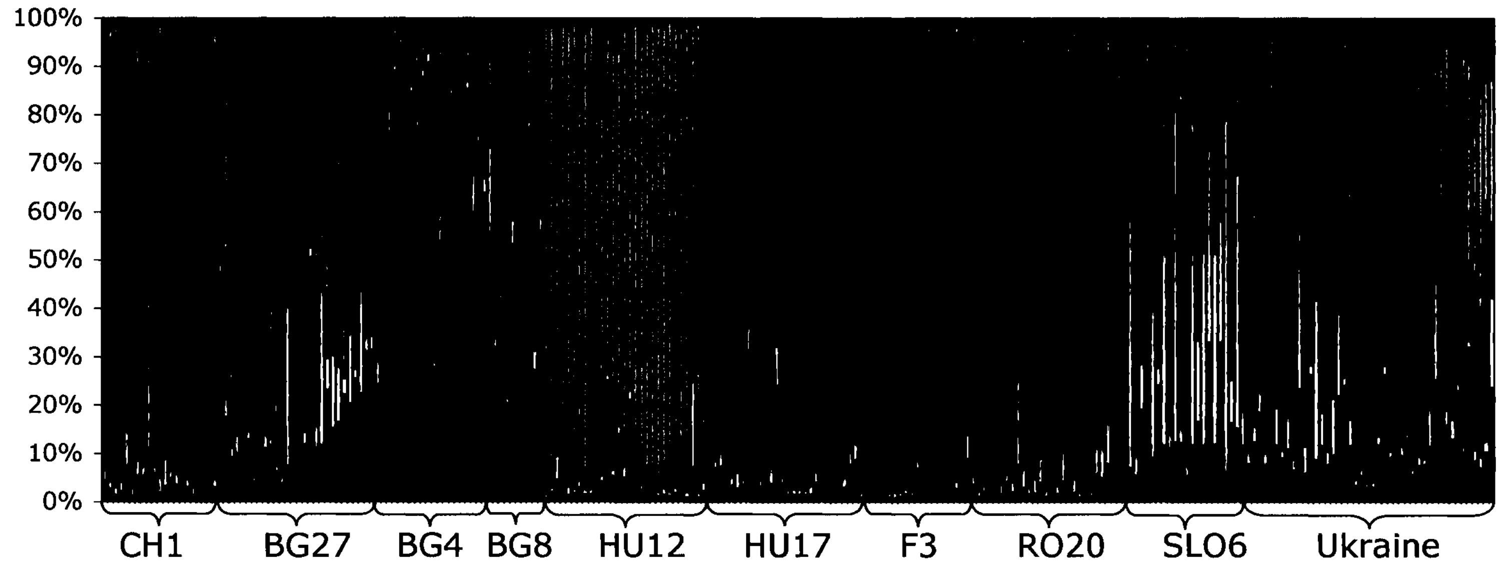


Figure 5.5:
Assignment test results: percentage assignment of each European individual (represented by vertical bars) to each of the 6 genetic clusters (represented by different colors) inferred by the program Structure (Pritchard *et al.* 2000). Site codes (Table 5.1) indicate the geographic location of individuals along the x-axis. Non-*C. stoebe* individuals excluded.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

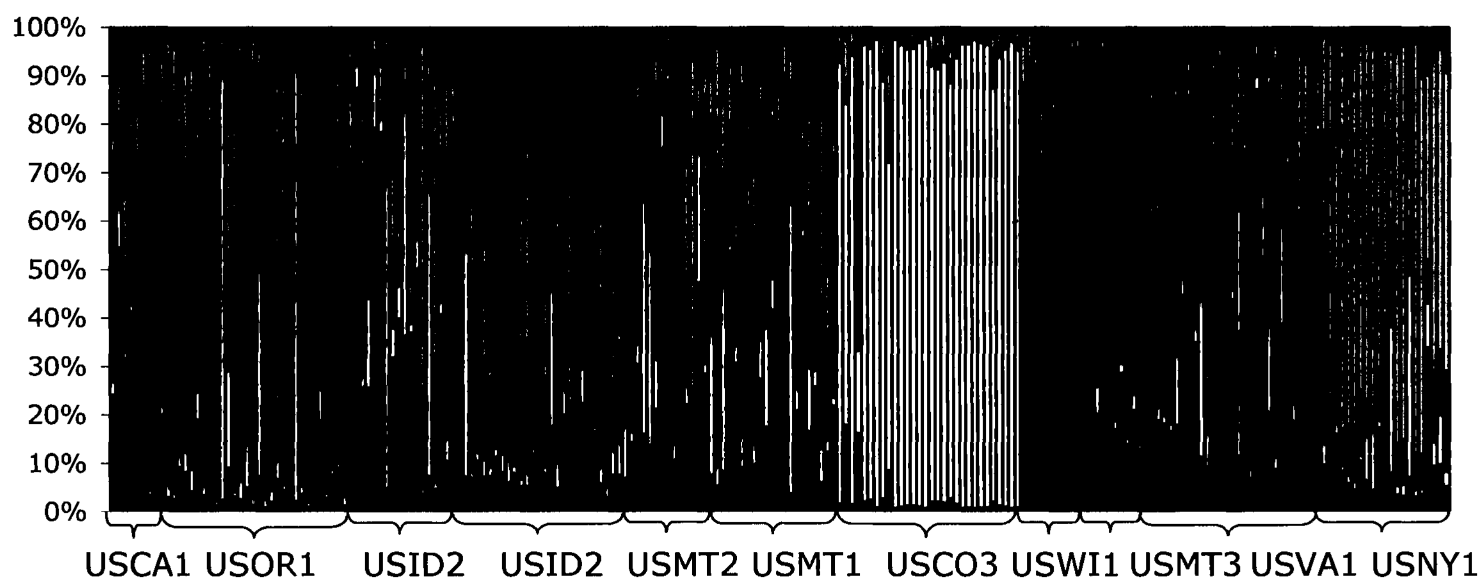


Figure 5.5, continued:
Assignment test results: percentage assignment of each North American individual (represented by vertical bars) to each of the 6 genetic clusters (represented by different colors) inferred by the program Structure (Pritchard *et al.* 2000). Site codes (Table 5.1) indicate the geographic location of individuals along the x-axis. Non-*C. stoebe* individuals excluded.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

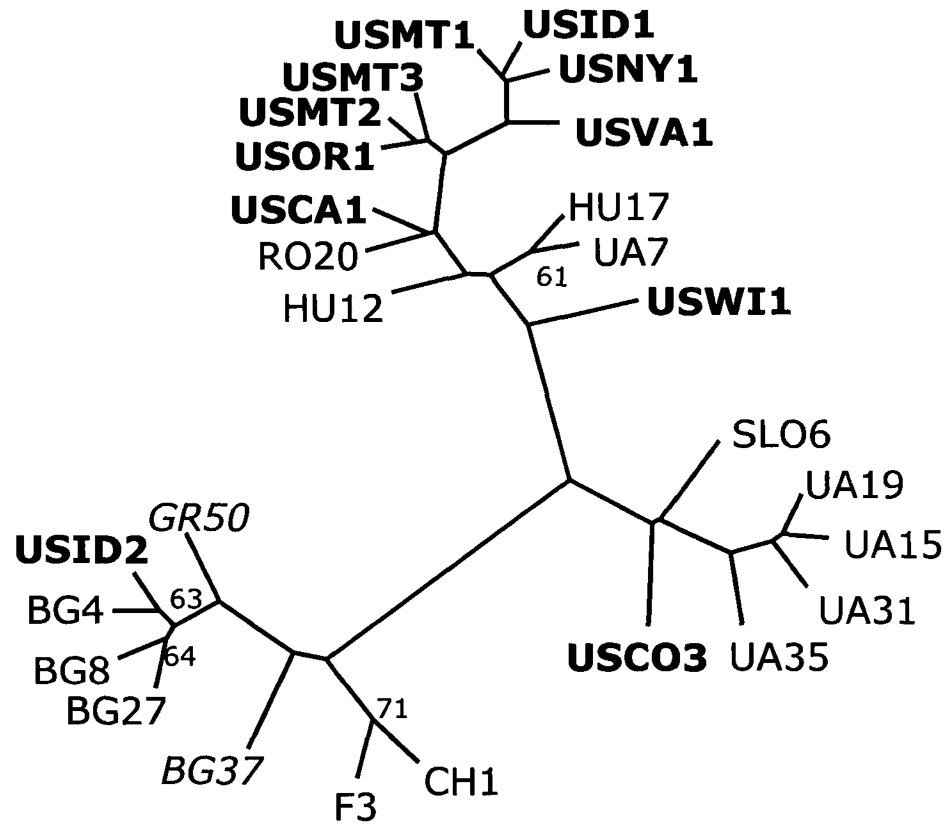


Figure 5.6:
Unrooted UPGMA tree constructed using Nei's standard genetic distance (1972). European site codes (Table 5.1) are given in regular text, while North American site codes are in bold. Italicized codes indicate the two other subspecies (GR50 = *C. grisebachii confusa,* BG 37 = *C. spp.* unknown). Bootstrap values greater than 50% are given.