THE INVASION HISTORY OF THE PARTIALLY CLONAL INVADER, KUDZU (*PUERARIA MONTANA VAR. LOBATA)*

By

KERIN E. BENTLEY

(Under the Direction of Rodney Mauricio)

ABSTRACT

Due to the future effects of climate change and anthropogenic activities that change habitat, it is of high priority from a conservation perspective to identify how species can adapt quickly to shifts in biotic and abiotic conditions. Invasive species serve as excellent model species for identifying the genetic, evolutionary, demographic and ecological variables that can lead to high fitness in new environments. These species are introduced to a new area and within relatively short time frames exhibit rapid population expansion and geographic spread. They are the success stories for thriving in a new environment, but the mechanisms responsible for adaptation are still unclear. Studies of invasion have highlighted the importance of high genetic diversity present in the invasive range, which is now generally attributed to multiple introductions or high diversity in founders. However, with high diversity there are several evolutionary mechanisms that could produce adaptation to the environment. This has sparked the debate as to how frequently and to what degree natural selection versus demographic process or stochastic events like genetic drift affect the evolutionary trajectory of an introduced species. With the application of next generation technologies to these non-model species, more cases of selection are becoming available to answer these questions, but these studies are still rare. There

is also increasing evidence that the mode of reproduction can greatly affect the distribution of genetic diversity in an invasion, influencing the ability of species to invade, but this issue is highly understudied.

Cumulative knowledge of invasion histories, including both introduction and evolutionary histories, across species can imply what and how genetic, biological or evolutionary factors may influence invasion. I reconstructed the invasion history of a highly successful invader, kudzu or *Pueraria montana var. lobata,* by comparing neutral genetic diversity patterns within and among the native and invasive ranges to make inferences about genetic bottlenecking, the effect of a mixed breeding system on genetic diversity, population structure and gene flow as well as test for evidence of natural selection. I found contrasting evidence of the amount of genetic diversity in the invasive range. However, there was consistent evidence across multiple marker types that kudzu was introduced into the United States multiple times from disparate sources across East Asia. These populations exhibited evidence of more sexual reproduction compared to the native range, but upon introduction into the United States, kudzu increased its rate of clonal reproduction, even higher than the rest of the native range. We detected possible evidence of selection on standing genetic variation in the putative source populations prior to the introduction of kudzu into the United States.

INDEX WORDS: Kudzu, population genetics, invasion, RADseq, microsatellite, partially clonal reproduction, genome scan, natural selection

THE INVASION HISTORY IN THE PARTIALLY CLONAL INVADER, KUDZU (*PUERARIA MONTANA VAR. LOBATA*)

by

KERIN E. BENTLEY

B.A., Vanderbilt University, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPY

ATHENS, GEORGIA

2015

THE INVASION HISTORY IN THE PARTIALLY CLONAL INVADER, KUDZU (*PUERARIA MONTANA VAR. LOBATA*)

by

KERIN E. BENTLEY

Major Professor:	Rodney Mauricio
Committee:	Travis Glenn Kelly Dyer Xiaoyu Zhang Shumei Chang

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

DEDICATION

I dedicate this dissertation to the late Dr. David E. McCauley. He was an exceptional scientist of population genetics in plants and he convinced me that a career in science was the most exciting career of all.

ACKNOWLEDGEMENTS

With respect to Chapter 1, the author thanks E. Bentley, G. Bentley, G-Q. Chen, H. Dahn, J. Haase, K. Hickman, S. Hoffberg, R. Kaul, J. Lee, M. Mauricio and K. Myhre for assistance with collecting. This material is based upon work supported by the National Science Foundation's Partnerships for International Research and Education (PIRE) program under grant no. OISE 0730218 as well as a National Science Foundation Graduate Research Fellowship under grant no. DGE 0903734.

With respect to Chapter 2, the author also thanks Blake Johnson, Heather Donaldson and Martha Marks for help with DNA extraction and genotyping; Sandra Hoffberg, Jared Lee, and Travis Glenn for experimental advice; Benjamin Law for help in figure preparation; Adam Bewick, Kathleen Pieper, and Michael McKain for bioinformatics help; and Robert Kaul, Karen Hickman, Guoqi Chen, Gregory Bentley, Elizabeth Bentley, Megan Bentley, and Edward Bentley for help in collection of samples.

With respect to Chapter 3, the author also would like to acknowledge the help of Sandra Hoffberg for assistance and support applying the 3RAD method, Sarah Sander for help with scripting and development of the STACKS pipeline for this study, Julian Catchen for updating the program STACKS, to accommodate 3RAD libraries, Benjamin Law for help in figure preparation, Katie Pieper for comments on the manuscript, Troy Kieran in normalizing the genomic libraries, the Georgia Genomics Facility for helping to meet sequencing needs, and the Georgia Advanced Computing Resource Center, particularly Yecheng Huang, for bioinformatics support.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

With increased biological invasions associated with global trade (Weber and Li 2008), degradation of suitable habitat from anthropogenic activities (Hoekstra et al. 2005), and large-scale environmental shifts from climate change that alter abiotic requirements over time and space (Chapin III et al. 2000), there is pressing need to understand how species adapt to new environments as well as maintain or change their geographic range limits. Invasive species serve as excellent model systems as they are an example of rapid evolution (Lee 2002), observable within decades (Huey et al. 2000). Typically, invasive species are introduced to a new environment, and, after a lag period, exhibit high fitness in the new habitat and begin a rapid population expansion to spread into new areas (Sakai et al. 2001). Understanding the genetics of invasion might illuminate the role of the evolutionary mechanisms that enable rapid adaptation to changing environments (Moran and Alexander 2014).

In this dissertation, I have studied the evolutionary genetic patterns of invasion in kudzu, *Pueraria montana var. lobata*. Kudzu is an invasive species whose invasion is not genetically well characterized but could illustrate the evolutionary history of a highly successful invasion. I have combined knowledge of the introduction history from historical accounts and utilized multiple molecular markers distributed across the genome and hundreds of populations across the native and invasive range to provide a more complete history of the movement of kudzu from its native range in Asia to the eastern United States that has been available to date.

Kudzu has spread to 5 million hectares (Blaustein 2001), since its introduction into the US from Asia in the late 1800's (Miller and Edwards 1983), outcompeting other native species on local scale across a myriad of habitats (Forseth and Innis 2004). Furthermore, kudzu appears to have undergone phenotypic differentiation from the native range, leading to aggressive invasiveness in the US. Such a phenotypic difference may be due to one or a suite of adaptations although no one has yet investigated the genetic basis of this invasive phenotype in kudzu. The United States and parts of its native range in Asia share similar biogeography (Jenkins and Mooney 2006), yet the transgressive phenotype of kudzu within the invasive range seems to be divergent from much of its native range. Studying the invasion history of kudzu can shed light on how kudzu has achieved rapid spread and adaptation in less than 150 years.

To determine possible evolutionary changes during the invasion process, a thorough assessment of the amount and distribution of genetic diversity in the invasive range must be established (Sakai et al. 2001). Without genetic variation, selection cannot result in evolutionary change (Fisher 1930). There are many ways in which genetic diversity can lead to increased fitness in the invasive range. For example, gene flow can enable low frequency alleles in the native source populations to have a higher likelihood of rising to high frequencies and enhance the potential for standing genetic variation to gain a novel advantageous function (Kolbe et al. 2004, Kolbe et al. 2008, Ascunce et al. 2011). Furthermore, heterotic multilocus genotypes can be produced from the population admixture of previously isolated populations now exchanging gene flow in the exotic range (Ellstrand and Schierenbeck 2000). Infusion of heterozygosity and exchange of different alleles from historically isolated populations might lead to hybrid vigor and reduced sensitivity to environmental stochasticity in the resulting generations (Robertson and Reeve 1952). Adaptation is even possible with small founder events (Dlugosch and Parker

2008), if the founders hold high levels of heterozygosity or contain diversity representative of the native range (Wares et al. 2005), or maintain fitness-related alleles in balancing selection after the initial bottleneck due to their intermediate frequency (Carson 1990). Previous studies have found adaptive changes across many fitness-related traits during invasion (Siemann and Rogers 2001, Blair and Wolfe 2004, Brown and Eckert 2005, Dlugosch and Parker 2007, Lavergne and Molofsky 2007, Vandepitte et al. 2014), but evolutionary histories are variable (Keller and Taylor 2008). Full characterization of population genetics of the invasion is the foundation upon which the importance of the amount, nature and distribution of genetic variation and its evolutionary impact can be assessed (Wares et al. 2005).

The first chapter of this dissertation evaluated the contemporary patterns of neutral genetic diversity across the introduced range in kudzu and discussed implications for the introduction and spread of this invasive species. The clonality of kudzu introduced challenges to population genetic analyses that might bias the assessment of spatial patterns of distinct multilocus genotypes (Arnaud-Haond et al. 2007); however, I assessed patterns of diversity while taking this into account. To perform a fine-scale assessment of the invasive range, I screened 87 populations in the invasive range with 15 microsatellite markers and a 789 bp chloroplast region. Detailed genetic and clonal analyses were performed, testing levels of diversity, population structure and phylogeographic relationships among lineages. I found that kudzu exhibited genetic evidence of clonal reproduction and exhibited higher heterozygosity compared to other long-lived perennials (Hamrick and Godt 1996). Only a few clonal lineages and haplotypes were present in the introduced range, implying that a large single introduction or few at most were involved in this invasion. Introduced lineages in the United States were geographically randomly distributed but isolated, suggesting that genotypes rarely expand into

already established populations. This lack of connectedness among pre-existing populations and evidence of clonality imply that population admixture was not important in this invasion. While it has been suggested that recombination between different source genotypes is the driving force behind invasion rather than the amount of genetic diversity levels *per se* (Dlugosch and Hays 2008), clonal reproduction appears to have inhibited this mechanism.

However, detailed knowledge of genetic patterns in the invasive range need to be put into context with the native range in order to determine whether genetic, and or clonal diversity patterns have changed upon introduction (Estoup and Guillemaud 2010, Handley et al. 2011). In the second chapter of this dissertation, I describe a detailed study of the population genetics of kudzu in the native range in Asia and how it compares to the invasive range. In addition to the 87 invasive populations in the first chapter of this dissertation, I screened an additional 142 native populations from across China, South Korea and Japan, a cumulative total of almost 5,000 individuals from across the entire native and invasive range. This dataset was able to better characterize the introduction history and reproductive biology of this species during invasion. Haplotype assignment, genetic clustering analysis, and tracing of clonal lineages back to Asia confirmed that kudzu was actually introduced to the United States multiple times. It was generally thought that species became invasive despite a lack of genetic diversity from small founder events, creating the historical "paradox" of invasion (Allendorf and Lundquist 2003). However, the frequency and consequences of multiple introductions has enabled scientists to show that the amount of genetic diversity increases proportionally with the number of introduction events (Kolbe et al. 2004). Consequently, no matter the origin of the source populations, multiple introductions can mitigate the negative effects of genetic bottlenecks (Novak and Mack 2005) and are common in invasion histories (Ellstrand and Schierenbeck

2000, Bossdorf et al. 2005). In fact, variation is often maintained or increased during the invasion process (Maderspacher 2011), generally depending on the number of introductions (Dlugosch and Parker 2008). It appears that kudzu is no exception: this study suggested that multiple introductions of highly heterozygous individuals did not suffer from severe genetic bottlenecks.

Additionally, different modes of reproduction alter how genetic diversity can be distributed across genomes (Stebbins 1957, Barrett 2011), and the comparison of genetic variation among ranges suggested that the significant levels of clonality of kudzu likely had a significant impact on the evolutionary history of this species. Patterns and differences in diversity within and among ranges indicated a shift to increased rates of vegetative reproduction in the US compared to the more sexual putative source populations as well as the native range as a whole. Increased clonal reproduction ensures the persistence of co-adapted gene complexes rather than the continuous production of new genetic variants over time from hybridization (Stebbins 1957, Barrett et al. 2008). Consequently, while adaptive evolution can either stem from selection on adaptive standing genetic variation in the exotic range or new mutation (Barrett and Schluter 2008, Prentis et al. 2008), the likelihood of new alleles reaching high frequencies across multiple genetic backgrounds is heavily influenced by the frequency of asexual versus sexual reproduction (Barrett et al. 2008). Unless an invasion was a case of pre-adaptation of standing genetic variation to the novel environment, clonal reproduction would constrain the evolvability of the species (Barrett et al. 2008). Alternatively, sexual reproduction could prevent adaptation by breaking up co-adapted gene complexes suited to the environment (Crow and Kimura 1965). In this case, it would advantageous to switch to clonal reproduction to avoid this issue (Vallejo-Marín et al. 2010). Such an advantage may have been the case for

kudzu. Specifically, the putative source populations identified in this comparative study were significantly more sexually reproducing and thus more diverse, but upon introduction to the United States, kudzu shifted to increased clonal reproduction not seen in its original range. This could be a means by which kudzu could maintain highly adapted multilocus genotypes in this new environment.

After determination of patterns of diversity and their evolutionary implications, evolutionary mechanisms that lead to successful invasions still need to be established. While it is possible to observe high fitness in the novel range from phenotypic plasticity (Ghalambor et al. 2007) or random changes in allele frequencies from genetic drift (Keller and Taylor 2008), natural selection is believed to an important aspect of producing invasiveness (Lavergne and Molofsky 2007). With the recent development of next generation technology and methods that enable genome-wide analyses in non-model species, it is now possible to investigate the signatures of natural selection in any genome (Ekblom and Galindo 2011). A population genetics approach to scanning for selection can identify loci or linked loci that underwent a selective sweep (Storz 2005), specifically non-neutral loci that exhibit population differentiation from the average genomic baseline levels (Beaumont and Balding 2004). This baseline is useful in avoiding spurious findings as the global average accounts for demographic or evolutionary processes that may have affected the entire genome (Hohenlohe et al. 2010). Founding populations can become adapted from natural selection on standing genetic variation that have new beneficial function or selection on new mutation (Prentis et al. 2008). The rate of fixation of an adaptive allele depends on the effective population size, strength of selection and initial frequency of the adaptive allele (Hartl and Clark 1997). As new mutations are expected to be rare and start out at a frequency of 1/2N, assuming diploidy, it is possible for new mutations to

take longer to reach high frequency as compared to standing genetic variation (Barrett and Schluter 2008). Conversely, standing variation from the native range can reach fixation faster because alleles already in existence have a higher initial frequency in the introduced range (Barrett and Schluter 2008). Our ability to detect each event depends on the magnitude of the selective sweep (Storz 2005). Sweeps from pre-existing alleles in the native range or sweeps that are historically old can be difficult to identify under the model of genetic hitchhiking (Barrett and Schluter 2008). Alleles at any outlier loci that are detected may be adaptive, linked to a target of selection, swept to fixation by genetic drift (Storz 2005, Bierne et al. 2011). These alleles serve as candidates of adaptation and deserve further study to confirm their role in the kudzu invasion.

By comparing genomic diversity between native populations that contributed to the invasion and those that did not in addition to a comparison between those putative source populations and the invasive range, my third chapter utilizes an F_{ST}-based genome scan for selection (Foll and Gaggiotti 2008) to simultaneously assess the timing and nature of adaptive change among 52 native and 31 invasive populations of kudzu which were previously characterized in Chapter 3 of this dissertation. Loci with significant deviation from the genomic average represent potential products of non-neutral processes and can be used to identify possible targets of selection (Storz 2005). Targets of selection in invasive kudzu could be associated with genes underlying clonal reproduction to explain the increased clonal reproduction detected in the invasive range, or evolution could have occurred at genes underlying traits unrelated to reproduction. The shift to higher levels of vegetative reproduction in this case may then be a means to preserve adaptive genotypes in those traits. Furthermore, natural selection could have also acted before or after introduction to the United States (Huey et

al. 2000, Vandepitte et al. 2014). The genome scan identified 27 outlier loci when comparing putative source populations and the rest of the native range. Most loci represented significant changes in standing genetic variation rather than new mutation and implied that evolutionary changes occurred prior to introduction to the United States. I also identified 4 outlier loci between the putative source populations and the invasive range; however, these would have represented soft sweeps due to minor changes in allele frequencies if they were adaptive. If these outlier loci are truly the products of natural selection, it is likely that all outlier loci detected were linked to adaptive polymorphism rather than being the target themselves as these sequences were highly repetitive. Further study of these loci and genomic sequences linked to these regions for detection of any genes that could be responsible for phenotypic trait variation in this invasion history.

I also assessed whether genetic and clonal diversity patterns of the thousands of single nucleotide polymorphisms discovered mirrored the findings using microsatellite markers between the native and invasive ranges. Patterns of single nucleotide polymorphisms among diversity measures confirmed higher clonal reproduction in the invasive range compared to Asia and evidence of higher sexual reproduction in the putative source populations, but kudzu was remarkably less heterozygous in all groups using this marker type, calling into question the importance of gene diversity in shaping this invasion history.

References

Allendorf, F. W. and L. L. Lundquist. (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17:24-30.

Arnaud-Haond, S., C. M. Duarte, F. Alberto, and E. A. Serrão. (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology* 16:5115-5139.

Ascunce, M. S., C. C. Yang, J. Oakey, L. Calcaterra, W. J. Wu, C. J. Shih, J. Goudet, K. G. Ross, and D. W. Shoemaker. (2011) Global Invasion History of the Fire Ant Solenopsis invicta. *Science* 331:1066.

Barrett, R. D. H. and D. Schluter. (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23:38-44.

Barrett, S. C. (2011) Why reproductive systems matter for the invasion biology of plants. 1st edition. Blackwell Publishing Ltd, Hoboken.

Barrett, S. C. H., R. I. Colautti, and C. G. Eckert. (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* 17:373-383.

Beaumont, M. A. and D. J. Balding. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13:969-980.

Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* 20:2044-2072.

Blair, A. C. and L. M. Wolfe. (2004) The evolution of an invasive plant: an experimental study with Silene latifolia. *Ecology* 85:3035-3042.

Blaustein, R. (2001) Kudzu's invasion into Southern United States life and culture. *The Great Reshuffling: Human dimensions of invasive species*:55-62.

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

Brown, J. S. and C. G. Eckert. (2005) Evolutionary Increase in Sexual and Clonal Reproductive Capacity during Biological Invasion in an Aquatic Plant Butomus umbellatus (Butomaceae). *American Journal of Botany* 92:495-502.

Carson, H. L. (1990) Increased genetic variance after a population bottleneck. *Trends in Ecology & Evolution* 5:228-230.

Chapin III, F. S., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, and S. E. Hobbie. (2000) Consequences of changing biodiversity. *Nature* 405:234-242.

Crow, J. F. and M. Kimura. (1965) Evolution in sexual and asexual populations. *American Naturalist:*439-450.

Dlugosch, K. and I. Parker. (2007) Molecular and quantitative trait variation across the native range of the invasive species Hypericum canariense: evidence for ancient patterns of colonization via pre adaptation? *Molecular Ecology* 16:4269-4283.

Dlugosch, K. and I. Parker. (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17:431-449.

Dlugosch, K. M. and C. G. Hays. (2008) Genotypes on the move: some things old and some things new shape the genetics of colonization during species invasions. *Molecular Ecology* 17:4583-4585.

Ekblom, R. and J. Galindo. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107:1-15.

Ellstrand, N. C. and K. A. Schierenbeck. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* 97:7043.

Estoup, A. and T. Guillemaud. (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* 19:4113-4130.

Fisher, R. A. (1930) The genetical theory of natural selection: a complete variorum edition. Oxford University Press.

Foll, M. and O. Gaggiotti. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977.

Forseth, I. N. and A. F. Innis. (2004) Kudzu (Pueraria montana): history, physiology, and ecology combine to make a major ecosystem threat. *Critical reviews in plant sciences* 23:401-413.

Ghalambor, C., J. McKay, S. Carroll, and D. Reznick. (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* 21:394-407.

Hamrick, J. and M. Godt. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical transactions: biological sciences* 351:1291-1298.

Handley, L.-J. L., A. Estoup, D. Evans, C. Thomas, E. Lombaert, B. Facon, A. Aebi, and H. Roy. (2011) Ecological genetics of invasive alien species. *BioControl* 56:409-428.

Hartl, D. L. and A. G. Clark. (1997) Principles of population genetics. Sinauer associates Sunderland.

Hoekstra, J. M., T. M. Boucher, T. H. Ricketts, and C. Roberts. (2005) Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters* 8:23-29.

Hohenlohe, P. A., P. C. Phillips, and W. A. Cresko. (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences* 171:1059.

Huey, R. B., G. W. Gilchrist, M. L. Carlson, D. Berrigan, and L. Serra. (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science* 287:308.

Jenkins, P. T. and H. A. Mooney. (2006) The United States, China, and invasive species: present status and future prospects. *Biological Invasions* 8:1589-1593.

Keller, S. R. and D. R. Taylor. (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters* 11:852-866.

Kolbe, J. J., R. E. Glor, L. R. Schettino, A. C. Lara, A. Larson, and J. B. Losos. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177-181.

Kolbe, J. J., A. Larson, J. B. Losos, and K. De Queiroz. (2008) Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biology Letters* 4:434.

Lavergne, S. and J. Molofsky. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences* 104:3883.

Lee, C. E. (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* 17:386-391.

Maderspacher, F. (2011) The benefits of bottlenecks. *Current Biology* 21:R171-R173.

Miller, J. H. and B. Edwards. (1983) Kudzu: Where did it come from? And how can we stop it? *Southern Journal of Applied Forestry* 7:165-169.

Moran, E. V. and J. M. Alexander. (2014) Evolutionary responses to global change: lessons from invasive species. *Ecology Letters* 17:637-649.

Novak, S. and R. Mack. (2005) Genetic bottlenecks in alien plant species: influence of mating systems and introduction dynamics. *Species invasions: insights into ecology, evolution, and biogeography*:201-228.

Prentis, P. J., J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. (2008) Adaptive evolution in invasive species. *Trends in Plant Science* 13:288-294.

Robertson, F. W. and E. Reeve. (1952) Heterozygosity, environmental variation and heterosis.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, and N. C. Ellstrand. (2001) The Population Biology of Invasive Species. *Annual Review of Ecology and Systematics*:305-332.

Siemann, E. and W. E. Rogers. (2001) Genetic differences in growth of an invasive tree species. *Ecology Letters* 4:514-518.

Stebbins, G. L. (1957) Self fertilization and population variability in the higher plants. *American Naturalist* 91:337-354.

Storz, J. F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14:671-688.

Vallejo-Marín, M., M. E. Dorken, and S. C. H. Barrett. (2010) Ecological and Evolutionary Consequences of Clonality for Plant Mating. *Annual Review of Ecology, Evolution, and Systematics* 41.

Vandepitte, K., T. Meyer, K. Helsen, K. Acker, I. Roldán-Ruiz, J. Mergeay, and O. Honnay. (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology* 23:2157-2164.

Wares, J. P., A. R. Hughes, and R. K. Grosberg. (2005) Invasive species and evolutionary processes. Pages 229-257 *in* J. S. D. Sax, and S. D. Gaines, editor. Species Invasions: Insights into Ecology, Evolution and Biogeography Sinauer Publishing.

Weber, E. and B. Li. (2008) Plant invasions in China: What is to be expected in the wake of economic development? *Bioscience* 58:437-444.

CHAPTER 2

CONTEMPORARY PATTERNS OF NEUTRAL GENETIC DIVERSITY ACROSS THE INTRODUCED RANGE OF A HIGHLY CLONAL VINE, KUDZU (*PUERARIA MONTANA VAR. LOBATA)*: IMPLICATIONS FOR UNDERSTANDING THE INTRODUCTION AND SPREAD OF AN INVASIVE SPECIES[1]

[1] Bentley, K.E. and Mauricio, R. Submitted to American Journal of Botany on October 11, 2015.

Abstract

Premise of the study

Pueraria montana var. lobata, or kudzu, is an invasive species whose invasion is not genetically well characterized but could illustrate the evolutionary history of a highly successful invasion. The clonality of kudzu introduces challenges to population genetic analyses that might bias the assessment of spatial patterns of distinct multilocus genotypes. Assessing patterns of genetic diversity while taking this into account is necessary to understand the invasion and spread of kudzu in its North American invasive range.

Methods

We screened 87 populations in the invasive range with 15 microsatellite markers and a 789 bp chloroplast region. We performed detailed clonal analyses and tested levels of genetic diversity, population structure and phylogeographic relationships among lineages.

Key results

Kudzu exhibited a clonal rate of 80%, and was more heterozygous than other long-lived perennials. Only 353 distinct clonal lineages were detected with over 60% sharing the same maternal haplotype. Populations were established with few genotypes, many being monoclonal. No route of expansion was detected as the spread of genotypes lacked spatial patterning.

Conclusions

Kudzu is highly clonal, but exhibits high heterozygosity. Only a few clonal lineages and haplotypes exist in the introduced range. Our data are consistent with a large single introduction or few at most. Introduced lineages are geographically randomly distributed but isolated, suggesting that genotypes rarely expand into already established populations.

Key words: Clonal reproduction; Invasive species; Kudzu; Microsatellite

Introduction

Invasive species provide an excellent model for understanding evolutionary processes (Reznick and Ghalambor, 2001). Individuals are moved to new locations and, depending on a number of basic population genetic parameters, might have the opportunity to adapt and spread. Thus, population genetics can provide a powerful forensic tool for understanding the geographic distribution of species. Although genetic variation in the invasive range depends on the introduction history and life history traits of a species, patterns of genetic diversity can provide details as to the number and size of introduction events, genetic bottlenecks, the dynamics of population establishment, the route of spread, and patterns of reproduction (Barrett, Colautti and Eckert, 2008). This information is central to understanding the process of invasion (Estoup and Guillemaud, 2010).

Pueraria montana (Lour.) Merr. var. *lobata* (Willd.) Maesen & Almeida (Ward, 1998), or kudzu, is an example of an invasive species where contemporary patterns of genetic diversity might provide some clues as to the evolutionary history of the invasion. Kudzu is an iconic and significant invasive species in North America, causing substantial ecological disruption (Forseth and Innis, 2004; Hickman et al., 2010) as well as substantial economic impact (Mitich, 2000).

Two aspects of kudzu's biology provide additional ability to detect genetic patterns of invasion: it is a long-lived perennial and it is at least partially clonal. Thus, it is likely that relatively few generations have passed since its introduction into North America, and present-day genetic diversity might reflect genetic patterns that were produced during colonization and expansion. Such contemporary remnants of the past could be used to understand evolutionary history more clearly.

It has been widely reported that kudzu was first introduced into North America as an ornamental from Japan in 1876 and then again in quantity to the southeastern United States in the 1930's and 1940's to control erosion (Lindgren et al. 2013). United States government records from the Soil Conservation Service describe the cultivation and large-scale distribution of kudzu across the southeastern United States during that period (Stewart, 1997): one hundred million seedlings were propagated and distributed within 7 years (Alderman and Alderman, 2001). However, the introduction history, including the number, size, source, and route of expansion into the United States, is not well established.

High genetic diversity observed in the invasive range of kudzu has led several researchers to infer that multiple introductions (Pappert, Hamrick and Donovan, 2000; Jewett et al., 2003; Sun et al., 2005), likely from different sources, were required since genetic diversity generally increases with the number of introductions (Kolbe et al., 2008). However, there is another possible cause of high genetic diversity in the invasive range. High genetic diversity could be observed from one large or multiple introductions from a single genetically diverse region of the native range.

Due to the intense cultivation and dissemination by root crowns in the early 20th century (Stewart, 1997), it is possible that kudzu was not continuously imported during this time period and the hypothesis of multiple introductions from multiple sources may not be correct. Once the Soil Conservation Service established centralized cultivation of kudzu, additional import of seed from Asia would be unnecessary, especially since seed set, germination rates, and seedling survival are very low in kudzu (Abramovitz, 1983) and vegetative propagation is relatively quick (Forseth and Innis, 2004). Thus, the initial native sources of those initial propagules might have been limited.

The fact that kudzu is so easily propagated derives from the fact that it is capable of both vegetative and sexual reproduction, although the frequency of either means of reproduction is unknown. Different reproductive strategies can alter the distribution of genetic diversity and pace of evolution (Barrett, Colautti and Eckert, 2008). Sexual and clonal reproduction distribute genetic variation differently across geographic space and time, affecting gene flow across populations, population admixture, and the potential for rapid adaptation to the colonized environment (Barrett, 2011). Consequently, genetic data can determine the extent to which kudzu clonally reproduces. Pappert *et al.* (2000) observed high clonality in the Southeast, but patterns of reproduction could vary across the entire invasive range. Furthermore, investigating the geographic distribution of clonal lineages could aid in the identification of the number of introductions and the route of expansion, because closely related clonal lineages likely originated from the same source and should be easily traceable in the absence of recombination.

Previous studies have investigated patterns of genetic diversity in the United States (Pappert, Hamrick and Donovan, 2000; Jewett et al., 2003; Sun et al., 2005), but those studies were based on limited geographic range. The range of kudzu has grown from an original 1.2 million hectares in the Southeast (Everest et al., 1999) to over 3 million hectares across the eastern United States (Blaustein, 2001). Furthermore, none of these studies has quantified the number of genetic clusters of kudzu, which can identify the number and size of introductions, or the route of expansion, which can highlight the trajectories of introduced lineages better suited to certain environments.

In this study, we expand upon previous studies with greatly increased sample sizes, broader geographic sampling across the invasive range and a detailed investigation of the patterns of diversity and clonality using both microsatellite and chloroplast sequencing data. We

address the following questions: 1) How clonal is kudzu? 2) How much genetic diversity exists in the entire invasive range? 3) How much gene flow occurs? 4) Can we infer the history of the kudzu invasion? 5) How has kudzu spread?

Materials and Methods

Sampling

We sampled from 87 locations across the invasive range of *Pueraria montana* var. *lobata* in the United States. Collection sites were generally separated by at least 100 km. For each site, we collected approximately 24 young leaves, sampling along a linear transect that ran the length of the population. The distance between samples was proportional to the transect length; generally each sample was separated by several meters. A single GPS coordinate was collected for each population (Appendix 2.1). Leaves were dried for genomic DNA extraction.

Genotyping and sequencing

Each sample was extracted using either a Qiagen DNeasy Kit (Qiagen, Germantown, Maryland) or Axyprep Multisource Genomic DNA Miniprep Kits (Axigen Biosciences, Union City, California) with the centrifuge protocol for plant tissue. Each sample was screened against 15 microsatellite markers (Hoffberg et al., 2014). PCR amplification was based on a 3-primer PCR approach using CAG tagging (Hauswaldt and Glenn, 2003) but modified for kudzu by Hoffberg *et al.* (2014). All primer amplification followed PCR protocols from Hoffberg *et al.* (2014). Amplification was verified with 1.5% agarose gels. Amplicons were diluted 1:6 in deionized water and visualized with an Applied Biosystems-3730xl DNA sequencer (Applied Biosystems, Foster City, California) with a 500 ROX-labeled size standard in each well to ensure accurate allele calls. Genotyping results were scored using GENEMARKER version 2.4 (SoftGenetics, State College, Pennsylvania). All data were run through Micro-Checker (Van

Oosterhout et al., 2004) before starting analysis to correct any errors as well as test for null alleles. Possible null alleles existed in 12%, 13% and 16% of populations for KZ06, KZ46 and KZ50, respectively.

We calculated the genotyping error rate (Pompanon et al., 2005) by randomly choosing 34 individuals without any missing data for PCR amplification of all markers in two blind replicates. Mean error rate per locus and the multilocus genotype error rate were calculated based on allele mismatches between the first and second allele scores identified through custom Perl scripts. Missing data were excluded from the error rate calculations. The average error rate per locus was 9.80×10^{-3}, while the multilocus genotype error rate was 8.82×10^{-2}, which is well below error rates for other microsatellite-based studies (Selkoe and Toonen, 2006). While error was infrequent, 6 of the 10 allelic mismatches were due to mistaken allele calling in KZ56; therefore, all alleles were re-scored for this marker with consistent scoring rules to minimize this issue.

Additional genetic data were collected from sequencing a single 789 bp intergenic spacer chloroplast region, rpoB-trnC. Primers and the PCR protocol followed Shaw *et al.* (2005) for the universal primer pair, $trnC^{GCA}R$ (CAC CCR GAT TYG AAC TGG GG) and rpoB (CKA CAA AAY CCY TCR AAT TG). The PCR protocol followed a touchdown procedure: 80°C for 5 min, followed by 30 cycles of 96°C for 1 min, 50°C for 2 min, 72°C for 3 min, followed by 72°C for 5 min. PCR product of sequences was prepared according to Glenn and Schable (2005) and sequenced using an Applied Biosystems 3730xl 96-capillary DNA Analyzer. Nucleotide calling and quality of raw chloroplast sequence data were verified in Sequencher 4.7 and aligned using default parameters with MUSCLE (Edgar, 2004) in Geneious version 7. Any detected

polymorphisms were confirmed both in the chromatograms and the alignment to reduce scoring error.

Assessing clonality

Clonal species present a special challenge to DNA-based genetic marker analysis because it is easy to overestimate the number of unique multilocus genotypes; scoring error, for example, might lead us to score two identical individuals as different if the samples differ at any position. Therefore, we used GENODIVE (Meirmans and Van Tienderen, 2004) to determine a clonal "threshold" that allowed us to delineate samples as distinct multilocus genotypes or clones. This threshold represents the maximum number of allowable mutational differences between two samples to remain categorized as clones within populations, while accounting for somatic mutation and scoring errors (Douhovnikoff and Dodd, 2003; Meirmans and Van Tienderen, 2004). The most appropriate clonal threshold was located between the first and second peak of a frequency histogram of pairwise genetic distances between samples, generated by a stepwise mutation model that ignored missing data. A full dataset with both distinct multilocus genotypes within populations and their associated clones (identical multilocus genotypes) was used for all clonal analysis.

In order to identify genetic patterns of clonal lineages within and among populations, clonal diversity measures were estimated. The number of unique multilocus genotypes (MLG), effective number of MLGs based on rarefaction (eff), Nei's genetic diversity index corrected for sample size (div), evenness (eve) and the Shannon-Weiner index corrected for sample size (shc) were estimated in GENODIVE. Lastly, the "proportion distinguishable" (R) was calculated as (MLG-1)/(N-1) (Ellstrand and Roose, 1987), where N is sample size. This measure indicates the observed frequency of unique multilocus genotypes in a given group of samples.

Additional evidence of clonal reproduction was investigated by testing for linkage disequilibrium (LD) and the violation of Hardy-Weinberg equilibrium (HWE). Since there is not a single accurate measure of linkage disequilibrium in systems with significant asexual reproduction, de Meeûs and Balloux (2004) recommend the use of both single and multiple loci estimators. Multilocus LD, measured by the index of association (I_A), was tested with 1000 permutations using the software, poppr (Kamvar, Tabima, and Grünwald, 2014). This index uses the ratio of observed to expected variance in the number of loci in which two individuals differ to ascertain to what extent individuals demonstrate recombination indicative of panmixia (Agapow and Burt, 2001). The modified index of association, rbarD, was also measured to account for unequal sample size (Agapow and Burt, 2001). A single locus measure of linkage disequilibrium, was also performed with Bonferroni corrections in GENEPOP (Raymond and Rousset, 1995; Rousset, 2008) for a dataset of 15 populations that had a sample size of at least 7 multilocus genotypes with clones removed so as to maintain sufficient statistical power. The same dataset was used to test for Hardy-Weinberg equilibrium in GENEPOP (Raymond and Rousset, 1995; Rousset, 2008) with 1000 batches.

Genetic diversity analyses

To better understand patterns of genetic variation across the invasive range for both microsatellite and chloroplast markers, clones of multilocus genotypes were excluded from all subsequent analyses. While the inclusion of clones can offer a biologically realistic understanding of population structure, clones can bias the assessment of spatial patterns of distinct multilocus genotypes. For microsatellite data, a single representative genotype with the least amount of missing data was randomly chosen for each clonal lineage within populations.

For chloroplast sequences, if multiple samples from the same clonal lineage in a population were sequenced, only the longest sequence was retained.

Genetic diversity measures of the 15 microsatellite markers were estimated, including the percentage of polymorphic loci (P%), effective number of alleles (N_E), observed heterozygosity (H_O), Nei's unbiased expected heterozygosity (uH_E), the fixation index (F) within populations and the inbreeding coefficient (F_{IS}) within loci using GenAlEx version 6.501 (Peakall and Smouse, 2006, 2012). As allelic richness is the number of alleles expected given the number of genes sampled in a population (Hurlbert, 1971; Petit, El Mousadik, and Pons, 1998; Kalinowski, 2004), allelic richness (A_R) was calculated using a rarefaction method in HP-RARE (Kalinowski, 2005) in order to standardize for uneven sample sizes. Rarefaction was standardized to the smallest population size with an even number of alleles per marker. A least squares regression model was tested between H_O and MLG as well as H_O and N in JMP Pro Version 10 (SAS Institute Inc., Cary, North Carolina) to see if there was a relationship between genetic diversity and the number of individuals sampled. The relationship between the difference between the observed and expected heterozygosities (H_O - uH_E) and MLG was also tested to see if excess observed heterozygosity occurred in populations with fewer genotypes as found in Pappert *et al.* (2000).

To assess genetic diversity in the chloroplast region, the number of haplotypes, the number of polymorphic sites, the number of parsimony informative sites and nucleotide diversity per site were calculated using DnaSP (Librado and Rozas, 2009).

Population structure and gene flow analysis

Evaluation of population structure can identify how genetic diversity is distributed across populations to infer the means of population establishment and gene flow in this species.

Population differentiation was determined using an AMOVA (Excoffier, Smouse and Quattro, 1992) with 999 permutations in GenAlEx v.6.501 (Peakall and Smouse, 2006, 2012). Isolation by distance was tested using a Mantel test with 999 permutations in GenAlEx v.6.501 (Peakall and Smouse, 2006, 2012) with Nei's genetic distance matrix and a kilometer-based geographic distance matrix produced from the Geographic Distance Matrix Generator version 1.2.1 (Ersts, 2013). Using the full dataset, genotype sharing across populations was assessed using GENODIVE (Meirmans and Van Tienderen, 2004), by re-defining clones across populations.

Inferring introduction history and spread

To identify the putative number and size of independent introduction events from the native range, genetic clusters (K) were detected using Discriminant Analysis of Principal Components (DAPC) implemented in the R package DAPC (Jombart, Devillard and Balloux, 2010). This program does not assume Hardy-Weinberg equilibrium or LD, which is more appropriate for partially clonal species than Bayesian-based clustering methods that include such assumptions.

To find the appropriate number of clusters in DAPC, Bayesian Information Criterion (BIC) values were regressed against cluster number (Jombart, 2013) with all principal components included (Appendix 2.2). Only 9 principal components were retained for final cluster assignment based on the alpha-score, encompassing 55.4% of the cumulative variance. All five discriminant analysis eigenvalues were retained. Scatterplots for each pair of axes were generated. Visualization of genetic clustering was achieved with pie charts representing percent membership of that population to a given cluster overlaid on GPS coordinates in ArcMap 10.2 (Environmental Systems Research Institute, Redlands, California), which we call a “structure map”.

Phylogeographic relationships of the invasive range were determined using the chloroplast sequencing data. To view the evolutionary relatedness of different haplotypes, a minimum-spanning haplotype network was created using statistical parsimony with a 95% connection limit and gaps treated as a fifth state in TCS v. 1.21 (Templeton, Crandall and Sing,1992; Clement, Posada and Crandall, 2000). Haplotype networks allow visualization of non-bifurcating genealogical relationships at the population level and can be more appropriate than more traditional phylogenetic methods, such as gene trees, when divergence levels are low (Clement, Posada and Crandall, 2000) as is expected in an invasion. Pie charts representing percent membership of that population to a given haplotype were overlaid on GPS coordinates in ArcMap 10.2 (Environmental Systems Research Institute, Redlands, California) to visualize the distribution of haplotypes across the landscape, which we call a "haplotype map". This map complements the haplotype network to see if there are any geographic patterns to the distribution of haplotypes across the range, and was used in combination with the structure map to determine the number of introductions from the native range.

Similarly, a microsatellite-based neighbor-joining tree, using Nei's genetic distance from gene frequencies generated in GenAlEx version 6.501 (Peakall and Smouse, 2006, 2012), was created with Phylip (Felsenstain, 2005) with 1000 bootstrap replications, to view phylogenetic relationships among genotypes. Bootstrap values less than 50% were collapsed in MEGA (Tamura et al., 2013).

Results

Clonality

Using the clonal threshold provided by GENODIVE, we estimated that 79.7% of the 1747 samples we analyzed across 15 microsatellites were clones of distinct multilocus

genotypes. We found 353 distinct clonal lineages across the entire invasive range. After clonal assignment, the number of unique multilocus genotypes in a population ranged from 1 to 15 with an average of 4 (Appendix 2.1), but more than 5 multilocus genotypes in a population was rare (Figure 2.1). The effective number of multilocus genotypes (eff) per population was 2.20 (Table 2.1). Twenty-two of the 87 populations sampled contained only a single clone. The average "proportion distinguishable" (R) of samples per population was 16.5%, ranging from 0% to 70% (Appendix 2.1).

Kudzu is significantly inbreeding: the average fixation index, F_{IS}, was -0.575. Both indices of association, I_A and rbarD, demonstrated that genotypes within this dataset were not the expected products of panmixia, with values of 1.23 ($P < 0.001$) and 0.10 ($P < 0.001$) respectively, rejecting the null hypothesis of sexual reproduction. Linkage disequilibrium was detected even when clones were removed and the single locus estimate was used. Approximately 5.7% of loci pairs exhibited linkage disequilibrium. KZ57 and KZ68 were linked in 40% of tested populations. Tests for Hardy Weinberg equilibrium indicated that 4.0% locus-population tests deviated significantly from Hardy-Weinberg expectations. The average inbreeding coefficient (F_{IS}) was negative at -0.403 (Table 2.2), with a variance of 0.062.

Genetic diversity

Individuals exhibited observed heterozygosity higher than the expected (with mean H_O = 0.444, uH_E = 0.398, respectively, Table 2.1; Appendix 2.1). The Shannon diversity index and Nei's genetic diversity index were largely concordant with average values of 0.406 and 0.361, respectively (Table 2.1; Appendix 2.3). Observed heterozygosity did not increase with sampling intensity ($F = 0.342$; $df = 86$; $P = 0.5602$). Observed heterozygosity did increase with the number of distinct multilocus genotypes in a population ($F = 5.16$; $df = 86$; $P = 0.026$), although the

linear model only explained 5.7% of the variation. There was no significant correlation between the number of unique multilocus genotypes in a population and the difference between observed and expected heterozygosity ($F = 2.41$; $df = 86$; $P = 0.1241$). Heterozygosity levels significantly deviated across markers ($F = 25.4$; $df = 1304$; $P < 0.001$), ranging from 0.145 to 0.741 with the exception of KZ42 that was almost completely monomorphic ($H_O = 0.004$). Approximately 62% of loci in a given population were polymorphic with the exception of marker KZ42, which was almost monomorphic across the entire range (Table 2.2). There were 8 alleles per locus on average, ranging from 2 to 17 (Table 2.2). Given the low genotypic diversity in a population, allelic richness was low with 1.41 alleles and 1.70 effective alleles on average (Table 2.1).

The 789 bp chloroplast region was successfully sequenced for 271 clonal lineages across 80 populations. There were 14 polymorphic sites, 11 of which were parsimony informative. Nucleotide diversity per site was 1.58×10^{-3}. There were 13 haplotypes detected. There were 1.36 haplotypes per population on average ranging from 1 to 5 haplotypes. All but two haplotypes were only one or two mutational steps away from Haplotype 1 or Haplotype 2 (Figure 2.2). Those two haplotypes, Haplotype 5 and 8, represented only 4% and 1% of sequenced individuals, respectively. Descriptions of mutational differences among haplotypes can be found in Appendix 2.4.

Population structure and gene flow

Kudzu exhibited population structure with 40% of the total genetic variation distributed among populations ($\phi_{ST} = 0.395$). While almost 50% of populations contained two clonal lineages or fewer, the average evenness (eve = 0.658) demonstrated that there was generally equal representation of clones when multiple distinct genotypes occurred within a population (Table 2.1). Only eight of 279 multilocus genotypes (2.3%) were shared across populations.

Three clonal lineages were widespread, making up 27% of the entire dataset. One shared lineage was a long distance dispersal event between Virginia (VA10) and Texas (TX6), a distance of 1559 km. Otherwise, shared clonal lineages were dispersed regionally. Overall, 40 of 87 populations, including 632 samples or 36% of the dataset, contained shared multilocus genotypes with an average of 1.3 shared clones in those populations. No populations contained more than two shared multilocus genotypes.

On the other hand, chloroplast haplotypes were highly shared. Approximately 63% of sequenced individuals shared Haplotype 1 (Figure 2.2). Haplotype 2 and 3 represented 10% and 9% of the data with the rest of the haplotypes making up less than 5%.

Introduction history and spread

While there was no one best K for genetic clustering in DAPC, 6 genetic clusters was the only K in the stabilized range of the regression that produced the highest proportion of successfully assigned clones overall (>96%) (Figure 2.3). The size of each genetic cluster ranged from 15% to 26% with the exception of Cluster 1, which encompassed 5% of the genotyped individuals. Cluster 1 and 5 appeared to be genetically distant from the other clusters (Figure 2.3).

There were no discernible geographic trends in the structure map or haplotype map, including at the edges of the range (Figures 2.4 and 2.5). Similarly, the neighbor-joining tree of microsatellite data formed a large polytomy (Appendix 2.5), providing no evidence as to the relatedness of multilocus genotypes across the invasive range. No isolation by distance was detected (Rxy = -0.018, P = 0.363).

Discussion

High clonality in kudzu

Pueraria montana var. *lobata* is highly clonal throughout its invasive range in the eastern United States, with 80% of our samples sharing identical multilocus genotypes with at least one other sample in their respective populations. Clonal diversity measures indicate that the majority of populations in the invasive range of kudzu are clones of a few distinct multilocus genotypes, with many populations containing just a single clonal lineage. Clonality in invasive kudzu follows a distinctive genetic signature of almost complete vegetative reproduction as described in a simulation study of clonality (Balloux, Lehmann, and de Meeûs, 2003) . Specifically, this signature is defined by the observation of excess observed heterozygosity, highly negative inbreeding coefficients with large variance, low genotypic diversity within populations, yet high allelic diversity across loci.

Our estimate of clonality across the invasive range is consistent with an estimate by Pappert et al. (2000) who measured the clonality of kudzu in the southeastern U.S. at 80% using allozyme markers. These estimates are also consistent with the fact that kudzu has a low sexual reproductive effort for a long-lived perennial (Forseth and Innis, 2004) and low seedling recruitment in the invasive range (Wechsler, 1977; Abramovitz, 1983; Forseth and Innis, 2004).

High clonality might have implications for the history of this invasion since clonal reproduction restricts the means to adapt to a new environment during an invasion (Barrett, Colautti and Eckert, 2008). Population admixture, or the generation of novel multilocus genotypes from different sources highly fit for the new environment is believed to a common means of adaptation in an introduced range (Ellstrand and Schierenbeck, 2000), but is contingent on sexual reproduction. Without admixture, adapted genotypes can arise only through mutation

or standing genetic variation (Prentis et al., 2008). The invasion of kudzu does not seem to have been dominated by a single genotype, thus it is unlikely that a new mutation after introduction could have been shared across all invasive clonal lineages. Pre-adaptation of standing genetic variation and high phenotypic plasticity of this species become more likely candidates in explaining the success of the invasion of kudzu.

Whether there has been a shift in breeding system from the native range to the invasive is not known. Population structure has been reported to be higher in the native range (Sun et al., 2005); however, no studies have investigated the distribution of genetic and clonal diversity in Asia in great detail. Any possible shift towards higher levels of vegetative reproduction might be causally related to the success of the spread of kudzu in the introduced range and should be investigated in future studies.

High genetic diversity in the invasive range—

Our estimates of genetic diversity (H_O = 0.444; shc = 0.406) were significantly higher than those previously reported for kudzu in the United States (H_O = 0.267 in Pappert *et al.* (2000); shc = 0.221 in Sun *et al.* (2005)). Furthermore, observed heterozygosity in kudzu was higher than other microsatellite-based studies of invasive perennial vines, such as *Fallopia japonica* (H_O = 0.321) (Grimsby et al., 2007) and *Mikania micrantha* (H_O = 0.303) (Yan et al., 2011).

Paradoxically, high gene diversity is common for largely asexual species (Ellstrand and Roose, 1987) due to mutation accumulation, where the lack of recombination of alleles creates heterozygote excess across the genome. In fact, the high level of clonality found in this species is possibly driving observed levels of diversity (Balloux, Lehmann, and de Meeûs, 2003).

This excess heterozygosity could have negative ramifications for the fitness of an individual, specifically in the form of genetic load, but research suggests that this may not be the case for invasive species (Bazin et al., 2014). The accumulation of mutations without the ability to purge alleles has the possibility to create genetic load in largely asexual species; however, a previous study found that biomass production in invasive kudzu positively correlated with levels of heterozygosity (Pappert, 1998). Similarly, species with high but not complete asexuality (80-95% clonality), which are most subject to genetic load, were found to be the most invasive with a higher probability of invasion and shorter time to invasion in a recent simulation study (Bazin et al., 2014).

Dynamics of population establishment

Despite substantial range expansion, it appears that large-scale gene flow across populations is rare with 50% of populations containing only 1 or 2 clonal lineages (Appendix 2.1). The high population structure and low genotypic diversity within populations detected in this study imply that individuals do not commonly establish in pre-existing populations. Half of the populations in this study contained two distinct multilocus genotypes or fewer, and sharing of clonal lineages across populations was uncommon, making up less than 3% of clonal lineages in this study. Consequently, migration events of a single individual into novel areas is likely accountable for a large portion of geographic spread. Additional gene flow into that area would occur with decreasing probability, as demonstrated by the Poisson-like distribution of multilocus genotypes across populations. This is not unexpected given the means of dispersal in this species. Kudzu seeds are dispersed by primarily by wind (Byrd, Maddox, and Westbrooks, 2009), but do not appear to play a large role in the invasion given the lack of evidence of sexual reproduction. Individuals from vegetative reproduction are found only as far as the last point of connection to

its parent root. Aboveground tissue can establish in the soil easily and could be distributed farther by animals or wind; however, it is unknown how often and how far such dispersal would be (Simberloff, 2010).

Introduction history

The six genetic clusters detected in the microsatellite data contain only a few hundred genetically distinct individuals introduced into the United States and they appear to share the same maternal lineage; therefore, it is more likely that one large or a few introductions occurred from the same region in Asia. If the amount of genetic diversity and population structure in the native range are similar to that of the invasive range, then sampling a single region could result in the patterns of genetic diversity observed in our study. The six genetic clusters in the invasive range would be artifacts of fine-scale population structure in the native range, not introduction events from disparate sources. This is supported by the fact that almost a third of the invasive range is made up of only three clonal lineages.

That being said, it is possible that chloroplast diversity may be simply uninformative to establish the number of introduction events. Specifically, if there was insufficient diversity in the chloroplast of this species, or if chloroplast diversity was well admixed across the native range, then those data would not provide any means in which to elucidate the introduction history. Consequently, comparisons of genetic diversity at both nuclear and chloroplast loci between both the native and invasive range will be necessary to identify the introduction history more definitely.

Patterns of range expansion

Lastly, it remains unclear from where and how kudzu geographically expanded its range from its original plantings in the Southeast. Locations of the first introduction events could not

be established, but the three most common clonal lineages spread widely across the range could be founding lineages distributed by the Soil Conservation Service. Furthermore, a route of expansion could not be implied given the population structure and lack of spatial patterning of genotypes. There was also no evidence of isolation by distance. If the invasion stemmed from a single large introduction, a genetic signature of expansion may be absent. Expansion could have simply diffused outwards. However, in the event of multiple introductions, random dissemination of individuals across the Southeast by the Soil Conservation Service could have simply erased our ability to detect any signature of geographic spread. Nonetheless, the lack of any particular clonal lineages dominating the expansion at the edge of the range does imply that no one clonal lineage or genetic cluster is responsible for invasion success. As the competitive traits exhibited by kudzu make it a significant invasive species in North America (Forseth and Innis, 2004; Lindgren et al., 2013), kudzu may be pre-adapted and thus pre-disposed to invasiveness.

To summarize, kudzu in the United States is highly genetically diverse, which is likely due to high clonal reproduction across its range. With a "proportion distinguishable" of only 20%, low genotypic diversity, high allelic diversity and highly negative inbreeding coefficient, kudzu exhibits a classic signature of high asexual reproduction. While populations grouped into multiple clusters, the introduction of this species into the United States appears to be limited with the detection of only a few hundred clonal lineages, the majority of which share the same chloroplast haplotype. During initial introduction, it appears that the Soil Conservation Service distributed propagules from each genetic cluster randomly across the range. Subsequent geographic expansion of the range seems to have occurred through rare immigration events into novel areas. As no particular clonal lineage dominated the invasion success of this species, it

appears that kudzu is either pre-adapted to its novel environment or exhibits extensive phenotypic plasticity to environmental heterogeneity.

Tables

Table 2.1. Global averages of genetic and clonal diversity measures of all 87 populations in the invasive range.

Diversity Measure	Value	Standard Error
N	20.0	0.572
MLG	4.06	0.381
R	0.165	0.0194
A_R	1.41	0.0191
N_E	1.70	0.0418
H_O	0.444	0.0124
uH_E	0.398	0.0123
F	-0.575	0.044
eff	2.20	0.187
div	0.361	0.0346
shc	0.406	0.0428
eve	0.658	0.0258

Note: A_R, allelic richness; div, Nei's genetic diversity index; eff, effective number of MLGs based on rarefaction; eve is evenness; F, fixation index; H_O, observed heterozygosity; N, sample size; N_E, average effective number of alleles; MLG, distinct multilocus genotype; R, "proportion distinguishable"; shc, Shannon-Weiner index corrected for sample size; uH_E, Nei's unbiased expected heterozygosity; SE, standard error for each measure.

Table 2.2. Genetic diversity across loci for kudzu in the invasive range.

Locus	**A**	**A_R**	**N_E**	**H_O**	**uH_E**	**F_{IS}**
KZ06	7	1.21	1.33	0.145	0.213	0.144
KZ12	2	1.39	1.56	0.510	0.390	-0.691
KZ31	5	1.29	1.41	0.336	0.291	-0.504
KZ35	12	1.59	2.14	0.626	0.588	-0.468
KZ37	5	1.57	1.90	0.700	0.568	-0.656
KZ42	2	1.00	1.00	0.004	0.004	-0.097
KZ46	9	1.43	1.80	0.459	0.427	-0.329
KZ50	11	1.37	1.67	0.354	0.368	-0.191
KZ56	13	1.41	1.82	0.439	0.406	-0.370
KZ57	15	1.42	1.93	0.395	0.423	-0.199
KZ68	5	1.36	1.67	0.385	0.357	-0.356
KZ73	4	1.53	1.92	0.638	0.532	-0.555
KZ75	7	1.57	1.99	0.641	0.572	-0.479
KZ78	17	1.62	2.04	0.741	0.621	-0.615
KZ79	4	1.21	1.31	0.285	0.206	-0.684
Mean	8	1.40	1.70	0.444	0.398	**-0.403**

Note: A is the total number of alleles at that locus. A_R is average allelic richness. N_E is the effective number of alleles. H_O is observed heterozygosity. uH_E is Nei's unbiased expected heterozygosity. F_{IS} is the inbreeding coefficient averaged across populations at each locus.

Figures

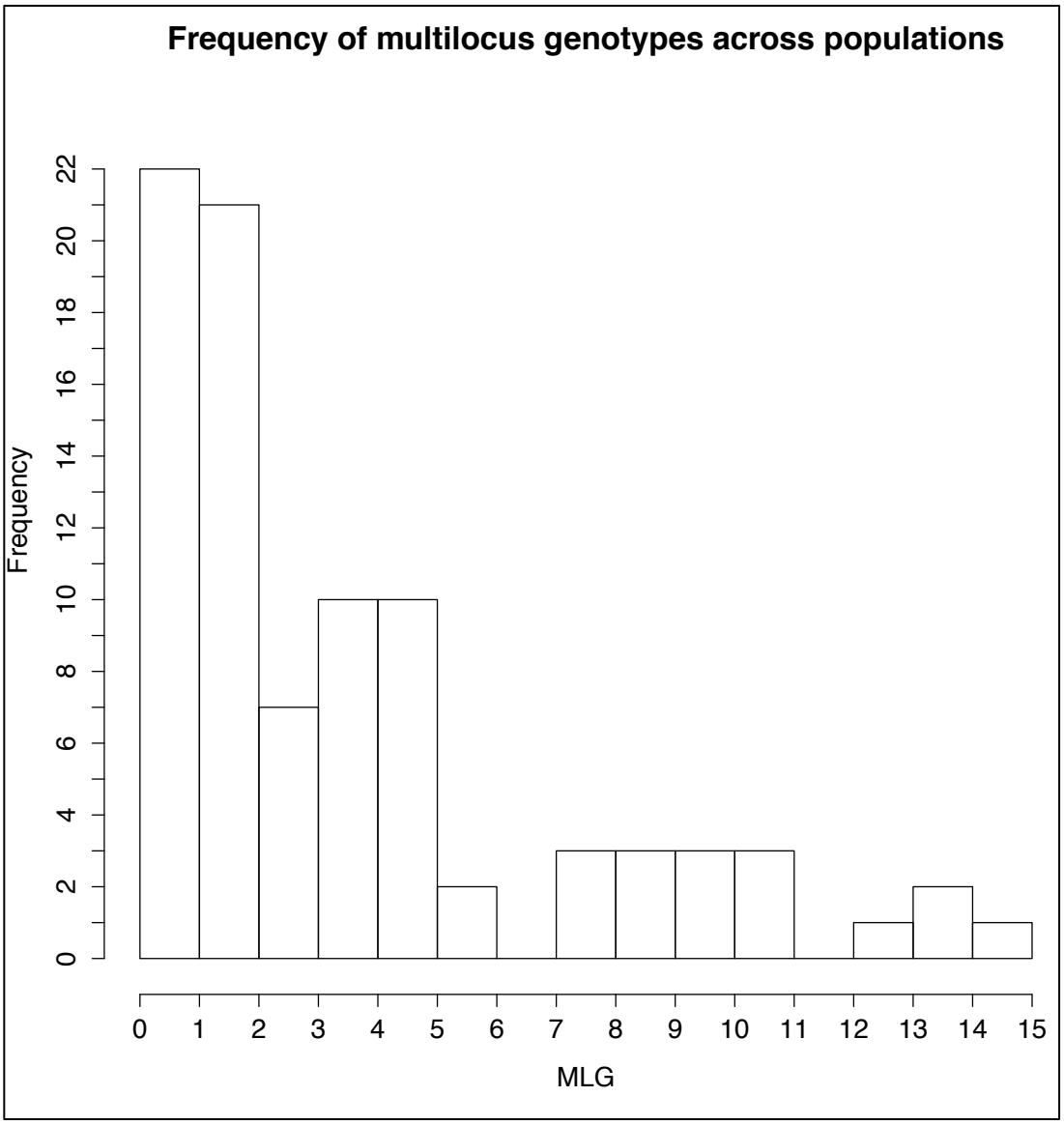


Figure 2.1. Histogram depicting the number of multilocus genotypes in a given population across the invasive range.

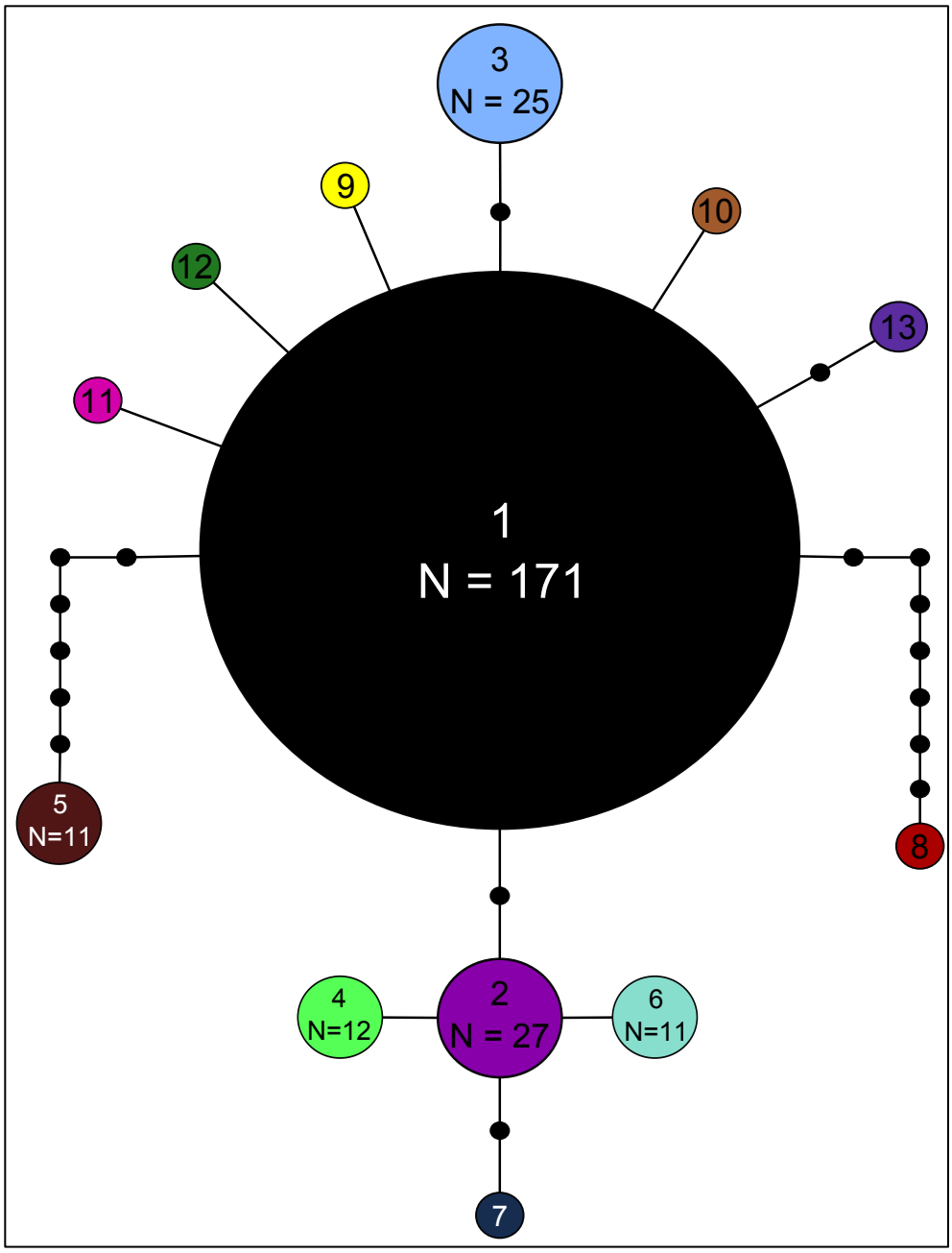


Figure 2.2. A haplotype network of the 13 haplotypes detected at the rpoB-trnC chloroplast marker. Smallest circles represent 1 to 4 individuals. Intermediate sizes represent 10 to 27 individuals. N indicates the sample sizes of haplotypes greater than 10 individuals.

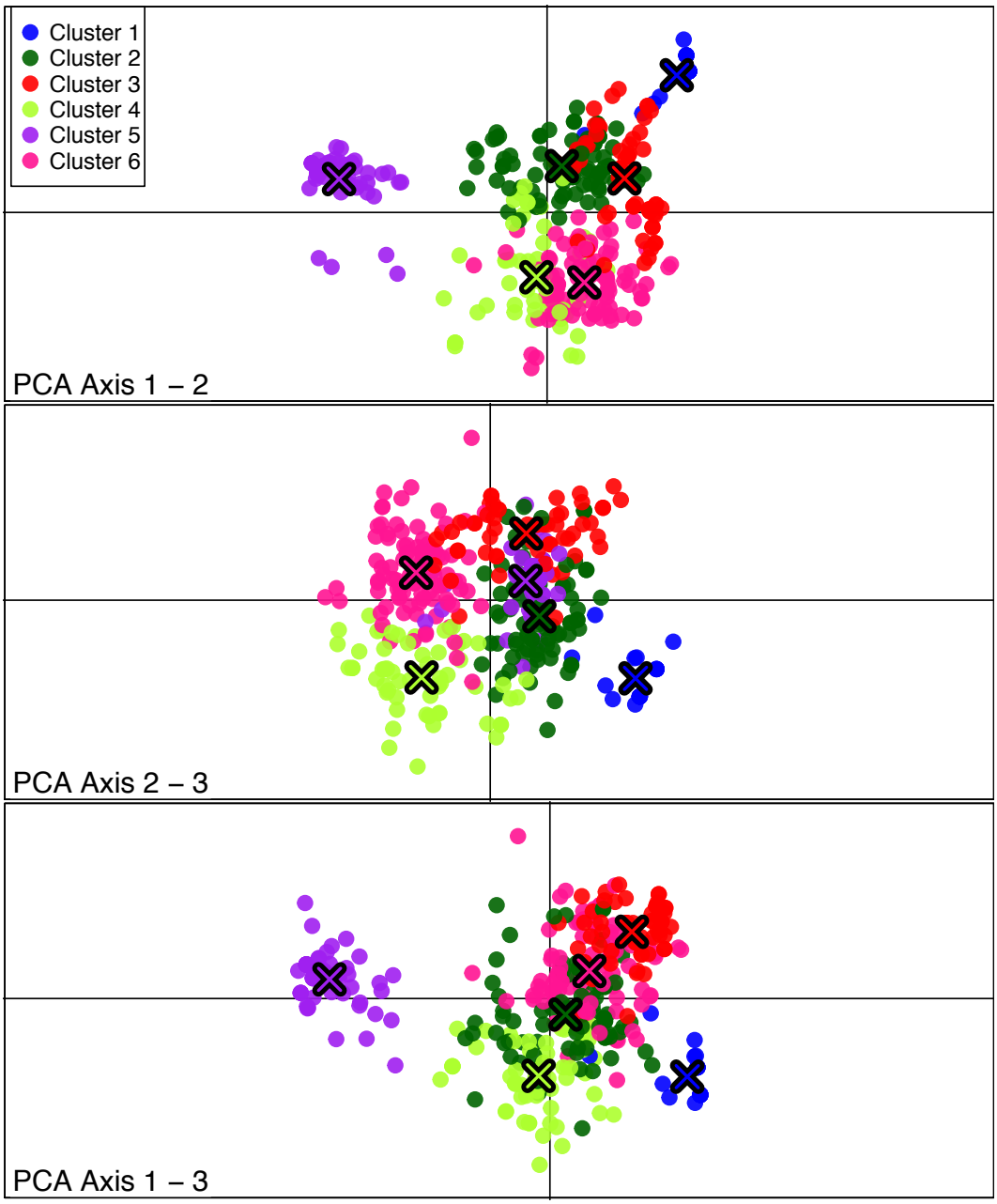


Figure 2.3. A DAPC plot the six genetic clusters detected in the microsatellite data. The color of the X matches the color of the corresponding cluster and delineates the centroid of that cluster.

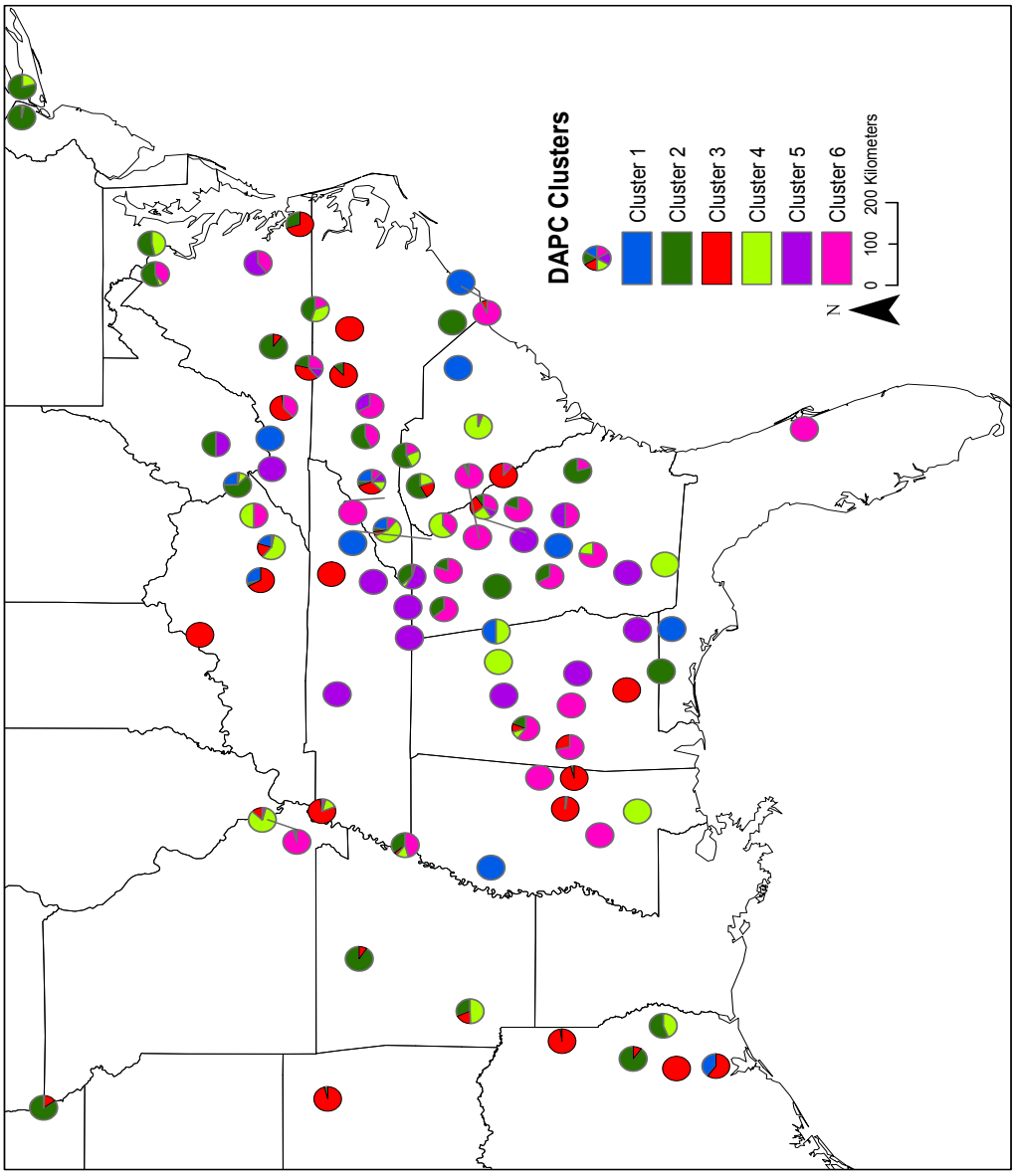


Figure 2.4. Structure map highlighting the geographic distribution of genetic clusters across the introduced range based on DAPC results for K = 6 clusters. Pie charts represent the proportion of a population assigned to a given cluster. Colors match those of the clusters in Figure 3.

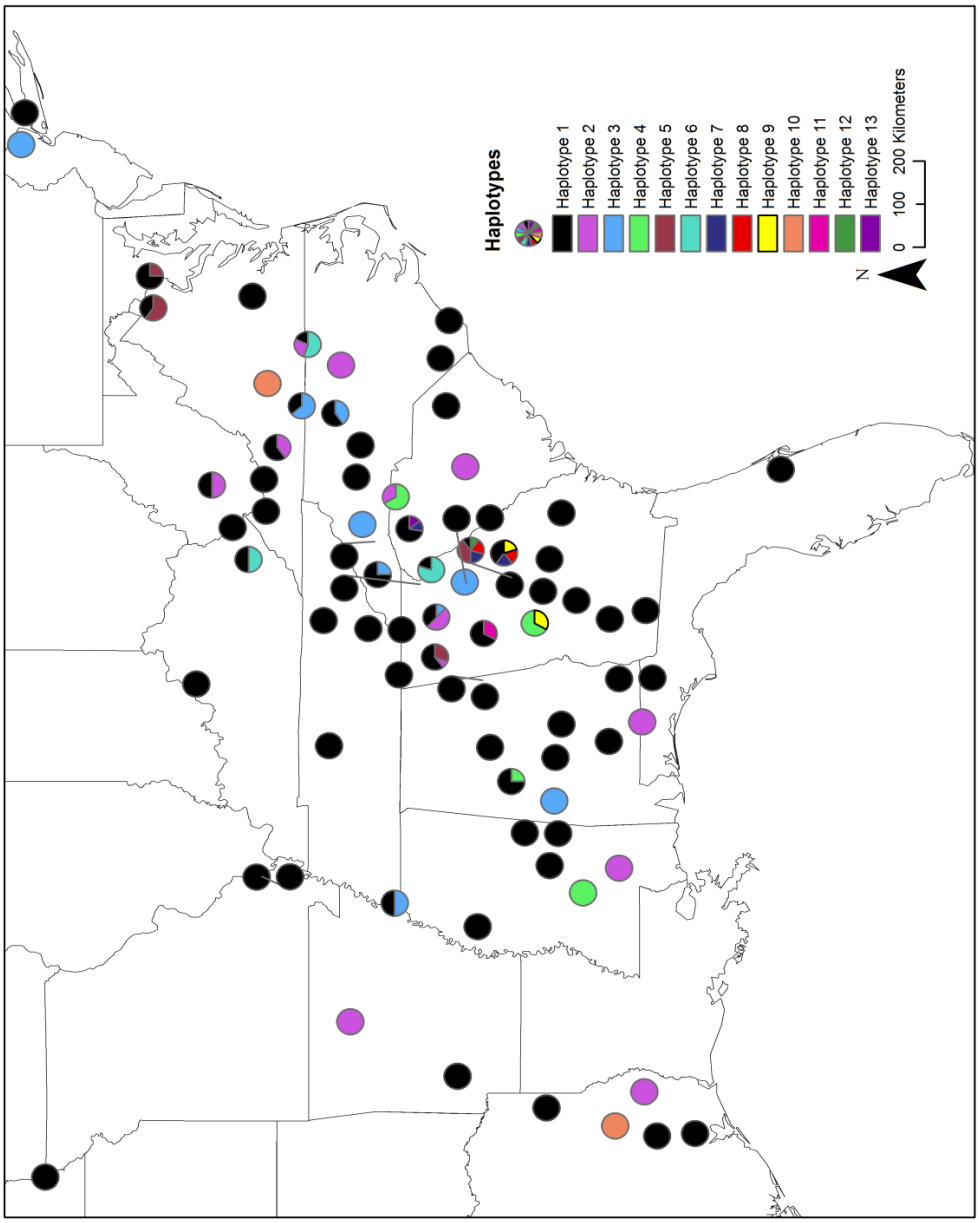


Figure 2.5. A haplotype map exhibiting the geographic distribution of 13 haplotypes dtected at the rpoB-trnC chloroplast marker with corresponding colors to Figure 2. Pie charts indicate the proportion by which a population contains to a given haplotype.

References

ABRAMOVITZ, J. N. (1983) Pueraria lobata Willd. (Ohwi), Kudzu: limitations to sexual reproduction. M.S. thesis, University of Maryland, College Park, Maryland, USA.

AGAPOW, P. Ä., AND A. BURT. (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1: 101-102.

ALDERMAN, D. H., AND D. G. S. ALDERMAN. (2001) Kudzu: A tale of two vines. *Southern Cultures* 7: 49-64.

BALLOUX, F., L. LEHMANN, AND T. DE MEEÛS. (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635-1644.

BARRETT, S. C. (2011) Why reproductive systems matter for the invasion biology of plants. In D.R. Richardson [ed.], Fifty Years of Invasion Ecology: The Legacy of Charles Elton, 195-210. Blackwell Publishing, Hoboken, New Jersey, USA.

BARRETT, S. C. H., R. I. COLAUTTI, AND C. G. ECKERT. (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* 17: 373-383.

BAZIN, É., H. MATHÉ-HUBERT, B. FACON, J. CARLIER, AND V. RAVIGNÉ. (2014) The effect of mating system on invasiveness: some genetic load may be advantageous when invading new environments. *Biological Invasions* 16: 875-886.

BLAUSTEIN, R. (2001) Kudzu's invasion into Southern United States life and culture. *The Great Reshuffling: Human dimensions of invasive species*: 55-62.

BYRD, J. D., V. MADDOX, AND R. WESTBROOKS. (2009) Kudzu [Pueraria montana (Lour.) Merr.]

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

DE MEEÛS, T., AND F. O. BALLOUX. (2004) Clonal reproduction and linkage disequilibrium in diploids: a simulation study. Infection, *Genetics and Evolution* 4: 345-351.

DOUHOVNIKOFF, V., AND R. DODD. (2003) Intra-clonal variation and a similarity threshold for identification of clones: application to Salix exigua using AFLP molecular markers. *Theoretical and applied genetics* 106: 1307-1315.

EDGAR, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32: 1792-1797.

ELLSTRAND, N. C., AND M. L. ROOSE. (1987) Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*: 123-131.

ELLSTRAND, N. C., AND K. A. SCHIERENBECK. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* 97: 7043.

ERSTS, P. J. (2013) Geographic Distance Matrix Generator (version 1.2.3).

ESTOUP, A., AND T. GUILLEMAUD. (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* 19: 4113-4130.

EVEREST, J. W., J. H. MILLER, D. M. BALL, AND M. PATTERSON. (1999) Kudzu in Alabama history, uses, and control. Alabama A&M and Auburn Universities, Alabama Cooperative Extension System ANR-65.

EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

FELSENSTEIN, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

FORSETH, I. N., AND A. F. INNIS. (2004) Kudzu (Pueraria montana): history, physiology, and ecology combine to make a major ecosystem threat. *Critical reviews in plant sciences* 23: 401-413.

GLENN, T. C., AND N. A. SCHABLE. (2005) Isolating microsatellite DNA loci. *Methods in enzymology* 395: 202-222.

GRIMSBY, J. L., D. TSIRELSON, M. A. GAMMON, AND R. KESSELI. (2007) Genetic diversity and clonal vs. sexual reproduction in Fallopia spp. (Polygonaceae). *American Journal of Botany* 94: 957-964.

HAUSWALDT, J. S., AND T. C. GLENN. (2003) Microsatellite DNA loci from the Diamondback terrapin (Malaclemys terrapin). *Molecular Ecology* Notes 3: 174-176.

HICKMAN, J. E., S. WU, L. J. MICKLEY, AND M. T. LERDAU. (2010) Kudzu (Pueraria montana) invasion doubles emissions of nitric oxide and increases ozone pollution. *Proceedings of the National Academy of Sciences* 107: 10115.

HOFFBERG, S. L., K. BENTLEY, J. B. LEE, K. E. MYHRE, K. IWAO, T. C. GLENN, AND R. MAURICIO. (2014) Characterization of 15 microsatellite loci in kudzu (Pueraria montana var. lobata) from the native and introduced ranges. *Conservation Genetics Resources*: 1-3.

HURLBERT, S. H. (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

JEWETT, D., C. JIANG, K. BRITTON, J. SUN, AND J. TANG. (2003) Characterizing Specimens of Kudzu and related taxa with RAPD's. *Castanea*: 254-260.

JOMBART, T. (2013) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1.4-0. Available at: http://cran.r-project.org/web/packages/adegenet/

JOMBART, T., S. DEVILLARD, AND F. BALLOUX. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.

KALINOWSKI, S. T. (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics* 5: 539-543.

KALINOWSKI, S. T. (2005) HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187-189.

KAMVAR, Z. N., J. F. TABIMA, AND N. J. GRÜNWALD. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.

KOLBE, J. J., A. LARSON, J. B. LOSOS, AND K. DE QUEIROZ. (2008) Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biology Letters* 4: 434.

LIBRADO, P., AND J. ROZAS. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

LINDGREN, C. J., K. L. CASTRO, H. A. COINER, R. E. NURSE, AND S. J. DARBYSHIRE. (2013) The Biology of Invasive Alien Plants in Canada. 12. Pueraria montana var. lobata (Willd.) Sanjappa & Predeep. *Canadian journal of plant science* 93: 71-95.

MEIRMANS, P. G., AND P. H. VAN TIENDEREN. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4: 792-794.

MITICH, L. W. (2000) Kudzu [Pueraria lobata (Willd.) Ohwi]. *Weed Technology* 14: 231-235.

PAPPERT, R. A. (1998) Population genetic variation and heterotic patterns of Pueraria lobata Ohwi (kudzu). M.S. thesis, University of Georgia, Athens, GA.

PAPPERT, R. A., J. HAMRICK, AND L. A. DONOVAN. (2000) Genetic variation in Pueraria lobata (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany* 87: 1240-1245.

PEAKALL, R., AND P. E. SMOUSE. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

PEAKALL, R., AND P. E. SMOUSE. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* 28: 2537-2539.

PETIT, R. J., A. EL MOUSADIK, AND O. PONS. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844-855.

POMPANON, F. O., A. L. BONIN, E. BELLEMAIN, AND P. TABERLET. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6: 847-846.

PRENTIS, P. J., J. R. U. WILSON, E. E. DORMONTT, D. M. RICHARDSON, AND A. J. LOWE. (2008) Adaptive evolution in invasive species. *Trends in Plant Science* 13: 288-294.

RAYMOND, M., AND F. ROUSSET. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248-249.

REZNICK, D. N., AND C. K. GHALAMBOR. (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112: 183-198.

ROUSSET, F. (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106.

SELKOE, K. A., AND R. J. TOONEN. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615-629.

SHAW, J., E. B. LICKEY, J. T. BECK, S. B. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142-166.

SIMBERLOFF, D. (2010) Kudzu. In D. Simberloff AND M. Rejmanek [eds.], Encyclopedia of biological invasions, 396-399. University of California Press, Berkeley, California, USA.

STEWART, M. A. (1997) Cultivating kudzu: the Soil Conservation Service and the kudzu distribution program. *The Georgia Historical Quarterly*: 151-167.

SUN, J., Z. LI, D. JEWETT, K. BRITTON, W. YE, AND X. GE. (2005) Genetic diversity of Pueraria lobata (kudzu) and closely related taxa as revealed by inter-simple sequence repeat analysis. *Weed Research* 45: 255-260.

TAMURA, K., G. STECHER, D. PETERSON, A. FILIPSKI, AND S. KUMAR. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* 30: 2725-2729.

TEMPLETON, A. R., K. A. CRANDALL, AND C. F. SING. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619-633.

VAN OOSTERHOUT, C., W. F. HUTCHINSON, D. P. WILLS, AND P. SHIPLEY. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535-538.

WARD, D. B. (1998) Pueraria montana: the correct scientific name of the kudzu. *Castanea* 63: 76-77.

WECHSLER, N. R. (1977) Growth and Physiological Characteristics of Kudzu: Pueraria Lobata (Willd.) Ohwi, in Relation to Its Competitive Success. M.S. thesis, University of Georgia, Athens, Georgia, USA.

YAN, Y., Y. HUANG, X. FANG, L. LU, R. ZHOU, X. GE, AND S. SHI. (2011) Development and characterization of EST-SSR markers in the invasive weed Mikania micrantha (Asteraceae). *American Journal of Botany* 98: e1-e3.

CHAPTER 3

INVASIVE KUDZU, *PUERARIA MONTANA VAR. LOBATA*, EXHIBITED A SHIFT TO HIGHER RATES OF CLONALITY AFTER MULTIPLE INTRODUCTIONS FROM EAST ASIA INTO THE UNITED STATES

Introduction

Across the globe, biological invasions are a serious threat to forestry, agriculture and environmental health (Pimentel et al. 2005). Biological invasions destroy habitat (Molnar et al. 2008), disrupt the balance of abiotic functions (Gordon 1998), and displace flora and fauna in rapid time frames (Mooney and Cleland 2001, Hejda et al. 2009). Consequently, the ability of invasive species to populate massive areas so successfully has been an area of intense research for decades (Baker and Stebbins 1965), but few invasions are thoroughly understood (Prentis et al. 2008). More comparative studies of genetic patterns among the native and invasive ranges are needed to identify the ecological, demographic or evolutionary factors that produce successful invasions.

Pueraria montana (Lour.) Merr. var. *lobata* (Willd.) Maesen & Almeida (Ward 1998), kudzu, is an iconic invasive species since its first introduction into the United States almost 150 years ago (Lowe et al. 2000) and is an opportunity to understand the genetic and evolutionary factors that lead to an aggressive invader. In addition to rapid geographic spread on the order of 50,000 hectares per year (Forseth and Innis 2004), invasive kudzu forms large, dense canopies that homogenize large areas of habitat. The density of its extant range threatens ozone quality above highly infested regions through cumulative emissions of nitrous oxide (Hickman et al.

2010). Yet despite its significant impact in eastern North America, little is known about the evolutionary history of this invasion.

One such factor is the amount of genetic variation introduced into the invasive range which influences the evolutionary potential for adaptation in the new environment (Prentis et al. 2008). It was historically assumed that species invasions evolved after large genetic bottlenecks whereby reduced standing genetic variation in the novel environment could constrict adaptive evolution (Allendorf and Lundquist 2003). However, recent studies have shown increasing evidence that genetic bottlenecks do not always occur (Dlugosch and Parker 2008). Genetic diversity may not be reduced if individuals in source populations are themselves highly diverse (Maderspacher 2011). Additionally, similar or even increased amounts of genetic variation in the invasive range can be produced from multiple introductions (Kolbe et al. 2004). Multiple introductions provide the opportunity for new, heterotic multilocus genotypes from population admixture (Ellstrand and Schierenbeck 2000) as well as increased frequencies of beneficial standing genetic variation in the new environment (Prentis et al. 2008). Consequently, the comparison of patterns of genetic diversity between the native and invasive range can identify differences in the amount of genetic diversity and how these differences may have been produced.

Another factor that is critical to invasion is the means of reproduction in an organism. Invasive species that exhibit multiple modes of reproduction can illuminate how reproductive strategies affect genetic variation in an invasion (Barrett 2011). Kudzu is capable of both sexual and vegetative reproduction (Lindgren et al. 2013), each of which has different genetic effects and impacts on invasion. Sexual reproduction generates novel multilocus genotypes through recombination, offering a means for rapid adaptation to a new environment (Barrett et al. 2008).

Conversely, a lack of recombination from clonal reproduction essentially preserves multilocus genotypes, facilitating the spread of fit lineages (Barrett et al. 2008). In its invasive range, kudzu has been found to be highly clonal (Bentley, Chapter 2; Pappert et al. 2000); however, it is unclear to what degree kudzu exhibits vegetative versus sexual reproduction and if there has been a shift to higher clonal reproduction upon introduction to the United States. It is not uncommon to observe reproductive shifts after colonization in plants, particularly in favor of selfing or asexuality, which provides reproductive assurance (e.g. *Elodea canadensis*, Sculthorpe 1967; *Fallopia japonica*, Hollingsworth & Bailey 2000; *Oxalis pes-caprae*, Ornduff 1987; *Eichhornia paniculata*, Barrett et al., 2008; *Butomus umbrellatus*, Lui et al. 2005).

The influence of these genetic or biological factors can only be accurately determined by first identifying the native sources of invasion and number of introduction events. It has been widely reported that kudzu was first introduced into the United States from Japan at the Centennial Exposition in Philadelphia, Pennsylvania, in 1876 (Li et al. 2011), but this has never been evaluated. Previous genetic studies have provided no indication as to possible sources and disagree on how many introductions were involved. Pappert *et al.* (2000) and Sun *et al.* (2005) concluded that high genetic diversity in the invasive range likely stemmed from multiple introductions, while the within-range study in Chapter 2 found evidence of a possible single large introduction due to relatively few clonal lineages and haplotypes found in the invasive range. However, none of these studies directly compared populations in the United States to the native range to identify putative sources or quantify the number of introductions. Such comparisons are necessary to understand the introduction history of an invasive species accurately (Kolbe et al. 2008).

Fortunately, the biology of kudzu facilitates detection of putative sources. Due to kudzu's longevity and infrequent recombination due to high rates of vegetative reproduction (Mitich 2000), the lifespan of a given genotype is longer relative to short-lived, sexually reproducing organisms (Arnaud-Haond et al. 2012). Taken together, evolutionary change of clonal lineages in 150 years should be relatively less than species of other life histories, making forensic phylogeographic analyses potentially more accurate.

Thus, to understand the invasion history of this species, we address the following questions. 1) How many introductions were there and where did the founders originate? 2) Are there differences in genetic diversity between the native and invasive ranges? 3) Was there a shift towards increased asexuality in the invasive range? To do so, we included the entire geographical breadth of both ranges to present a better understanding of this invasion.

Materials and Methods

Study species

Pueraria montana var. lobata, or kudzu, is a long-lived leguminous vine native to eastern Asia that was widely introduced into the United States in the early 1900s (Mitich 2000). Since then, it has expanded to 5 million hectares of land, growing its range by 50,000 hectares per year (Forseth and Innis 2004). Kudzu is capable of both sexual and vegetative reproduction; however, seed germination and seedling survival is very low in both the native and introduced range (Abramovitz 1983, Tsugawa et al. 1988, Lindgren et al. 2013). Once established, however, it exhibits several traits conducive to long-term survival, including extensive root structures capable of storing large resources (Wechsler 1977), rapid growth (Sasek and Strain 1988), little structural allocation needed for support (Wechsler 1977), and high vegetative reproductive capability of hundreds and possibly thousands of ramets per individual (Forseth and Innis 2004).

Overall, these competitive traits facilitate the production of large patches that blanket affected areas, eliminating light and excluding native species. They also make control efforts extremely difficult often lasting over a decade to eradicate a single population (Everest et al. 1999). Kudzu is considered one of the worst invasive species in the world (Lowe et al. 2000), causing up to $500 million in damages every year in forestry and agriculture (Blaustein 2001, Forseth and Innis 2004).

Sampling

Almost 5000 samples spanning 229 populations across the native and invasive ranges were included in this study. Specifically, we collected approximately 22 young leaves per population from 87 populations across the invasive range in the United States and 142 populations across the native range, including 91 from China, including Taiwan, 29 from Japan and 22 from South Korea. Collection sites were generally separated by at least 100 km. We sampled along a transect that ran the length of the population. Distance between samples was proportional to the transect length, separated by several meters. A single GPS coordinate was collected for each population (Table 3.2). Leaves were dried for genomic extraction.

Genotyping and sequencing

Each sample was extracted using either a Qiagen DNeasy Kit (Qiagen, Germantown, MD) or Axyprep Multisource Genomic DNA Miniprep Kits with the centrifuge protocol for plant tissue (Axigen Biosciences, Union City, CA). Each sample was screened against 15 microsatellite markers (Hoffberg et al. 2014). PCR amplification was based on a 3-primer PCR approach using CAG tagging (Hauswaldt and Glenn 2003), but modified for kudzu by Hoffberg et al. (2014). All primers followed PCR protocols from Hoffberg et al. (2014). Amplification was verified with 1.5% agarose gels. Amplicons were diluted 1:6 in deionized water and visualized

with an Applied Biosystems-3730xl DNA Analyzer (Applied Biosystems, Foster City, CA) with a 500 ROX-labeled size standard in each well to ensure accurate allele calls. Genotyping results were scored using GENEMARKER version 2.4 (SoftGenetics, State College, PA). Prior to analysis, all genotypes were checked for errors and tested for null alleles in Micro-Checker (Van Oosterhout et al. 2004). Less than 10% of all tests suggested possible evidence of null alleles, so no correction was taken. Possible null alleles existed in 17%, 17%, 20% and 23% of populations for KZ06, KZ57, KZ46 and KZ50, respectively.

We calculated the genotyping error rate (Pompanon et al. 2005), by randomly choosing 96 individuals without any missing data for PCR amplification of all markers in two blind replicates. Mean error rate per locus and the multilocus genotype error rate were calculated based on allelic mismatches between the first and second allele scores identified through custom Perl scripts. Missing data were excluded from the error rate calculations. The average error rate per locus was 5.90×10^{-3}, while the multilocus genotype error rate was 7.81×10^{-2}. While error was low compared to other studies, 80% of error was due to mistaken allele calling in KZ50 and KZ56; therefore, all alleles were rescored for these markers with consistent scoring rules to minimize this issue.

Additional genetic diversity was assessed with sequences from a single 830 bp intergenic spacer chloroplast region, rpoB-trnC. Primers and the PCR protocol followed (Shaw et al.) (2005) for the universal primer pair, $trnC^{GCA}R$ (CAC CCR GAT TYG AAC TGG GG) and rpoB (CKA CAA AAY CCY TCR AAT TG). The PCR recipe includes a 25 μL reaction [1 mM Tris-HCl pH 8.3, 5 mM KCl, 100 μg/mL BSA, 1.8 mM $MgCl_2$, 0.2 μM of each dNTP, 0.2 μM of each primer, 0.625 units of Taq DNA Polymerase (New England Biolabs) and ~20ng DNA template]. The PCR protocol followed a touchdown procedure: 80°C 5 min, followed by

30 cycles of 96°C 1 min, 50°C 2 min, 72°C 3 min, followed by 72°C 5 min. PCR product of sequences was prepared according to Glenn and Schable (2005) and sequenced using an Applied Biosystems 3730xl 96-capillary DNA Analyzer. Nucleotide calling and quality of raw chloroplast sequence data were verified in Sequencher 4.7 (Gene Codes Corp, Ann Arbor, MI) and aligned using default parameters with MUSCLE (Edgar 2004) in Geneious version 7 (Biomatters Ltd.). Any detected polymorphisms were confirmed both in the chromatograms and the alignment to eliminate scoring error.

Assigning distinct multilocus genotypes and clones

Because *Pueraria montana* var. *lobata* exhibits significant clonal reproduction, it is necessary to carefully define a multilocus genotype as simply counting unique multilocus genotypes will greatly overestimate their number. Multilocus genotypes and their associated clones were identified in the software, GENODIVE (Meirmans and Van Tienderen 2004). Based on the histogram of the distribution of allelic pairwise distances under a stepwise mutation model where missing data were ignored, a clonal threshold of three maximum allowable mismatches was determined as most appropriate. This clonal threshold defined any two samples as the same clonal lineage if they exhibited three or fewer mutational differences between them. This method is a standard means to account for scoring error and somatic mutation for clonal species in identifying distinct multilocus genotypes in a population (Arnaud-Haond et al. 2007).

Identifying putative sources and number of introduction events

Genetic clustering of native and invasive populations was achieved using Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010). The appropriate number of clusters (K) was determined by regressing the Bayesian Information Criteria against cluster number to identify cluster numbers in the elbow region of the curve (Jombart 2013). All

eigenvalues and principal components, based on the alpha score method, were retained. Scatterplots for each pair of axes were generated. Spatial patterns of genetic clustering was visualized with pie charts representing percent membership of that population to a given cluster overlaid on GPS coordinates in ArcMap, hereafter called a structure map. Population admixture through sexual reproduction was assessed using a structure plot by charting the proportion membership of each individual to a given cluster.

The distribution of shared clonal lineages was utilized to retrace clonal lineages directly back to the native range. Clones were re-defined across populations rather than within populations in GENODIVE (Meirmans and Van Tienderen 2004), maintaining the same clonal threshold. Clonal lineages shared across populations were then observed to identify putative sources from the native range.

Evidence of introduction events and putative source populations was also assessed with chloroplast sequencing data. Haplotype networks are often more accurate than traditional gene trees when divergence is low as in the case of recent colonization (Clement et al. 2000). Consequently, a minimum-spanning haplotype network of chloroplast sequences was created using statistical parsimony with a 95% connection limit and gaps treated as a 5th state in TCS v. 1.21 (Templeton et al. 1992, Clement et al. 2000). Pie charts representing percent membership of a population to a given haplotype were overlaid on GPS coordinates in ArcMap to visualize the distribution of haplotypes across both ranges, hereafter referred to as a haplotype map. This map complements the haplotype network, visualizing any geographic patterns to the distribution of haplotypes.

To summarize the evidence linking native populations to the US, each native population was assigned a score based on the degree to which that population grouped with any US

populations. Scores were based on the results of the haplotype network, the DAPC analysis for both K =12 and K =18, and the shared multilocus genotype analysis. The scoring obeyed the following rules for each analysis: a native population was given a single point if any individuals from that population grouped with individuals from the US in groups that contained four or more other native populations ("indirect link"), two points if any individuals from that population grouped with individuals from the US in groups that contained three or fewer other native populations ("semi-direct link"), three points if any individuals from that population grouped with individuals from the US as the only native source in that group ("direct link"). Scoring results of each analysis were then summed to produce a total score. The scores were overlaid on a map using GPS coordinates as a heat map. A hotter color, nearing red, reflected relatively more evidence linking that native population to the US. A cooler color, nearing green, reflected little evidence of that population contributing to the US invasion. Populations were considered putative sources if they had a score of 5 or 6. To observe whether all populations deemed to be "putative sources" accounted for all of the genetic identify of invasive populations, a DAPC-based structure plot was generated with both invasive populations and putative source population with scores of 5 and 6.

After identification of putative source populations, differences in genetic and clonal diversity measures were compared between putative and non-putative sources to determine how the putative founders may have affected the invasive range as compared to other areas of the range. Differences between the invasive range and its putative sources were also compared for both genetic and clonal diversity measures to observe how the invasion process may have led to genetic changes since introduction. Due to unbalanced sample sizes, the invasive range and non-putative sources in each comparison were sub-sampled 1000 times to match the number of

number of putative sources populations prior to one-tailed Wilcoxon rank-sum tests for each diversity measure in JMP Pro 10 (SAS Institute, Cary, North Carolina).

Comparing genetic diversity among ranges

Genetic diversity statistics were measured for each population in each range and compared among ranges. These diversity measures included the percentage of polymorphic loci (P%), effective number of alleles (N_E), observed heterozygosity (H_O), and Nei's unbiased expected heterozygosity (uH_E) using GenAlEx version 6.501 (Peakall and Smouse 2006, 2012). As allelic richness is the number of alleles expected given the number of genes sampled in a population (Hurlbert 1971, Petit et al. 1998, Kalinowski 2004), allelic richness (N_A) was calculated with a rarefaction method in HP-RARE (Kalinowski 2005) in order to standardize for uneven sample sizes. Rarefaction was standardized to the smallest population size (2 genes) with an even number of alleles per marker. Deviations from Hardy-Weinberg equilibrium were tested in GENEPOP (Raymond and Rousset 1995, Rousset 2008), using 500 batches and Bonferonni correction for the 15 US and 42 Asian populations with 7 multilocus genotypes or greater to maintain sufficient sample size. Clones were excluded as the inclusion of clones may bias these results inappropriately. A single representative genotype with the least amount of missing data was retained for each clonal lineage in a population.

Significant differences in standard measures of genetic diversity between the invasive and native ranges were tested in JMP Pro 10 (SAS Institute Inc., Cary, North Carolina) using a Wilcoxon rank-sum test for all diversity statistics, except for observed and expected heterozygosities which were each tested with a two-sample t-test. Differences in the percent polymorphic loci were compared using a one-way ANOVA after square root transformation.

Lastly, significant population differentiation was tested within and among ranges (F_{RT}) with an AMOVA using 999 permutations in GenAlEx version 6.501 (Peakall and Smouse 2006, 2012).

Excess heterozygosity common after genetic bottlenecking was tested using a one-tailed Wilcoxon signed rank test under the stepwise mutation model (SMM) in BOTTLENECK (Piry et al. 1999). A chi-square test was used to determine whether there was a significant association between origin (native or invasive) and excess heterozygosity. The strength of association was measured using Pearson's phi coefficient in JMP Pro 10 (SAS Institute Inc., Cary, North Carolina). Both distinct multilocus genotypes and their associated clones were included in this analysis as this test depends on the allele frequency spectrum of a population.

To assess any change in chloroplast diversity, the number of haplotypes, the number of segregating sites, the number of parsimony informative sites and nucleotide diversity per site were calculated for each range using DNaSP (Librado and Rozas 2009).

Assessing clonal diversity within and among ranges

Clonal diversity was measured based on the clonal threshold of three mutational differences set in GENODIVE described above. The number of unique multilocus genotypes, Nei's genetic diversity index corrected for sample size (div), also known as Simpson's diversity index, and evenness (eve) were estimated in GENODIVE (Meirmans and Van Tienderen 2004). The effective number of multilocus genotypes based on rarefaction (eff) was also calculated to account for uneven and small sample sizes (Grünwald et al. 2003). This measure can be informative in identifying variation in genotypic frequency distribution within a population much equivalent to the effective number of alleles (Meirmans and Van Tienderen 2004). The "proportion distinguishable" (R) was calculated as ((MLG-1)/(N-1)) (Ellstrand and Roose 1987) to assess the frequency of sexual reproduction in a population. We used the complement of that

measure, 1-R, to infer rates of asexual reproduction. The fixation index (F) was also calculated in GenAlEx version 6.501 (Peakall and Smouse 2006, 2012). Clonal diversity statistics were tested for significant differences between the invasive and native ranges with Wilcoxon rank-sum tests except for observed and expected heterozygosities, which were two sample t-tests in JMP (SAS Institute Inc., Cary, North Carolina).

We also tested the null hypothesis that the observed diversity in each range was due to sexual reproduction to see if the native range also exhibits deviations from panmixia as was described for populations in the invasive range in Chapter 2. This was achieved by randomizing alleles 1000 times over individuals using the corrected Nei's diversity index for the two datasets in GENODIVE (Meirmans and Van Tienderen 2004). Secondly, a multilocus linkage disequilibrium measure, the index of association (I_A), was tested with 1000 permutations in poppr (Kamvar et al. 2014). This index uses the ratio of observed to expected variance in the number of loci in which two individuals differ to ascertain to what extent individuals demonstrate recombination indicative of random mating (Agapow and Burt 2001). The modified index of association, rbarD, was also measured to account for unequal sample size (Agapow and Burt 2001).

Genetic and clonal comparisons among ranges and pairwise comparisons between the putative sources, non-putative sources and invasive range were made with a one-tailed Wilcoxon rank-sum test in JMP Pro 10 (SAS Institute Inc., Cary, North Carolina). The test between the putative and non-putative sources as well as putative and invasive populations were re-sampled 1000 times in order to account for sample sizes.

Results

Putative sources and number of introduction events

The best number of genetic clusters in the microsatellite dataset was determined to be within the range of 12 to 18 clusters (Figure 3.16). Within that range, K of 12 was chosen as most appropriate as additional K did not greatly improve the proportion of correct assignment. Twenty-one principal components were the optimal number based on the alpha score method, which encompassed 62.8% of the cumulative variance. 94% of individuals were correctly assigned for K = 12 with the individual clusters exhibiting a percentage of correct assignment above 82%. 93% of individuals were correctly assigned for K = 18 with individual clusters exhibiting a percentage of correct assignment above 89%. The overall DAPC plot and DAPC structure plots for K = 18 can be found in material (Figures 3.17, 3.18, 3.19, 3.20, 3.21), but is not discussed in detail here. While K = 18 provided slightly finer-scale clustering, figures based on K = 18 did not reveal any patterns not observed in K = 12.

Of the 12 genetics clusters detected in DAPC (Figure 3.1), two clusters largely did not include the US, located in the southern half of China (Figure 3.2). Cluster 4 was present in southern China with only one individual clustering in the US, NC3-19. Cluster 10 was spread across central-southern China (Figure 3.2), but was completely absent in the US.

There were no clusters found in the invasive range (Figure 3.3) that did not include native individuals (Figure 3.2). Native genetic clusters found in the US clustered with populations across central-northern China, South Korea and Japan. Two clusters were spread across the US but clustered with few native populations. Specifically, Cluster 2 was spread across the entire invasive range, but was linked to only two populations in central China, specifically in Hubei and Jiangsu provinces, KHU2 and KJI5 respectively, and one population in KKO10 (Kangwon

province, South Korea). Cluster 11 was spread widely across the US, but was linked solely to KJP27 in Iwate in northern Japan. The microsatellite-based population tree using Nei's genetic distance did not show any US multilocus genotypes being more closely related to certain Asian genotypes over others as the entire tree was a large polytomy (Figure 3.22). The structure plots exhibited less admixture among clusters in the invasive range (Figure 3.4) with more admixture in the native range (Figures 3.5 (China), Figure 3.6 (Japan), Figure 3.7 (South Korea)).

When clonal lineages were re-defined across populations, it was also possible to detect shared clonal lineages between the US and Asia (Table 3.5). Four of the 18 shared clonal lineages between populations were shared between the US and Asia. One clone was shared across 22% of all US samples (379 samples) and 5 Chinese populations in central-northern China (KAN9, KB1, KHU2, KSA2 and KSD1). A second clonal lineage linked two US populations (TX4 and AL12) with Huangshan, Anhui province, China (KAN1). The third clonal lineage linked KY11 and NC17 to Aomori, Japan (KJP28). Lastly, a clonal lineage in NC1 was found in both Kagoshima, Japan (KJP6) and South Gyeongsang Province, Korea (KKO18).

Furthermore, US chloroplast haplotypes shared the same haplotypes as Asian samples (Table 3.3). While 16 haplotypes were found only in Asia, 10 haplotypes were assigned to both US and Asian populations. Haplotypes 1 and 2 were the most common (Figure 3.9), both spread across the entire US (Figure 3.10). In Asia, Haplotype 1 spanned the eastern coast of China up through Korea and Japan (Figure 3.11). Over 69% of all US samples and 33% of all Asian samples were assigned to Haplotype 1 (Table 3.3). Haplotype 2 was widespread across the entire Asian range albeit less common in Korea and northern China (Figure 3.11). Haplotype 2 was comprised of 10% of all US samples and 43% of all Asian samples.

Of all other haplotypes shared between the US and Asia, four haplotypes did not involve Japan, with two haplotypes linked to only China, one linked to only Korea and one linked to both China and Korea (Table 3.3). The two haplotypes that traced back to only Chinese populations were Haplotype 9 from the Shaanxi region of central China and Haplotype 12 from Suzhou, Jiangsu province on the eastern coast. The single haplotype in the US that was traced back to Korea was Haplotype 6 from Jido, South Korea on the southeastern coast. Haplotype 5 was linked to both China and Korea, specifically linked to Beijing, Hubei and Hunan provinces in China and four populations spread across Korea. Haplotype 4 was found in both Hubei, China and southern Japan. Additionally, two haplotypes were traced back to Japan alone, specifically Haplotype 3 from Miyagi in northeast Japan and Haplotype 7 from Ōita in southern Japan. Haplotype 11 in the US was linked back to Cheorwon in northern South Korea as well in addition to Hyogo in central Japan. Only 1 haplotype was found in the US with no native sources, Haplotype 27 in population SC6. This rare US-specific haplotype was different from Haplotype 26 from KZH8 in Zhejiang, China by only a single nucleotide polymorphism (Figure 3.9). Aside from haplotypes 1 and 2, there was little regional patterning of haplotypes in the US (Figure 3.10) or the native range (Figure 3.11) but there were more haplotypes in the northern part of the range.

When all analyses were combined to produce a heat map showing the number of links of native populations to the US, 11 native populations demonstrated the most evidence as putative sources with scores of 5 and 6. These populations were located in central-northern China, Japan and South Korea (Figure 3.12). Native populations with moderate ties to the US in yellow delineated by a score of four were largely located in central-northern China (7 populations) with

the addition of two populations in northern South Korea and one in central Japan (Figure 3.12). The rest of the native range exhibited few ties to the United States.

Of the putative source populations with scores of 5 and 6, four populations were located in central China, in Anhui (KAN1), Hubei (KHU2), Jiangsu (KJI3) and Shaanxi (KSA2) provinces. Two populations, KJP6 and KJP8, were located in the island of Kyushu in southern Japan. Three populations were in North Tohoku of northern Japan, specifically Iwate (KJP25, KJP27) and Aomori (KJP28) prefectures. Lastly, two populations were in in South Gyeongsang (KKO18) and South Jeolla (KKO15) provinces in southern South Korea. It is important to note that of the populations on this short list, almost all had ties to Haplotype 1 that made up over two-thirds of invasive haplotypes and Haplotype 2 that consisted of 10% of invasive haplotypes. Additionally, KHU2 and KSA2 were grouped with the clonal lineage shared across 22% of the invasive range. Putative sources with scores of 5 and 6 contained 9 of the 10 haplotypes found in the United States. Haplotype 11 was missing. KJP15 and KKO11 with scores of 4 contained this haplotype, but the structure plots with native populations of scores 4, 5 and 6 did not improve resolution of genetic clustering with the invasive range than populations with scores of 5 and 6 (results not shown).

A DAPC structure plot including all invasive populations and the 11 putative sources with an optimal 10 principal components retained illustrated that all invasive individuals clustered with putative sources with a best K of 11 (Figures 3.13 and 3.14). 90% of individuals were assigned correctly with individual cluster assignments all above 75%. While KJP27-24 was the only native individual to cluster with invasive populations in Cluster 10, several DAPC runs demonstrated that this individual consistently clustered with the same individuals (results not shown).

Assessment of genetic diversity

As 63.35% of loci were polymorphic, it was possible to compare measures of genetic diversity among ranges to test for differences in genetic diversity (Table 3.1). Based on the clonal threshold of 3, there were 274 clonal lineages in the United States and 685 clonal lineages in the native range. There was no significant difference in the percent polymorphic loci between ranges ($F = 1.29$; $df = 228$; $P = 0.2567$). 73 of the 138 Asian populations (53%) and 53 of 79 US populations (67%) demonstrated significant excess of heterozygosity; however, excess was higher in the invasive range ($X^2 = 4.15$, $df = 1$, $P = 0.0416$) albeit the association between origin (invasive and native) and excess heterozygosity levels was weak (phi = 0.138). Neither of these individual heterozygosity measures, H_O and uH_E, was significantly different between the native and invasive ranges ($P = 0.0823$ and $P = 0.8193$, respectively) (Table 3.1).

The United States also demonstrated a significant 35% reduction in the number of multilocus genotypes per population than Asia ($P = 0.0047$) with 3.07 and 4.70 multilocus genotypes in the US and Asia, respectively (Table 3.1). All other comparisons of genetic diversity measures among ranges demonstrated no statistical differences ($P > 0.05$) (Table 3.1). Allelic richness did not vary with 1.44 alleles per population on average in both ranges. Effective number of alleles was higher in Asia ($N_E = 1.82$) than in the US ($N_E = 1.71$), but again this was not significantly different. Deviations from Hardy-Weinberg equilibrium were detected in 33 tests (4%) of 775 tests with only two in the United States. Lastly, the analysis of molecular variance using F_{ST} revealed that 20% and 30% of genetic variation was distributed among populations in the invasive range and native range, respectively. Differentiation among ranges was very low ($F_{RT} = 0.02$).

Genetic diversity analysis for the 830 bp rpoB-trnC chloroplast region identified 27 haplotypes in the successfully sequenced 2316 samples (Table 3.4), including 96% of Asian and 100% of US populations. 16 haplotypes were found in Asia, while 11 haplotypes were found in the US (Table 3.3). There were 48 segregating sites in Asia but only 13 in the US, all of which were parsimony informative. The nucleotide diversity per site in Asia was 2.05×10^{-3} and 1.34×10^{-3} in the US. All haplotypes were highly related; all haplotypes were within two mutational steps from Haplotypes 1 and 2 except Haplotypes 5 and 12 (Figure 3.9). Haplotype 5 was likely a single 7 bp deletion event and the multiple mutations within the 10 base pair region also likely occurred from a single mutation event (Table 3.4).

All genetic diversity measures were significantly more diverse in the putative source populations than the rest of the native range with the exception of expected heterozygosity, which was not significantly different ($P = 0.587$) (Table 3.1). The putative sources exhibited a significantly higher number and effective number of multilocus genotypes with 9.91 multilocus genotypes and 6.90 effective multilocus genotypes per population than the non-putative sources with 4.27 multilocus genotypes and 2.72 effective multilocus genotypes per population. That represents an 80% and 87% difference among ranges, respectively. There were significantly more alleles ($P = 2.35 \times 10^{-4}$), a higher effective number of alleles ($P = 1.53 \times 10^{-3}$) and higher observed heterozygosity (8.25×10^{-3}) in putative sources than the non-putative sources.

When the putative sources were compared to the invasive range, all genetic diversity measures in the source populations were significantly more diverse (Table 3.1). While there were 9.91 multilocus genotypes in the sources, there were only 3.07 in the invasive range ($P = 1.28 \times 10^{-4}$). The invasive range exhibited lower allelic richness ($N_{A(PUTATIVE)} = 1.51$, $N_{A(INVASIVE)} = 1.44$; $P = 1.01 \times 10^{-3}$), lower effective number of alleles ($N_{E(PUTATIVE)} = 2.22$, $N_{E(INVASIVE)} = 1.71$;

$P = 5.77 \times 10^{-6}$), and both elevated observed and expected heterozygosities ($H_{O(PUTATIVE)} = 0.472$, $H_{O(INVASIVE)} = 0.449$, $P = 0.580$; $uH_{E(PUTATIVE)} = 0.349$; $H_{O(INVASIVE)} = 0.308$, $P = 0.134$).

Shift in reproductive strategy

At the marker level, both the invasive and native ranges demonstrated evidence of clonality. Tests for panmixia failed in 89% and 86% of populations in Asia and the US, respectively, in addition to the rejection of the null hypothesis of linkage equilibrium among markers for both indices of association, I_A ($I_{A(US)} = 0.888$, $P = 9.99 \times 10^{-4}$; $I_{A(ASIA)} = 1.67$, 9.99×10^{-4}) and rbarD ($rbarD_{(US)} = 0.0680$, 9.99×10^{-4}; $rbarD_{(ASIA)} = 0.0840 = 9.99 \times 10^{-4}$).

Differences in clonality were observed at the range and population levels. Overall, there were 935 multilocus genotypes detected in the 4928 samples studied. Specifically, there were 685 genotypes in 3187 native samples, indicating that approximately 80% of Asian samples were clones of genetically distinct multilocus genotypes. 85% of invasive samples were clones with only 274 multilocus genotypes detected in 1741 samples. Thus, using the complement of the proportion distinguishable (R) as a proxy for vegetative rates, the estimated rate of clonal reproduction (1-R) was 79% in the native range but 8% higher in the United States with a clonal rate of 85%.

On the population level, the average proportion distinguishable in the United States was significantly lower than the entire native range ($R_{(US)} = 0.108$, $R_{(ASIA)} = 0.183$; $P = 0.0081$) (Table 3.1; Table 3.2). Moreover, populations in the US exhibited a significant 38% reduction in the number of effective multilocus genotypes per population ($P = 0.0037$) with an average of 1.87 effective multilocus genotypes in the US as compared to 3.04 in Asia (Table 3.1). Invasive kudzu also demonstrated a significant 37% loss of clonal diversity per population (div = 0.266) compared to the native range (div = 0.422) ($P = 0.0035$) (Table 3.1). The invasive range

exhibited a higher fixation index (P = 0.003) with an average value of -0.612 compared to -0.413 in Asia, which was 33% lower than the US. There was no significant difference in the evenness of multilocus genotypes in a population (P = 0.3591) (Table 3.1).

When the native range was parsed into two groups, the putative sources and the rest of the native (i.e. non-putative), genetic patterns changed. There were significant differences across all measures except evenness (Table 3.1). In addition to significantly lower proportion distinguishable ($R_{PUTATIVE}$ = 0.377, $R_{NONPUTATIVE}$ = 0.167; P = 0.0499) and higher clonal diversity ($div_{PUTATIVE}$ = 0.700, $div_{NONPUTATIVE}$ = 0.399; P = 4.92 x 10^{-4}) in the putative sources, the fixation index was also significantly higher (P = 1.72 x 10^{-5}) with a 128% difference between an $F_{PUTATIVE}$ of -0.0973 and $F_{NONPUTATIVE}$ of -0.440.

Clonal diversity measures were also significantly different between the invasive populations and putative sources (Table 3.1). The invasive range exhibited significant reductions in clonal diversity ($div_{INVASIVE}$ = 0.266, $div_{PUTATIVE}$ = 0.700; P = 1.93 X 10^{-4}) and the proportion distinguishable ($R_{INVASIVE}$ = 0.108, $R_{PUTATIVE}$ = 0.377; P = 1.29 x 10^{-3}). The evenness of clonal lineages in a population was significantly higher in the invasive range ($eve_{INVASIVE}$ = 0.764, $eve_{PUTATIVE}$ = 0.664; P = 0.0126).

Lastly, the 11 putative sources detected in Figure 3.12 exhibited a significantly lower rate of clonality than the rest of the native range (1-$R_{PUTATIVE}$ = 0.623; 1-$R_{NON\text{-}PUTATIVE}$ = 0.836; P = 4.71 x 10^{-2}) (Table 3.1). However, after introduction of those putative sources to the United States, clonal rates shifted significantly from 62% as observed in the putative sources to 89% clonal rates in the invasive range (1-$R_{PUTATIVE}$ = 0.623; 1-$R_{INVASIVE}$ = 0.892; P = 4.12 x 10^{-4}, W = 782), an invasive clonal rate that surpassed even the rates of the rest of the native range (1-$R_{NON\text{-}PUTATIVE}$ = 0.836; 1-$R_{INVASIVE}$ = 0.892; P = 0.0317).

Discussion

In this study, we utilized the geographic distribution of genetic diversity to infer the invasion history of this species. Genetic variation was well structured within both ranges but not well differentiated between ranges such that putative sources could be pinpointed and the number of introductions inferred. Genetic assignment of US populations to different groups across Asia from both nuclear and chloroplast data demonstrated that kudzu was introduced multiple times with as many as 11 introduction events (Figure 3.12). In fact, individuals from the 11 putative sources were found in all genetic clusters detected in the United States (Figure 3.4).

As the first study to link populations in the US back to their native sources, several previous hypotheses were refuted. Firstly, results negated the historical belief that invasive kudzu originated in Japan. The introduction of kudzu to the US involved China, Japan and South Korea (Figure 3.12). Secondly, the widespread geographic distribution of the putative sources eliminated the hypothesis of a single, large introduction (Figure 3.12), supporting multiple introductions from disparate sources instead. This finding highlight the importance of comparing genetic patterns between native and invasive ranges, because when taken in isolation, the detection of only a few hundred multilocus genotypes, two-thirds of which possessed only two haplotypes, suggested that a single, large introduction was possible in Chapter 2.

Despite multiple introductions, reduced diversity was evident in the United States, when genetic and clonal diversity patterns were compared among ranges. There was a significant reduction in the number and effective number of multilocus genotypes, higher fixation indices and lower clonal diversity per population in the invasive range as compared to Asia (Table 3.1). Additionally, chloroplast diversity was lower in the US with fewer haplotypes (Figure 3.10), segregating sites and nucleotide diversity per site than the native range.

However, the total number of multilocus genotypes was higher in the invasive range, indicating that diversity patterns in the US likely did not emerge from bottleneck effects. Moreover, an introduced pool of individuals is not bottlenecked if it is has more genetically distinct individuals than its sources. Even with the possibility of ghost populations, there was almost twice the number of multilocus genotypes in the United States than the 11 putative sources. Despite what looked like genetic bottlenecking at the chloroplast locus, reduced chloroplast diversity was due to sampling from two of most common haplotypes spread across the entire central-northern region of the native range (Figure 3.11). Moreover, measures most sensitive to genetic bottlenecks (Allendorf 1986), allelic richness and heterozygosity, did not change in the invasive range (Table 3.1). Changes in the other diversity measures likely emerged from other evolutionary or ecological factors.

These factors become clearer after understanding that the putative sources were significantly more diverse than the rest of the native range (Table 3.1). Putative sources were not only significantly more diverse than the rest of the native range, but also reflective of sexual reproduction. With a significant increase in multilocus genotypes per population, higher allelic richness, a larger effective number of alleles and a 128% difference in the fixation index close to 0 (Table 3.1), there was strong indication of increased sexual reproduction as compared to non-putative sources.

The higher diversity and sexual rates in the putative source populations shaped the high heterozygosity and doubling of multilocus genotypes found in the invasive range. Outcrossing among heterozygous individuals ($H_{O(PUTATIVE)} = 0.415$) would not necessarily reduce heterozygosity levels in progeny if parents were genetically differentiated across loci and would augment the number of distinct multilocus genotypes in the next generation. Introduction of

these seeds would explain why both observed and expected heterozygosities were not significantly altered from levels in the putative sources (Table 3.1). In fact, it would have taken less than a dozen viable seeds from half of the 125 multilocus genotypes in the putative sources to observe a similar number of multilocus genotypes as the invasive range. Occasional population admixture from sexual reproduction observed in the invasive range (Figure 3.4) as well as rapid range expansion could have contributed to the increase in invasive individuals as well.

While bottlenecking did not appear to be a main driver of diversity patterns, changes in diversity statistics in the invasive range is most likely the product of a shift towards higher asexuality in the US. As a species, kudzu was largely clonal with an average rate around 80%. Both ranges exhibited consistent evidence of high clonality with excess heterozygosity, low genotypic diversity and negative inbreeding coefficients (Balloux et al. 2003). Moreover, null hypotheses of panmixia and linkage equilibrium were rejected for both ranges. However, the rate of clonality (1-R) in the US was 8% higher than the entire native range and approximately 20% higher than the putative sources with a rate of 85%. Clonal rates even went significantly beyond those clonal rates of the rest of the native range ($P = 0.0317$; one-tailed Wilcoxon rank sum test). There was a significant decrease in fixation indices, multilocus genotypes per population, proportion distinguishable and clonal diversity and an increase in evenness of lineages after diverse putative sources colonized the United States (Table 3.1). These patterns imply that while there was more sexual reproduction in the putative sources than the rest of the native range, upon introduction to the United States, kudzu likely transitioned to extreme levels of clonality beyond what was observed for the rest of the species.

If this were the case, an increase in clonality in the United States would exert a major influence in driving the spread of kudzu. A single individual is capable of producing hundreds to thousands of ramets (Forseth and Innis 2004), and increased rates of vegetative reproduction would quicken geographic spread through reproductive assurance (Sakai et al. 2001, Kalisz et al. 2004). Additionally, if a particular lineage was well adapted to the novel environment, increased clonality would be highly beneficial by eliminating recombination (Facon et al. 2006, Barrett et al. 2008). Consequently, rapid adaptive evolution from new mutation would be an unlikely mechanism of invasion for kudzu, as a lack of recombination cannot spread new alleles into other genetic backgrounds (Barrett and Schluter 2008). Rather, pre-adaptation or selection on standing genetic variation would be more likely to produce invasiveness since high clonal reproduction preserves co-adapted gene complexes.

Alternatively, it is possible that genetic patterns in the United States simply reflect the nature of geographic spread in the invasive range. Since the invasive range expands 50,000 hectares per year, asymmetrical gene flow into new rather than pre-existing populations is expected and would decrease the number of multilocus genotypes in a population, thereby altering diversity statistics in Table 3.1. However, we observe similar overall proportion distinguishable ($R_{US} = 0.15$; $R_{ASIA} = 0.21$) and similar genetic diversity (Table 3.1) with the majority of the native range, in which patterns would not the product of rapid range expansion. Furthermore, we do not observe spatial patterning to the diversity statistics. If patterns of diversity were driven by gene flow towards the expanding edge, the values of each diversity statistic would not be so randomly distributed (Table 3.2). We would expect to see a reduction of diversity at the periphery of the range (Sexton et al. 2009), which we did not observe (results not shown). Thus, the hypothesis of increased vegetative reproduction rather than gene flow into

new areas better accounts for patterns in the invasive range. Common garden studies comparing fitness traits and the frequency of vegetative versus sexual reproduction among populations of different sample sizes and in a variety of invasive habitats are needed in order to test this hypothesis.

Tables

Table 3.1. Summary of means and standard deviations (SD) of each genetic and clonal diversity measure in addition to sample size. P-values for comparisons of means are included with significant P-values ($P < 0.05$) bolded. Comparisons included the "among ranges" which compares overall values between the native and invasive ranges, "among native groups", specifically the non-putative sources and putative sources, and "among US and sources" putative sources versus the invasive range. The means and standard deviations of these measures on the species level, "Total", are also displayed. * indicates that comparisons were based on re-sampling 1000 times to match sample size to the lower group number.

Diversity	Asia		US		Among ranges	Non-putative Sources		Among native groups	Putative Sources		Among US and putative sources	Total	
Measure	Mean	SD	Mean	SD	P-value	Mean	SD	P-value*	Mean	SD	P-value*	Mean	SD
N	22.4	4.26	20.0	5.33	**< 1.00 x 10^{-4}**	22.25	4.37	**6.45 x 10^{-7}**	24.73	1.27	**1.38 x 10^{-8}**	21.52	4.83
MLG	4.70	4.25	3.07	3.14	**4.70 x 10^{-3}**	4.27	3.74	**0.0123**	9.91	6.39	**1.28 x 10^{-4}**	4.08	3.94
N_A	1.44	0.133	1.44	0.126	0.984	1.43	0.135	**2.35 x 10^{-4}**	1.51	0.0676	**1.01 x 10^{-3}**	1.44	0.130
N_E	1.82	0.468	1.71	0.395	0.0956	1.79	0.460	**1.53 x 10^{-4}**	2.22	0.390	**5.77 x 10^{-6}**	1.78	0.444
H_O	0.419	0.140	0.449	0.118	0.0823	0.415	0.143	**8.25 x 10^{-3}**	0.472	0.0734	0.580	0.430	0.132
uH_E	0.337	0.138	0.308	0.137	0.819	0.336	0.139	0.587	0.349	0.122	0.134	0.326	0.138
F	-0.413	0.534	-0.612	0.429	**3.00 x 10^{-3}**	-0.440	0.540	**1.72 x 10^{-5}**	-0.0973	0.343	**3.16 x 10^{-6}**	-0.489	0.505
eff	3.04	2.92	1.87	1.51	**3.70 x 10^{-3}**	2.72	2.41	**0.0334**	6.90	5.18	**1.23 x 10^{-3}**	2.60	2.54
div	0.422	0.380	0.266	0.321	**3.50 x 10^{-3}**	0.399	0.375	**4.92 x 10^{-4}**	0.700	0.343	**1.93 x 10^{-4}**	0.363	0.366
eve	0.733	0.238	0.764	0.251	0.359	0.739	0.242	0.0760	0.664	0.172	**0.0126**	0.745	0.243
R	0.183	0.204	0.108	0.158	**8.10 x 10^{-3}**	0.167	0.190	**0.0499**	0.377	0.270	**1.29 x 10^{-3}**	0.154	0.191
1-R	0.817	0.204	0.892	0.158	**8.10 x 10^{-3}**	0.833	0.190	**0.0471**	0.623	0.270	**0.0317**	0.846	0.191

Table 3.2. Genetic and clonal diversity measures for all 229 populations. Information for each population includes latitude, longitude, country of origin, sample size (N), the number of multilocus genotypes (MLG), allelic richness (N_A), effective number of alleles (N_E), observed heterozygosity (H_O), expected heterozygosity (uH_E), fixation index (F), effective number of genotypes (eff), Nei's genetic diversity (div) in this case referring to clonal diversity, evenness (eve) and proportion distinguishable (R).

Population	Latitude	Longitude	Country of origin	N	MLG	N_A	N_E	H_O	uH_E	F	eff	div	eve	R
AL10	31.512783	-86.702148	United States	7	1	1.60	1.60	0.600	0.600	-1.00	1.00	0.00	1.00	0.00
AL11	31.222782	-85.394867	United States	23	1	1.53	1.53	0.533	0.533	-1.00	1.00	0.00	1.00	0.00
AL12	33.17025	-87.53775	United States	20	9	1.55	2.52	0.538	0.300	-6.10×10^{-2}	5.13	0.847	0.570	0.421
AL13	32.584816	-88.185951	United States	15	2	1.34	1.51	0.500	0.267	-0.917	1.99	0.533	0.996	7.14×10^{-2}
AL14	33.648048	-85.426285	United States	19	2	1.32	1.46	0.300	0.200	-0.489	1.11	0.105	0.555	5.56×10^{-2}
AL3	33.524685	-86.825119	United States	22	6	1.31	1.55	0.353	0.516	-0.192	4.17	0.797	0.596	0.238
AL5	33.610401	-85.76815	United States	22	12	1.36	1.77	0.423	0.258	-0.239	9.31	0.935	0.776	0.524
AL6	32.454117	-87.944	United States	15	6	1.47	1.87	0.579	0.278	-0.328	2.14	0.571	0.357	0.357
AL7	32.431782	-87.037987	United States	22	1	1.47	1.47	0.467	0.347	-1.00	1.00	0.00	1.00	0.00
AL9	32.328934	-86.341347	United States	24	2	1.34	1.51	0.500	0.428	-0.917	1.09	8.30×10^{-2}	0.543	4.35×10^{-2}
AR1	35.81686	-92.550194	United States	23	1	1.67	1.67	0.667	0.233	-1.00	1.00	0.00	1.00	0.00
AR2	34.066254	-93.689255	United States	25	2	1.59	2.00	0.533	0.258	-0.221	1.57	0.380	0.787	4.17×10^{-2}
C	21.498606	101.537704	China	10	8	1.50	2.32	0.202	0.333	0.630	8.33	0.978	0.926	0.778
FL1	30.958817	-86.296432	United States	23	1	1.47	1.47	0.467	0.442	-1.00	1.00	0.00	1.00	0.00
FL2	30.782818	-85.37545	United States	22	1	1.27	1.27	0.267	0.461	-1.00	1.00	0.00	1.00	0.00
FL3	28.560625	-81.02325	United States	18	2	1.26	1.37	0.367	0.233	-0.889	1.12	0.111	0.559	5.88×10^{-2}
GA10	33.62965	-84.451385	United States	24	1	1.27	1.27	0.267	0.133	-1.00	1.00	0.00	1.00	0.00
GA14	34.687099	-83.422363	United States	18	1	1.33	1.33	0.333	0.192	-1.00	1.00	0.00	1.00	0.00
GA17	34.446316	-83.119232	United States	26	5	1.49	2.02	0.537	0.133	-0.216	2.32	0.591	0.463	0.160
GA18	33.921417	-83.406235	United States	26	1	1.47	1.47	0.467	0.167	-1.00	1.00	0.00	1.00	0.00

GA2	33.169716	-83.2808	United States	19	10	1.57	2.64	0.464	0.443	0.119	4.57	0.825	0.457	0.500
GA3	32.329033	-81.933052	United States	22	2	1.62	2.31	0.467	0.233	5.60×10^{-2}	1.86	0.485	0.931	4.76×10^{-2}
GA32	30.896784	-83.969666	United States	24	1	1.53	1.53	0.533	0.536	-1.00	1.00	0.00	1.00	0.00
GA33	31.586634	-84.151535	United States	23	2	1.34	1.51	0.500	0.467	-0.917	1.19	0.166	0.594	4.55×10^{-2}
GA34	32.077068	-83.763084	United States	24	1	1.47	1.47	0.467	0.267	-1.00	1.00	0.00	1.00	0.00
GA35	32.8517	-83.619781	United States	25	1	1.27	1.27	0.267	0.258	-1.00	1.00	0.00	1.00	0.00
GA36	33.292648	-83.376297	United States	24	7	1.53	2.52	0.411	0.233	0.142	3.00	0.696	0.375	0.261
GA37	34.666569	-84.94297	United States	21	9	1.51	2.40	0.520	0.133	-7.80×10^{-2}	4.46	0.814	0.495	0.400
GA39	33.947475	-83.380424	United States	24	1	1.40	1.40	0.400	0.498	-1.00	1.00	0.00	1.00	0.00
GA4	32.530434	-82.901436	United States	13	2	1.62	2.02	0.533	0.484	-0.162	1.17	0.154	0.583	8.33×10^{-2}
GA6	32.78175	-84.234581	United States	16	2	1.62	2.20	0.533	0.467	-0.108	1.44	0.325	0.719	6.67×10^{-2}
GA7	33.528767	-82.034782	United States	21	1	1.27	1.27	0.267	0.208	-1.00	1.00	0.00	1.00	0.00
GA9	33.199219	-83.437714	United States	23	2	1.39	1.59	0.467	0.467	-0.585	1.09	8.70×10^{-2}	0.545	4.55×10^{-2}
GA96	34.413933	-84.109383	United States	25	7	1.51	2.15	0.562	0.133	-0.190	3.79	0.767	0.541	0.250
H	30.255589	120.162964	China	14	7	1.64	2.92	0.520	0.292	7.60×10^{-2}	5.44	0.879	0.778	0.462
KAN1	30.013973	118.776627	China	23	14	1.58	2.74	0.566	0.475	-4.40×10^{-2}	7.67	0.909	0.511	0.591
KAN2	29.691074	118.298035	China	29	3	1.38	1.63	0.578	0.589	-0.787	1.61	0.394	0.538	7.14×10^{-2}
KAN3	31.093113	118.282066	China	24	2	1.34	1.51	0.500	0.554	-0.917	1.18	0.159	0.590	4.35×10^{-2}
KAN4	30.494402	117.929306	China	26	2	1.26	1.37	0.367	0.319	-0.889	1.08	7.70×10^{-2}	0.540	4.00×10^{-2}
KAN5	30.457275	117.192535	China	24	1	1.47	1.47	0.467	0.258	-1.00	1.00	0.00	1.00	0.00
KAN6	30.837708	117.884674	China	12	2	1.28	1.37	0.367	0.192	-0.889	1.18	0.167	0.590	9.09×10^{-2}
KAN8	29.720846	118.322479	China	22	1	1.47	1.47	0.467	0.233	-1.00	1.00	0.00	1.00	0.00
KAN9	30.119846	117.494453	China	20	14	1.65	3.14	0.506	0.192	0.188	11.8	0.963	0.735	0.684
KB1	40.408054	116.546944	China	23	1	1.40	1.40	0.400	0.233	-1.00	1.00	0.00	1.00	0.00
KCQ1	30.652222	107.818054	China	24	7	1.55	2.31	0.360	0.623	0.357	2.97	0.692	0.424	0.261
KCQ2	29.8375	106.384445	China	22	1	1.67	1.67	0.667	0.200	-1.00	1.00	0.00	1.00	0.00
KFU1	27.756647	118.035309	China	22	2	1.43	1.77	0.433	0.510	-0.333	1.94	0.506	0.968	4.76×10^{-2}
KFU2	24.894583	118.5904	China	24	3	1.47	1.81	0.222	0.333	0.501	2.13	0.554	0.711	8.70×10^{-2}

KFU3	24.722841	118.152206	China	25	6	1.36	1.79	0.241	0.325	0.186	3.98	0.780	0.569	0.208
KFU4	27.331877	118.120461	China	24	1	1.27	1.27	0.267	0.393	-1.00	1.00	0.00	1.00	0.00
KFU5	25.056	118.186264	China	22	1	1.40	1.40	0.400	0.323	-1.00	1.00	0.00	1.00	0.00
KFU6	25.004435	117.534065	China	24	6	1.31	1.65	0.422	0.133	-0.471	2.88	0.681	0.480	0.217
KFU7	25.026758	117.692093	China	24	8	1.29	1.68	0.141	0.200	0.388	5.54	0.855	0.615	0.304
KFU8	24.625448	117.759155	China	24	3	1.56	2.07	0.400	0.282	0.176	1.19	0.163	0.395	8.70×10^{-2}
KGD1	23.09649	114.09604	China	24	1	1.13	1.13	0.133	0.268	-1.00	1.00	0.00	1.00	0.00
KGD2	23.406	113.417	China	22	4	1.33	1.70	0.283	0.467	1.90×10^{-2}	1.47	0.333	0.367	0.143
KGD3	24.902	112.416	China	24	6	1.36	1.73	0.206	6.70×10^{-2}	0.319	3.03	0.699	0.433	0.217
KGD4	23.96	113.606	China	24	9	1.60	2.47	0.294	0.290	0.467	3.43	0.739	0.381	0.348
KGD5	23.059	112.667	China	13	1	1.13	1.13	0.133	0.320	-1.00	1.00	0.00	1.00	0.00
KGU1	25.605947	110.66391	China	27	1	1.33	1.33	0.333	0.561	-1.00	1.00	0.00	1.00	0.00
KGU3	25.291628	110.354187	China	23	1	1.33	1.33	0.333	6.70×10^{-2}	-1.00	1.00	0.00	1.00	0.00
KGU4	24.142021	107.258652	China	3	2	1.49	1.85	0.378	0.167	-2.40×10^{-2}	3.00	1.00	1.00	0.500
KGU5	24.775145	110.499199	China	27	6	1.40	2.10	0.188	0.167	0.434	2.88	0.678	0.412	0.192
KGZH1	27.235033	108.879167	China	24	1	1.60	1.60	0.600	0.367	-1.00	1.00	0.00	1.00	0.00
KHAI1	19.52791667	109.5522167	China	23	5	1.54	2.28	0.312	0.368	0.419	2.83	0.676	0.566	0.182
KHE1	34.573271	112.940967	China	27	1	1.53	1.53	0.533	0.300	-1.00	1.00	0.00	1.00	0.00
KHK1	22.358999	114.117996	China	26	1	1.20	1.20	0.200	0.486	-1.00	1.00	0.00	1.00	0.00
KHN10	29.302222	110.435833	China	23	6	1.60	2.55	0.633	0.267	-0.156	4.52	0.814	0.754	0.227
KHN11	26.6904	113.253067	China	20	2	1.26	1.37	0.367	0.100	-0.889	1.11	0.100	0.552	5.26×10^{-2}
KHN12	26.244167	111.659333	China	20	1	1.47	1.47	0.467	0.549	-1.00	1.00	0.00	1.00	0.00
KHN2	28.189655	113.18631	China	24	1	1.47	1.47	0.467	0.192	-1.00	1.00	0.00	1.00	0.00
KHN3	27.254799	112.655746	China	24	8	1.60	2.81	0.512	0.233	7.80×10^{-2}	4.00	0.783	0.500	0.304
KHN4	28.636944	109.964996	China	23	7	1.53	2.33	0.470	0.233	-3.00×10^{-2}	3.55	0.751	0.507	0.273
KHN5	27.017117	112.837433	China	24	3	1.50	1.83	0.733	0.560	-0.744	1.88	0.489	0.627	8.70×10^{-2}
KHN6	27.705299	113.035484	China	23	1	1.33	1.33	0.333	0.494	-1.00	1.00	0.00	1.00	0.00
KHN7	27.039722	110.602501	China	23	1	1.27	1.27	0.267	0.411	-1.00	1.00	0.00	1.00	0.00

KHN8	27.310556	110.114998	China	24	1	1.20	1.20	0.200	0.167	-1.00	1.00	0.00	1.00	0.00
KHN9	28.164167	109.65139	China	23	6	1.62	2.76	0.633	0.133	-0.121	4.37	0.806	0.729	0.227
KHU2	31.718417	111.766383	China	23	17	1.54	2.70	0.392	0.100	0.231	14.3	0.972	0.794	0.727
KHU3	30.280583	108.83575	China	20	10	1.57	2.72	0.600	0.569	-9.70×10^{-2}	5.26	0.853	0.526	0.474
KJI1	31.283726	119.437347	China	12	3	1.42	1.72	0.422	0.544	-0.179	2.00	0.545	0.667	0.182
KJI2	31.321274	119.809509	China	25	8	1.45	2.04	0.508	0.183	-0.114	4.25	0.797	0.531	0.292
KJI3	30.976746	120.675201	China	25	1	1.60	1.60	0.600	0.352	-1.00	1.00	0.00	1.00	0.00
KJI4	32.057243	118.796089	China	22	4	1.37	1.68	0.500	0.421	-0.542	3.10	0.710	0.776	0.143
KJI5	31.804064	119.333678	China	22	9	1.49	2.26	0.329	0.300	0.278	6.05	0.874	0.605	0.381
KJIX1	27.220718	115.015915	China	20	7	1.37	1.84	0.377	0.323	-0.114	4.08	0.795	0.510	0.316
KJIX2	29.347786	116.728363	China	27	1	1.60	1.60	0.600	0.459	-1.00	1.00	0.00	1.00	0.00
KJP1	35.783031	140.379944	Japan	23	6	1.51	2.22	0.478	0.341	-1.30×10^{-3}	2.74	0.664	0.457	0.227
KJP10	33.872616	130.770782	Japan	25	6	1.42	2.02	0.467	0.300	-0.235	5.00	0.833	0.833	0.208
KJP11	34.198157	131.346245	Japan	26	12	1.49	2.12	0.468	0.464	-3.70×10^{-2}	8.45	0.917	0.704	0.440
KJP12	34.1981420	132.1199800	Japan	25	1	1.33	1.33	0.333	0.387	-1.00	1.00	0.00	1.00	0.00
KJP13	34.490881	133.313506	Japan	23	1	1.40	1.40	0.400	0.465	-1.00	1.00	0.00	1.00	0.00
KJP14	34.772608	134.083658	Japan	26	8	1.53	2.28	0.557	0.167	-0.142	2.49	0.622	0.311	0.280
KJP15	34.817131	134.980072	Japan	25	1	1.47	1.47	0.467	0.200	-1.00	1.00	0.00	1.00	0.00
KJP16	34.952258	135.7452	Japan	25	1	1.47	1.47	0.467	0.499	-1.00	1.00	0.00	1.00	0.00
KJP17	35.018889	136.591439	Japan	25	7	1.28	1.53	0.324	0.233	-0.151	3.98	0.780	0.569	0.250
KJP18	34.782522	137.606936	Japan	25	3	1.44	1.94	0.533	0.233	-0.430	2.04	0.530	0.679	8.33×10^{-2}
KJP19	34.776094	138.000653	Japan	25	3	1.56	2.29	0.578	0.257	-0.252	2.64	0.647	0.879	8.33×10^{-2}
KJP2	35.545953	139.762492	Japan	26	3	1.50	1.86	0.556	0.370	-0.278	1.17	0.151	0.390	8.00×10^{-2}
KJP20	35.143611	138.837461	Japan	24	1	1.53	1.53	0.533	0.467	-1.00	1.00	0.00	1.00	0.00
KJP21	36.317514	139.617706	Japan	24	10	1.50	2.26	0.484	0.419	-4.00×10^{-2}	4.24	0.797	0.424	0.391
KJP22	37.127164	140.179611	Japan	24	14	1.56	2.39	0.519	0.267	2.50×10^{-2}	6.70	0.888	0.478	0.565
KJP23	37.908303	140.5771	Japan	19	5	1.53	2.20	0.447	0.478	5.00×10^{-3}	1.78	0.462	0.356	0.222
KJP24	38.394494	141.070542	Japan	25	16	1.53	2.55	0.507	0.542	-2.30×10^{-2}	10.6	0.943	0.662	0.625

KJP25	38.853267	141.581086	Japan	25	16	1.52	2.30	0.452	0.473	0.120	12.3	0.957	0.766	0.625
KJP26	39.240347	141.077572	Japan	24	4	1.50	2.12	0.567	0.510	-0.311	3.56	0.750	0.889	0.130
KJP27	39.927781	140.959858	Japan	24	6	1.43	1.90	0.390	0.508	1.30×10^{-2}	3.35	0.732	0.558	0.217
KJP28	40.457922	140.635097	Japan	26	18	1.53	2.54	0.445	0.442	0.124	13.5	0.963	0.751	0.680
KJP29	40.009281	140.085758	Japan	25	2	1.30	1.44	0.433	0.395	-0.905	1.37	0.280	0.684	4.17×10^{-2}
KJP3	33.579283	130.459439	Japan	25	12	1.41	1.90	0.357	0.512	2.70×10^{-2}	9.62	0.933	0.801	0.458
KJP4	32.808378	130.794017	Japan	20	1	1.53	1.53	0.533	0.225	-1.00	1.00	0.00	1.00	0.00
KJP5	32.069669	130.796642	Japan	24	7	1.58	2.69	0.552	0.388	-2.00×10^{-2}	3.27	0.725	0.468	0.261
KJP6	31.721436	130.800789	Japan	25	4	1.38	1.80	0.400	0.267	-0.200	2.00	0.520	0.499	0.125
KJP7	31.9	131.416667	Japan	24	6	1.56	2.55	0.522	0.539	-3.70×10^{-2}	3.69	0.761	0.615	0.217
KJP8	32.9018	131.788769	Japan	24	11	1.52	2.48	0.462	0.333	4.80×10^{-2}	6.70	0.888	0.609	0.435
KJP9	33.248128	131.247483	Japan	25	10	1.52	2.52	0.530	0.510	-0.117	6.44	0.880	0.644	0.375
KKO1	35.274719	129.181335	South Korea	25	4	1.35	1.73	0.433	0.495	-0.396	1.53	0.360	0.382	0.125
KKO10	38.184169	128.59238	South Korea	24	3	1.40	1.68	0.444	0.489	-0.337	1.41	0.304	0.471	8.70×10^{-2}
KKO11	38.236259	127.442061	South Korea	24	1	1.60	1.60	0.600	0.308	-1.00	1.00	0.00	1.00	0.00
KKO12	37.736275	126.487784	South Korea	24	1	1.73	1.73	0.733	0.333	-1.00	1.00	0.00	1.00	0.00
KKO13	36.786028	126.452057	South Korea	24	2	1.34	1.51	0.500	0.300	-0.917	1.09	8.30×10^{-2}	0.543	4.35×10^{-2}
KKO14	35.909856	126.631293	South Korea	24	2	1.30	1.44	0.433	0.367	-0.905	1.09	8.30×10^{-2}	0.543	4.35×10^{-2}
KKO15	35.089789	126.157468	South Korea	24	4	1.43	1.89	0.500	0.258	-0.333	2.72	0.659	0.679	0.130
KKO16	37.340006	127.931021	South Korea	24	1	1.73	1.73	0.733	0.225	-1.00	1.00	0.00	1.00	0.00
KKO17	36.422759	128.146956	South Korea	24	1	1.40	1.40	0.400	0.375	-1.00	1.00	0.00	1.00	0.00
KKO18	35.529485	127.731214	South Korea	26	3	1.55	2.03	0.556	0.367	-0.157	1.17	0.151	0.390	8.00×10^{-2}
KKO19	34.593831	127.801933	South Korea	24	9	1.48	2.31	0.452	0.200	7.30×10^{-2}	5.54	0.855	0.615	0.348
KKO2	35.665127	129.40506	South Korea	23	1	1.40	1.40	0.400	0.456	-1.00	1.00	0.00	1.00	0.00
KKO20	34.311039	126.76093	South Korea	24	4	1.36	1.73	0.550	0.450	-0.807	1.82	0.471	0.456	0.130
KKO21	33.428666	126.276898	South Korea	24	1	1.13	1.13	0.133	0.200	-1.00	1.00	0.00	1.00	0.00
KKO22	33.250904	126.43174	South Korea	23	7	1.25	1.50	0.286	0.315	-0.154	2.74	0.664	0.392	0.273
KKO3	37.497459	127.065323	South Korea	23	10	1.46	2.00	0.500	0.067	-0.158	7.45	0.905	0.745	0.409

KKO4	35.379822	129.25798	South Korea	23	1	1.53	1.53	0.533	0.233	-1.00	1.00	0.00	1.00	0.00
KKO5	35.116118	129.106865	South Korea	24	11	1.51	2.45	0.505	0.441	-3.20×10^{-2}	4.88	0.830	0.444	0.435
KKO6	36.029059	129.3477432	South Korea	23	10	1.48	2.09	0.461	0.267	7.00×10^{-3}	5.69	0.862	0.569	0.409
KKO7	36.980769	129.394677	South Korea	24	1	1.47	1.47	0.467	0.490	-1.00	1.00	0.00	1.00	0.00
KKO8	37.756466	128.878212	South Korea	24	2	1.39	1.57	0.567	0.454	-0.926	1.18	0.159	0.590	4.35×10^{-2}
KKO9	38.075139	128.629732	South Korea	24	1	1.40	1.40	0.400	0.233	-1.00	1.00	0.00	1.00	0.00
KLN1	41.862	123.651001	China	22	1	1.53	1.53	0.533	0.292	-1.00	1.00	0.00	1.00	0.00
KLN2	39.103611	121.819444	China	16	1	1.53	1.53	0.533	0.200	-1.00	1.00	0.00	1.00	0.00
KSA1	34.525742	110.103645	China	24	11	1.49	2.23	0.570	0.267	-0.245	7.02	0.895	0.639	0.435
KSA2	34.49778	110.093056	China	27	15	1.51	2.40	0.433	0.267	0.128	11.2	0.946	0.748	0.539
KSA3	34.0219	108.819038	China	21	9	1.55	2.43	0.431	0.470	0.140	8.32	0.924	0.925	0.400
KSA4	33.76239	108.763756	China	21	12	1.57	2.69	0.449	0.494	0.135	10.3	0.948	0.855	0.550
KSC1	29.601198	103.484505	China	19	5	1.59	2.30	0.461	0.514	0.128	2.36	0.608	0.337	0.222
KSC2	29.737631	103.571655	China	22	1	1.40	1.40	0.400	0.545	-1.00	1.00	0.00	1.00	0.00
KSC3	29.230425	103.262047	China	9	2	1.33	1.48	0.433	0.528	-0.848	1.53	0.389	0.764	0.125
KSC4	30.653355	104.079041	China	21	5	1.47	2.08	0.187	0.200	0.490	1.86	0.486	0.372	0.200
KSD1	36.480999	117.913002	China	23	2	1.34	1.51	0.500	0.233	-0.917	1.09	8.70×10^{-2}	0.545	4.55×10^{-2}
KSD2	36.407117	117.503006	China	21	1	1.47	1.47	0.467	0.424	-1.00	1.00	0.00	1.00	0.00
KSD3	36.205557	117.097436	China	23	3	1.43	1.94	0.356	0.258	-1.40×10^{-2}	2.04	0.534	0.681	9.09×10^{-2}
KSH1	31.24869	121.447029	China	14	1	1.33	1.33	0.333	0.233	-1.00	1.00	0.00	1.00	0.00
KTW1	23.83	120.782997	China	23	5	1.21	1.40	0.113	0.359	0.312	1.60	0.391	0.320	0.182
KTW4	25.155725	121.547745	China	24	3	1.27	1.42	0.178	0.167	0.164	1.52	0.359	0.508	8.70×10^{-2}
KTW5	22.719389	120.362472	China	24	2	1.08	1.11	0.0330	0.183	0.333	1.09	8.30×10^{-2}	0.543	4.35×10^{-2}
KY10	37.343876	-84.313889	United States	6	2	1.44	1.75	0.489	0.222	-0.561	2.57	0.733	0.857	0.200
KY11	37.179298	-83.596069	United States	22	3	1.48	1.83	0.467	0.0580	-0.220	1.47	0.333	0.367	9.52×10^{-2}
KY7	38.268234	-85.502586	United States	19	1	1.20	1.20	0.200	0.319	-1.00	1.00	0.00	1.00	0.00
KY8	37.38015	-82.432785	United States	24	6	1.37	1.64	0.341	0.382	4.80×10^{-2}	1.93	0.504	0.276	0.217
KY9	37.446583	-82.906631	United States	25	2	1.48	1.74	0.400	0.100	-0.127	1.08	8.00×10^{-2}	0.542	4.17×10^{-2}

KYU1	21.402008	101.617455	China	15	5	1.59	2.22	0.247	0.338	0.588	2.92	0.705	0.584	0.286
KYU10	25.7777	100.131833	China	23	1	1.40	1.40	0.400	0.358	-1.00	1.00	0.00	1.00	0.00
KYU2	21.922548	101.267967	China	18	6	1.25	1.67	0.237	0.528	-4.60×10^{-2}	2.89	0.693	0.413	0.294
KYU3	22.029682	100.852379	China	20	1	1.07	1.07	0.0670	0.200	-1.00	1.00	0.00	1.00	0.00
KYU4	21.877319	101.380898	China	23	8	1.49	2.19	0.389	0.225	0.265	3.21	0.719	0.401	0.318
KYU7	21.749426	100.276329	China	19	12	1.46	2.26	0.242	0.0330	0.529	9.76	0.947	0.751	0.611
KYU8	27.092175	100.060959	China	23	1	1.53	1.53	0.533	0.461	-1.00	1.00	0.00	1.00	0.00
KYU9	25.109817	101.416367	China	21	1	1.40	1.40	0.400	0.440	-1.00	1.00	0.00	1.00	0.00
KZH1	29.007999	119.084999	China	24	1	1.33	1.33	0.333	0.267	-1.00	1.00	0.00	1.00	0.00
KZH10	28.34935	121.065783	China	23	13	1.56	2.39	0.472	0.200	8.20×10^{-2}	5.34	0.850	0.411	0.546
KZH2	30.187133	120.084922	China	24	7	1.47	1.92	0.438	0.167	1.50×10^{-2}	2.38	0.605	0.340	0.261
KZH3	29.084358	119.194472	China	23	2	1.24	1.35	0.333	0.527	-0.778	1.09	8.70×10^{-2}	0.545	4.55×10^{-2}
KZH4	28.870035	120.969188	China	25	1	1.20	1.20	0.200	0.540	-1.00	1.00	0.00	1.00	0.00
KZH5	28.825021	120.765898	China	23	1	1.53	1.53	0.533	0.439	-1.00	1.00	0.00	1.00	0.00
KZH6	29.277767	120.902767	China	23	1	1.47	1.47	0.467	0.100	-1.00	1.00	0.00	1.00	0.00
KZH7	29.728117	120.222533	China	24	9	1.54	2.33	0.439	0.267	0.102	6.26	0.877	0.696	0.348
KZH8	29.7172	119.64935	China	23	9	1.52	2.27	0.475	0.233	2.70×10^{-2}	3.70	0.763	0.411	0.364
KZH9	31.249683	119.889233	China	24	1	1.13	1.13	0.133	0.505	-1.00	1.00	0.00	1.00	0.00
MD1	38.940334	-77.121368	United States	23	4	1.42	1.87	0.450	0.485	-0.218	2.37	0.605	0.593	0.136
MD2	38.992832	-76.991669	United States	22	3	1.51	2.04	0.533	0.300	-0.272	1.20	0.177	0.401	9.52×10^{-2}
MO1	37.234722	-89.521111	United States	18	2	1.32	1.47	0.333	0.067	-0.367	1.12	0.111	0.559	5.88×10^{-2}
MO2	37.318611	-89.527222	United States	9	1	1.47	1.47	0.467	0.369	-1.00	1.00	0.00	1.00	0.00
MS2	33.730831	-90.563904	United States	22	1	1.27	1.27	0.267	0.242	-1.00	1.00	0.00	1.00	0.00
MS3	32.341217	-89.136818	United States	23	2	1.39	1.59	0.533	0.233	-0.807	1.40	0.300	0.702	4.55×10^{-2}
MS4	31.967251	-89.859482	United States	13	2	1.48	1.76	0.567	0.133	-0.640	1.17	0.154	0.583	8.33×10^{-2}
MS5	31.352716	-89.342384	United States	8	1	1.73	1.73	0.733	0.292	-1.00	1.00	0.00	1.00	0.00
MS6	32.386185	-88.621902	United States	25	1	1.47	1.47	0.467	0.342	-1.00	1.00	0.00	1.00	0.00
NC1	34.345032	-78.708649	United States	21	1	1.13	1.13	0.133	0.367	-1.00	1.00	0.00	1.00	0.00

NC13	36.341019	-78.419914	United States	24	11	1.53	2.37	0.510	0.233	1.90×10^{-2}	4.72	0.822	0.429	0.435
NC15	35.966133	-78.847336	United States	26	2	1.17	1.24	0.233	0.0670	-0.833	1.08	7.70×10^{-2}	0.540	4.00×10^{-2}
NC17	35.426731	-82.528015	United States	18	2	1.40	1.62	0.567	0.509	-0.881	1.12	0.111	0.559	5.88×10^{-2}
NC19	34.205597	-77.922783	United States	17	1	1.53	1.53	0.533	0.125	-1.00	1.00	0.00	1.00	0.00
NC20	36.39465	-79.698265	United States	33	14	1.52	2.28	0.505	0.300	1.00×10^{-3}	5.47	0.843	0.391	0.406
NC21	34.212399	-77.835297	United States	10	1	1.20	1.20	0.200	0.267	-1.00	1.00	0.00	1.00	0.00
NC3	35.375084	-83.221382	United States	19	8	1.56	2.67	0.317	0.499	0.340	2.80	0.678	0.311	0.389
NC4	35.619469	-82.173416	United States	16	7	1.47	2.13	0.343	0.100	0.168	5.57	0.875	0.795	0.400
NC5	35.719601	-81.153816	United States	21	14	1.52	2.66	0.520	0.521	-6.60×10^{-2}	5.73	0.867	0.409	0.650
NC6	35.650284	-80.469467	United States	22	2	1.60	2.09	0.467	0.439	-4.60×10^{-2}	1.10	9.10×10^{-2}	0.548	4.76×10^{-2}
NC7	36.062199	-79.856621	United States	22	4	1.41	1.80	0.500	0.499	-0.429	2.07	0.541	0.517	0.143
NE1	40.595171	-95.787825	United States	8	1	1.33	1.33	0.333	0.450	-1.00	1.00	0.00	1.00	0.00
NY1	40.987316	-73.805351	United States	23	1	1.40	1.40	0.400	0.350	-1.00	1.00	0.00	1.00	0.00
NY4	40.906551	-73.584579	United States	15	3	1.36	1.67	0.511	0.167	-0.661	1.51	0.362	0.503	0.143
OK1	36.306821	-95.595824	United States	24	1	1.40	1.40	0.400	0.200	-1.00	1.00	0.00	1.00	0.00
S	30.653351	104.078979	China	4	3	1.54	1.95	0.422	0.200	2.90×10^{-2}	2.67	0.833	0.889	0.667
SC1	33.933933	-80.97023	United States	18	1	1.40	1.40	0.400	0.422	-1.00	1.00	0.00	1.00	0.00
SC4	35.078365	-81.717468	United States	22	6	1.50	2.04	0.502	0.446	-0.109	2.88	0.684	0.480	0.238
SC6	34.855518	-82.26638	United States	11	7	1.54	2.66	0.479	0.200	2.50×10^{-2}	5.26	0.891	0.658	0.600
SC8	34.255383	-79.699432	United States	24	1	1.27	1.27	0.267	0.452	-1.00	1.00	0.00	1.00	0.00
TN13	34.98735	-84.370117	United States	24	8	1.43	1.77	0.533	0.504	-0.328	3.10	0.707	0.387	0.304
TN14	35.527134	-84.348236	United States	24	1	1.60	1.60	0.600	0.133	-1.00	1.00	0.00	1.00	0.00
TN16	36.395649	-89.339897	United States	25	1	1.47	1.47	0.467	0.399	-1.00	1.00	0.00	1.00	0.00
TN2	35.026348	-85.312897	United States	7	1	1.53	1.53	0.533	0.300	-1.00	1.00	0.00	1.00	0.00
TN3	36.161068	-86.794365	United States	14	1	1.53	1.53	0.533	0.233	-1.00	1.00	0.00	1.00	0.00
TN6	35.095901	-90.076668	United States	17	2	1.53	2.01	0.567	0.267	-0.430	1.13	0.118	0.562	6.25×10^{-2}
TN8	36.249367	-84.18145	United States	25	1	1.67	1.67	0.667	0.267	-1.00	1.00	0.00	1.00	0.00
TN9	35.051533	-84.900452	United States	17	4	1.26	1.46	0.417	0.400	-0.800	1.45	0.331	0.363	0.188

TX2	30.704468	-94.934166	United States	18	1	1.53	1.53	0.533	0.333	-1.00	1.00	0.00	1.00	0.00
TX3	32.583782	-94.342934	United States	24	1	1.53	1.53	0.533	0.231	-1.00	1.00	0.00	1.00	0.00
TX4	30.924683	-94.007835	United States	20	2	1.46	1.61	0.433	0.267	-0.278	1.84	0.479	0.917	5.26×10^{-2}
TX5	30.049217	-94.881516	United States	25	4	1.45	2.03	0.433	0.267	-9.10×10^{-2}	3.31	0.727	0.827	0.125
TX6	31.420532	-94.723633	United States	20	1	1.27	1.27	0.267	0.342	-1.00	1.00	0.00	1.00	0.00
VA1	37.38335	-77.414047	United States	23	4	1.56	2.14	0.633	0.392	-0.295	2.37	0.605	0.593	0.136
VA10	37.147148	-79.23037	United States	25	1	1.27	1.27	0.267	0.133	-1.00	1.00	0.00	1.00	0.00
VA11	36.73859	-76.574051	United States	24	1	1.40	1.40	0.400	0.492	-1.00	1.00	0.00	1.00	0.00
VA4	36.990284	-80.790535	United States	18	5	1.50	2.06	0.573	0.133	-0.256	2.08	0.549	0.415	0.235
VA8	37.198532	-81.392899	United States	14	1	1.27	1.20	0.267	0.200	-1.00	1.00	0.00	1.00	0.00
VA9	37.168633	-81.898033	United States	9	1	1.53	1.53	0.533	0.450	-1.00	1.00	0.00	1.00	0.00
WV1	38.023884	-81.354553	United States	17	2	1.60	2.13	0.433	0.133	3.60×10^{-3}	1.99	0.529	0.997	6.25×10^{-2}
X	21.919767	101.277855	China	3	2	1.24	1.36	0.156	0.267	0.103	3.00	1.00	1.00	0.500

Table 3.3. Assignment of haplotypes to populations, separated by country. Numbers in parentheses indicate the number of individuals in that population that had that particular haplotype.

Haplotype	Country			
	US	China	Korea	Japan
1	AL10 (7), AL11 (7), AL12 (10), AL13 (16), AL14 (9), AL3 (12), AL5 (21), AL7 (11), AL9 (15), AR2 (13), FL2 (17), FL3 (4), GA10 (13), GA14 (4), GA17 (1), GA18 (1), GA2 (4), GA3 (17), GA32 (9), GA33 (6), GA34 (13), GA35 (11), GA36 (2), GA37 (12), GA4 (5), GA7 (10), GA9 (5), GA96 (9), KY10 (5), KY11 (19), KY7 (10), KY8 (22), KY9 (6), MD1 (3), MD2 (9), MO1 (12), MO2 (8), MS2 (13), MS3 (18), MS6 (7), NC1 (12), NC13 (2), NC17 (21), NC19 (1), NC20 (6), NC21 (4), NC3 (6), NC4 (6), NC5 (5), NC6 (3), NC7 (10), NE1 (1), NY4 (17), OK1 (4), SC6 (7), SC8 (14), TN13 (16), TN14 (15), TN16 (4), TN2 (3), TN3 (11), TN6 (1), TN8 (8), TN9 (5), TX2 (14), TX3 (20), TX5 (21), VA1 (12), VA4 (8), VA8 (10), VA9 (10), WV1 (6)	H (5), KAN1 (13), KAN3 (12), KAN5 (14), KAN8 (12), KAN9 (18), KFU1 (8), KFU7 (11), KGZH1 (1), KHN3 (1), KHU2 (8), KJI1 (10), KJI2 (18), KJI5 (3), KLN1 (2), KLN2 (3), KSD1 (3), KSD3 (2), KZH10 (12), KZH6 (1)	KKO10 (4), KKO14 (4), KKO17 (1), KKO2 (1), KKO21 (3), KKO3 (11), KKO5 (2), KKO6 (14), KKO7 (2), KKO8 (2), KKO9 (1)	KJP1 (7), KJP10 (19), KJP12 (4), KJP16 (12), KJP17 (18), KJP19 (9), KJP2 (13), KJP20 (16), KJP21 (1), KJP22 (7), KJP23 (1), KJP24 (5), KJP25 (10), KJP26 (15), KJP27 (20), KJP28 (19), KJP29 (11), KJP5 (14), KJP6 (13), KJP7 (18), KJP8 (7)
2	AR1 (4), FL1 (14), GA37 (1), GA96 (6), MS5 (11), NC13 (6), NC15 (13), SC1 (11), SC4 (2), TX4 (17), VA4 (4), WV1 (7)	C (1), H (6), KAN1 (1), KAN2 (13), KAN4 (13), KAN6 (8), KCQ1 (11), KCQ2 (1), KFU1 (6), KFU2 (10), KFU3 (14), KFU4 (13), KFU5 (14), KFU6 (17), KFU7 (4), KFU8 (12), KGD1 (2), KGD2 (4), KGD3 (12), KGD4 (3),	KKO12 (1), KKO19 (16)	KJP14 (17), KJP18 (19), KJP19 (5), KJP21 (19), KJP22 (11), KJP23 (13), KJP24 (17),

		KGD5 (2), KGU1 (9), KGU5 (2), KHAI1 (5), KHK1 (13), KHN10 (12), KHN11 (1), KHN12 (2), KHN3 (1), KHN4 (12), KHN5 (7), KHN6 (4), KHN7 (3), KHN8 (3), KHN9 (8), KSA1 (15), KSA2 (2), KSA3 (2), KSA4 (1), KSC1 (1), KSC2 (15), KSC3 (6), KSC4 (12), KSD1 (1), KSD3 (1), KSH1 (11), KTW1 (9), KTW4 (15), KTW5 (12), KYU1 (1), KYU10 (8), KYU2 (5), KYU4 (4), KYU7 (4), KYU9 (7), KZH1 (4), KZH10 (4), KZH2 (5), KZH3 (1), KZH4 (3), KZH5 (1), KZH7 (16), KZH8 (5), KZH9 (2), S (2)		KJP25 (4), KJP4 (14), KJP9 (20)
3	AL6 (17), GA39 (8), GA96 (1), NC20 (11), NC3 (1), NC4 (7), NC7 (3), NY1 (14), TN6 (17), VA11 (5)			KJP25 (5)
4	GA36 (1), GA6 (2), NC5 (7)	KHU2 (4), KHU3 (15)		KJP11 (22), KJP13 (11), KJP3 (22)
5	GA2 (4), GA37 (3), MD1 (12), MD2 (1)	KB1 (2), KHN3 (4), KHU2 (1)	KKO13 (5), KKO18 (6), KKO20 (7), KKO22 (10)	
6	AL12 (3), GA6 (3), MS4 (12), SC4 (13)		KKO15 (14)	
7	NC19 (10), TX6 (12), VA10 (11)			KJP8 (4)
8		KGU3 (16), KGU4 (3), KGU5 (8)		
9	GA17 (8), KY9 (1), NC13 (7)	KSA1 (3), KSA2 (2), KSA3 (2)		
10		KJIX2 (22)		
11	GA2 (2), GA36 (1), NC5 (2), SC6 (1)		KKO11 (2)	KJP15 (12)
12	GA2 (2), GA36 (1)	KJI3 (14)		

13		KJI4 (15)		
14		KCQ1 (2), KFU3 (2), KFU7 (1), KGD3 (1), KTW1 (2), KYU4 (2), KYU7 (2), KZH9 (1)		
15		KGU5 (11)		
16		KFU2 (3), KGD4 (7)		
17		KJIX1 (9)		
18		KHN7 (4), KYU1 (2), KYU2 (1)		
19				KJP8 (7)
20			C (5)	
21		KHE1 (3)	KKO11 (1)	KJP15 (1)
22		KHN2 (4)		
23		KSC4 (4)		
24			KKO1 (1), KKO16 (2)	
25		KSC1 (3)		
26		KZH8 (3)		
27	SC6 (2)			

Table 3.4. Description of mutational differences among haplotypes, using Haplotype 1 as a reference. N/A indicates not applicable. Position indicates the base pair position in the sequence. For example, TT for AA at position 599 means a TT substitution for AA at base pair position 599 in the sequence alignment.

Haplotype	**Description**
1	N/A
2	TT for AA at position 599
3	G for T at position 396; G for T at position 598
4	T for C at position 21
5	7 base pair deletion at position 453
6	C for A at position 149; TT for AA at position 599
7	G for T at position 187
8	TT for AA at position 599; T insertion at position 794

9	TT for AA at position 599; single base pair deletion at position 793
10	T for C at position 193; A for G at position 405
11	T for G at position 382; TT for AA at position 599; T insertion at position 794
12	C for T at position 567; TC for AA at position 569; CCCT insertion at position 572
13	T insertion at position 794
14	TT for AA at position 599; C for A at position 634
15	TT for AA at position 599; TT insertion at position 794
16	A for T at position 184; C for A at position 569; TT for AA at position 599
17	C for A at position 149
18	C for T at position 196; TT for AA at position 599
19	T for C at position 93; TT for AA at position 599
20	T for G at position 428; TT for AA at position 599
21	T for G at position 382;TT for AA at position 599; C for A at position 634; T insertion at position 794
22	G for T at position 309; TT for AA at position 599
23	T for C at position 696
24	C for T at position 41; TT for AA at position 599
25	AAATTACATG insertion at position 67; T for C at position 87; GAAATACTTATTAGTGTAT insertion at position 100; C for A at position 145; T for A at position 149; TTATGAAAAGAGGAATAAATAAAATAATA insertion at position 151; C for T at position 196; T for G at position 238; A for T at position 241; T for A at position 265; C for A at position 310; A for G at position 325; A for T at position 381; G insertion at position 383; G for A at position 407; single base pair deletion at position 427; A for G at position 437; single base pair deletion at position 459; 2 base pair deletion at position 468; G for T at position 478; G for A at position 502; C for A at position 536; G for A at position 556; TT for AA at position 599; A for G at position 624; single base pair deletion at position 627; G for C at position 638; G for T at position 646; A for G at position 666; T for G at position 697; C for T at position 713; C for T at position 731; C for A at position 811; G for A at position 818
26	G for A at position 634
27	G for A at position 634; C for A at position 672

Table 3.5. Clonal lineages shared across populations, including a list of populations for each clonal lineage and the number of affected samples in that population in the parentheses.

Shared Clonal Lineage	**Country**			
	US	**China**	**Korea**	**Japan**
1	AL10 (7), AL14 (19), FL2 (22), GA14 (18), GA17 (1), GA35 (25), GA39 (24), KY8 (2), KY9 (24), KY10 (6), KY11 (19), MO1 (18), MS2 (22), MS3 (23), NC3 (2), NC4 (4), NC20 (5), NC21 (10), NE1 (8), SC8 (24), TN8 (25), TX2 (18), TX5 (25), VA4 (14), VA8 (14)	KAN9 (1), KB1 (23), KHU2 (5), KSA2 (8), KSD1 (22), KSH1 (14)		
2	AL12 (7), TX4 (13)	KAN1 (7)		
3	KY11 (1), NC17 (17)	KJP28 (3)		
4	NC1 (21)	KJP6 (17), KKO18 (24)		
5		C (1), KGU4 (1)		
6		C (4), KFU2 (13), KFU3 (16), KFU7 (20), KGD3 (13), KGD4 (15), KGU1 (27), KGU3 (23), KGU4 (1), KGU5 (24), KHN7 (23), KTW1 (23), KTW4 (23), KTW5 (24), KYU1 (2), KYU2 (12), KYU3 (20), KYU4 (4), X (2)		
7		H (3), KZH2 (3)		
8		KAN3 (24), KAN8 (22), KHN3 (1)		
9		KAN4 (25), KZH3 (22)		
10		KAN5 (24), KKO9 (24), KKO20 (1)		
11		KGD1 (24), KGD2 (2), KGD5 (13), KHAI1 (12), KHK1 (26)		
12		KGD2 (18), KZH1 (24)		

13		KHN5 (7), KHN8 (24), KSC4 (3), KYU7 (4)		
14		KYU8 (23), KYU9 (21), KYU10 (23)		
15	TX6 (20), VA10 (25)			
16	AL7 (22), FL3 (17)			
17	GA32 (24), NC3 (11), NC4 (2)			
18	AL3 (9), AL9 (24), AL11 (23), GA4 (12), GA9 (23), GA33 (23), NC6 (1), NC20 (2), TN2 (7), TN3 (14), TN9 (17), TN13 (8), TN14 (24), VA1 (12), VA9 (9), WV1 (9)			

Figures

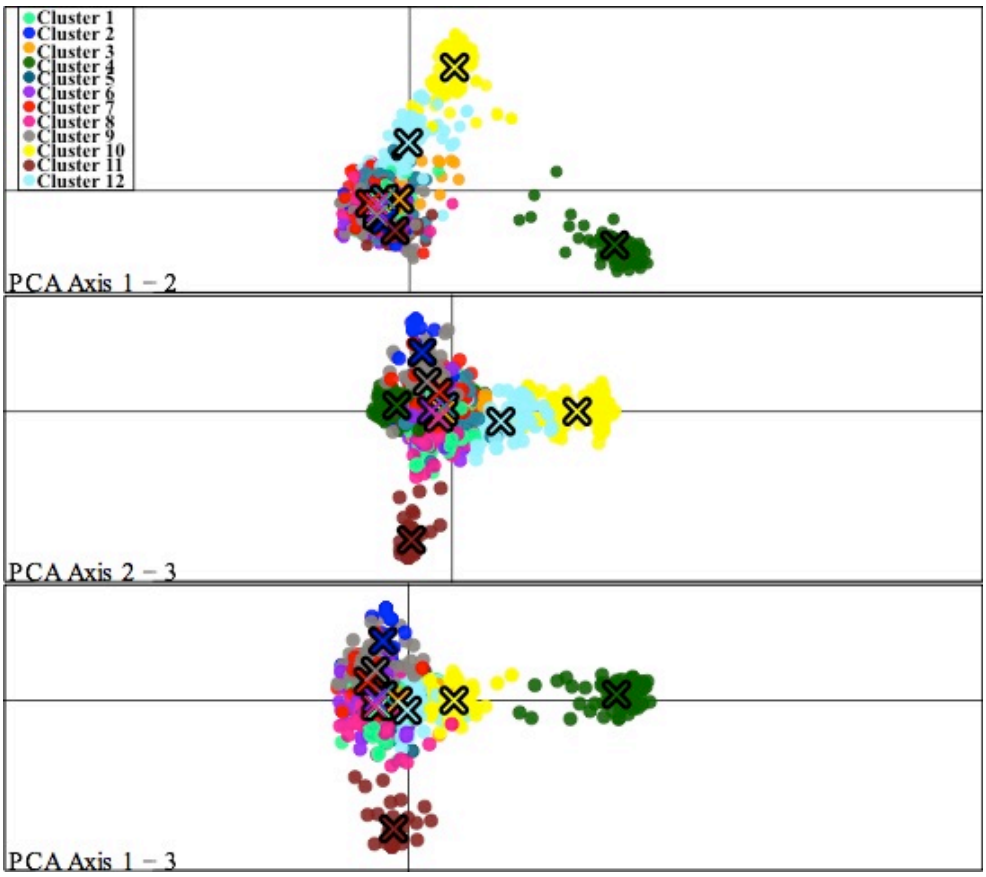


Figure 3.1. Distribution of genetic clusters based on Discriminant Analysis of Principal Components for all invasive and native populations based on a K of 12.

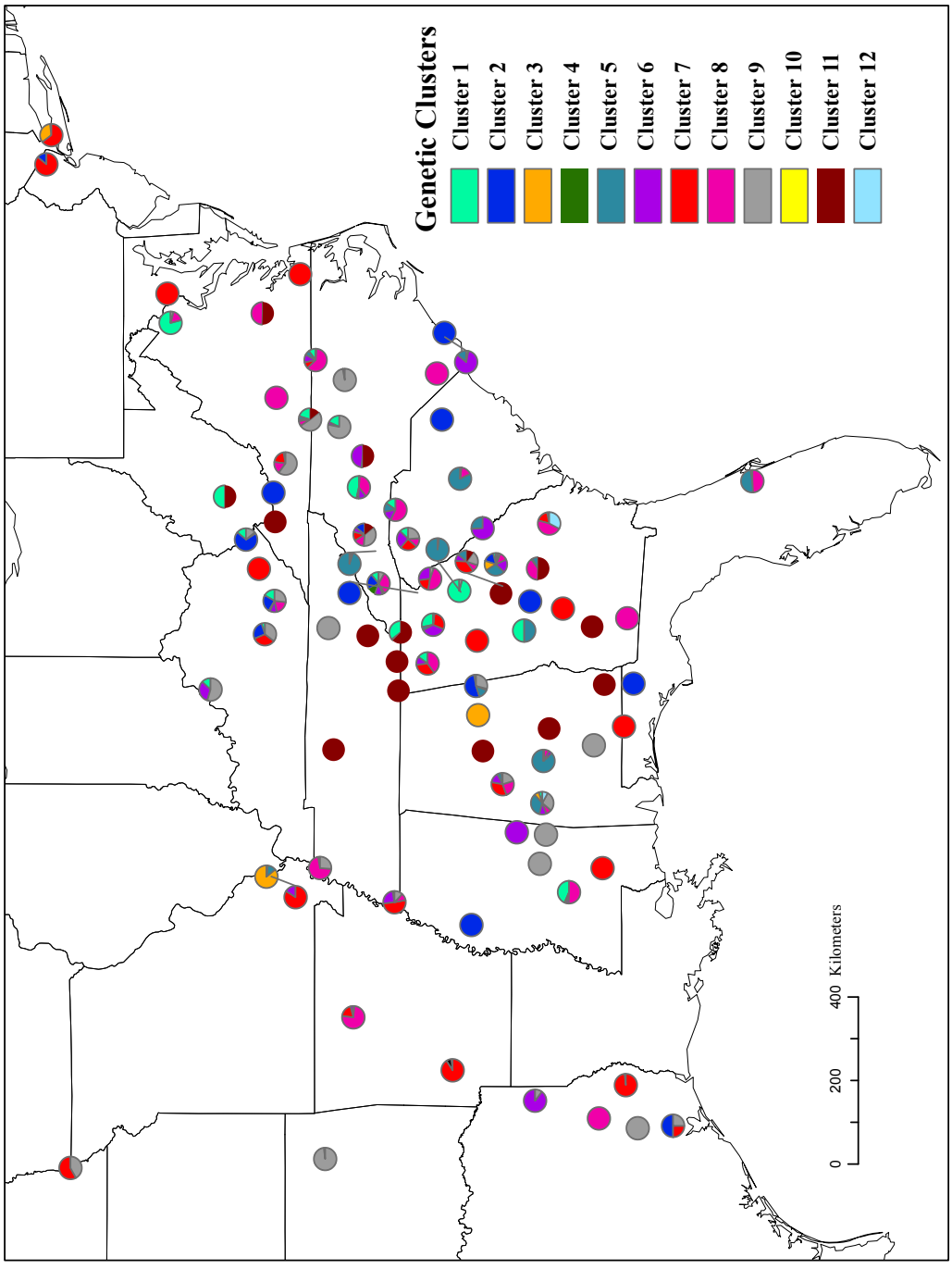


Figure 3.2. A structure map of the geographic distribution of genetic clusters across the invasive range from Discriminant Analysis of Principal Components with a K of 12.

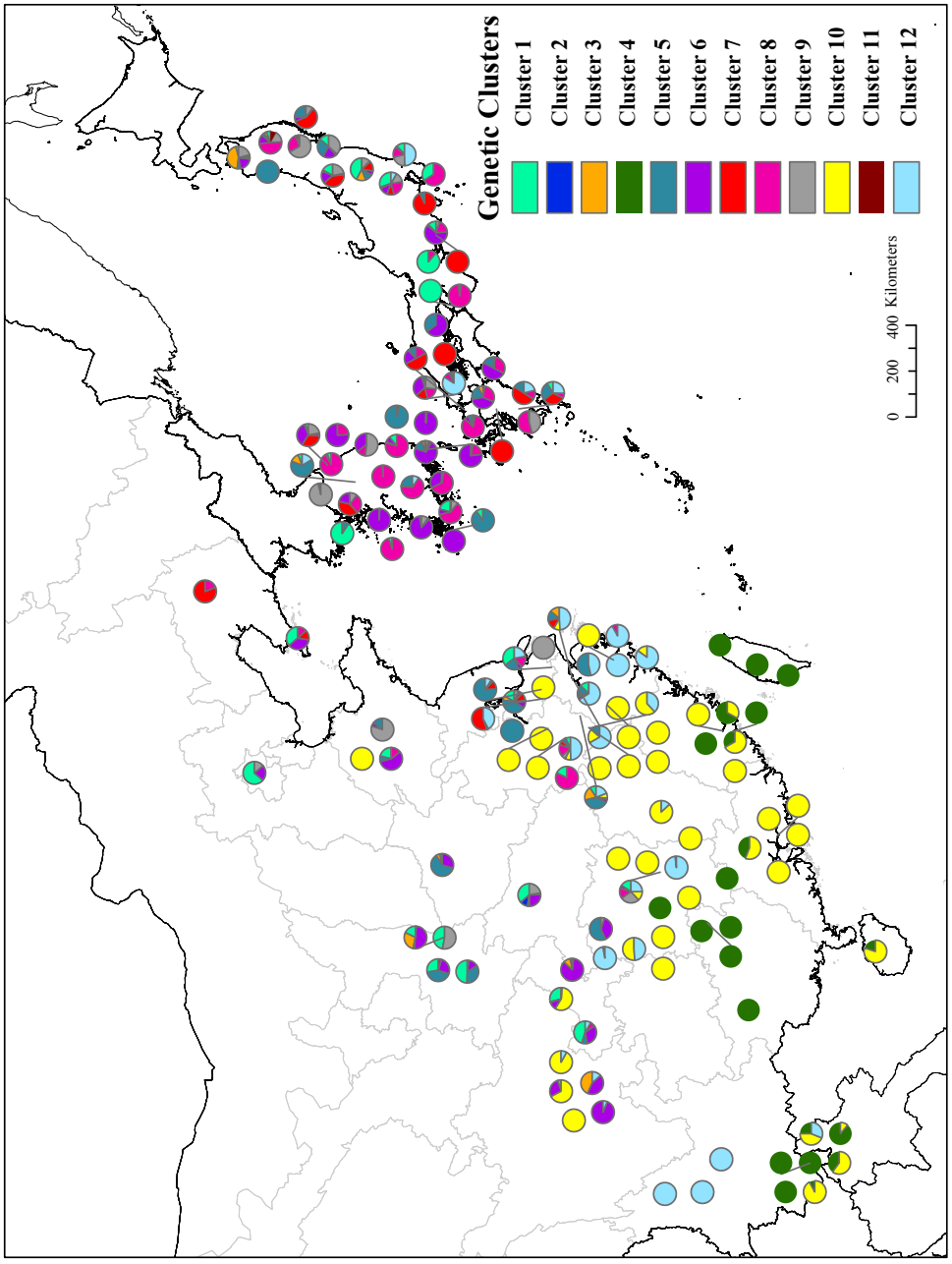


Figure 3.3. A structure map of the geographic distribution of genetic clusters across the native range from Discriminant Analysis of Principal Components with a K of 12.

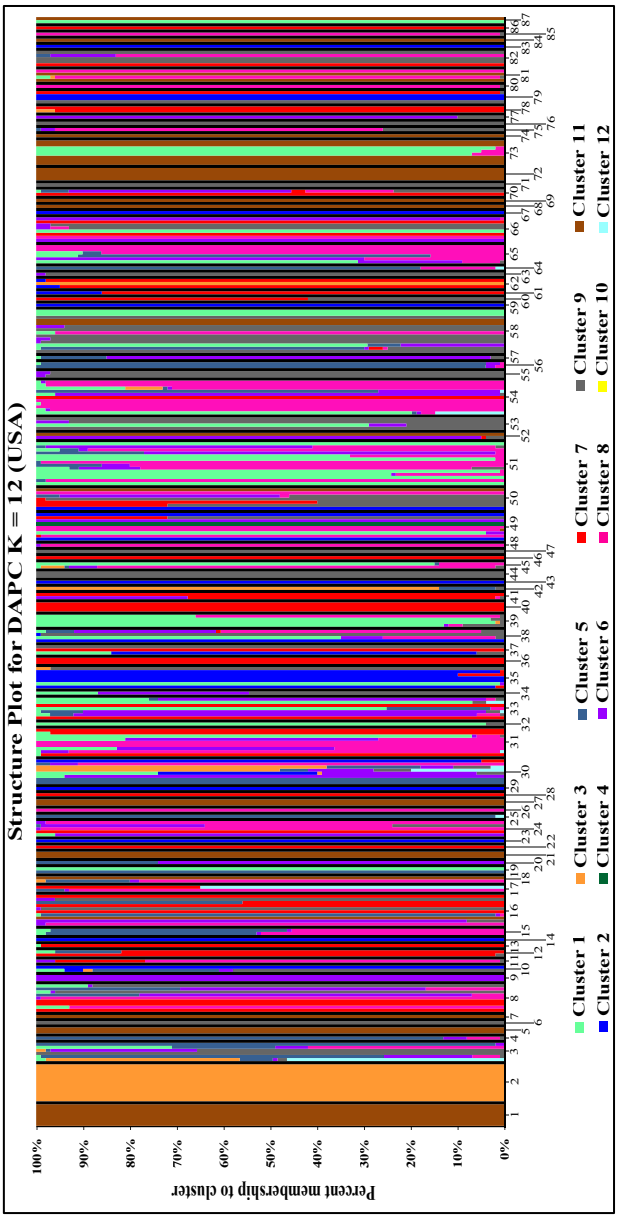


Figure 3.4. Structure plot of within-individual distribution of genetic clusters from a DAPC analysis with K of 12 in the United States. Legend for population numbers is found after Figure 3.7.

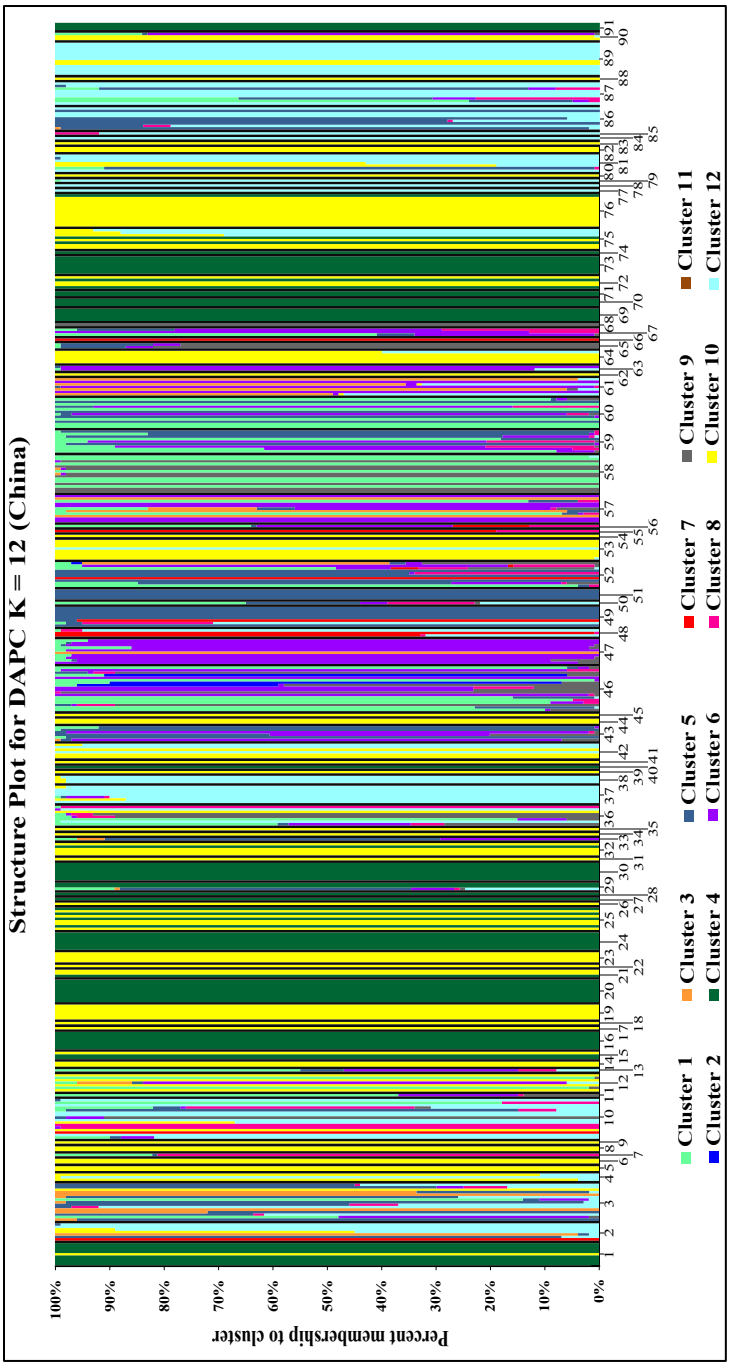


Figure 3.5. Structure plot of within-individual distribution of genetic clusters from a DAPC analysis with K of 12 in China. Legend for population numbers is found after Figure 3.7.

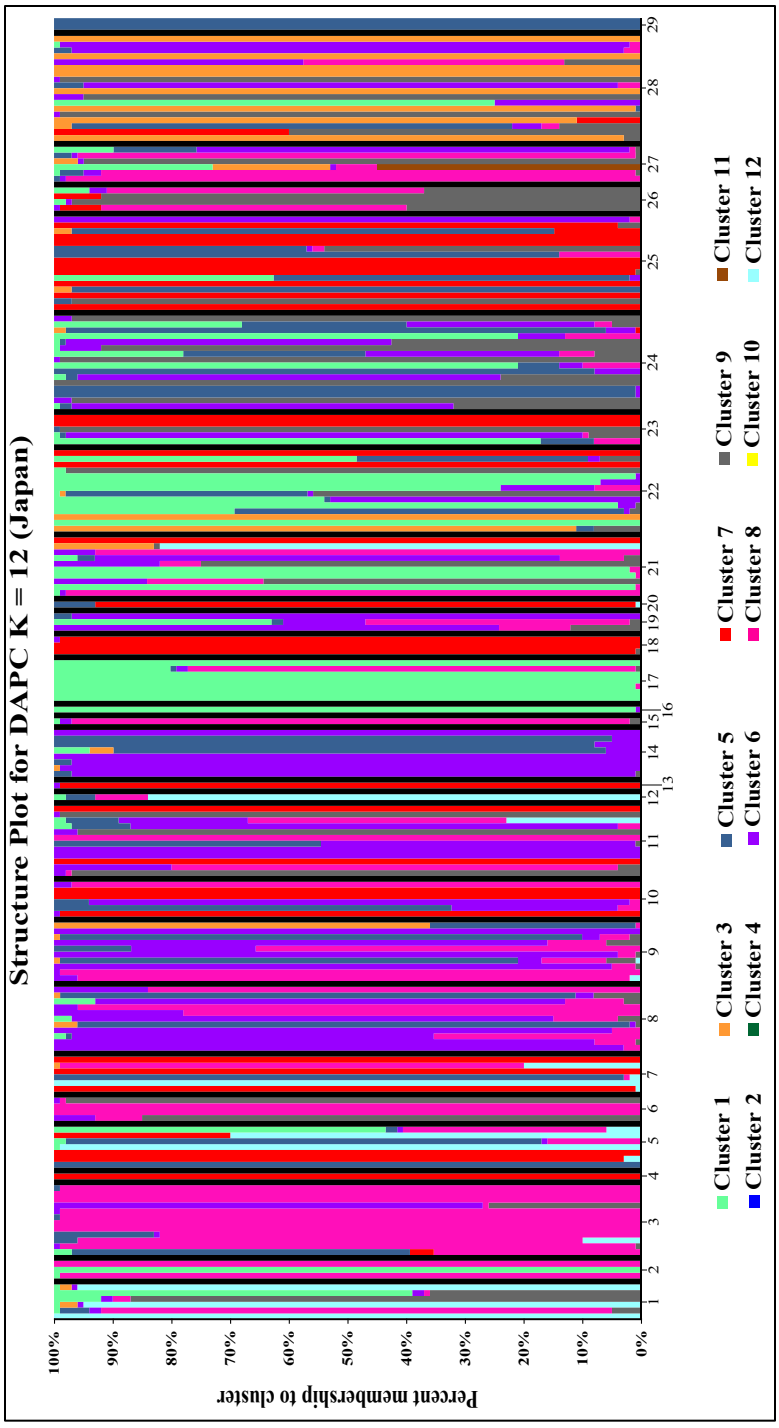


Figure 3.6. Structure plot of within-individual distribution of genetic clusters from a DAPC analysis with K of 12 in Japan. Legend for population numbers is found after Figure 3.7.

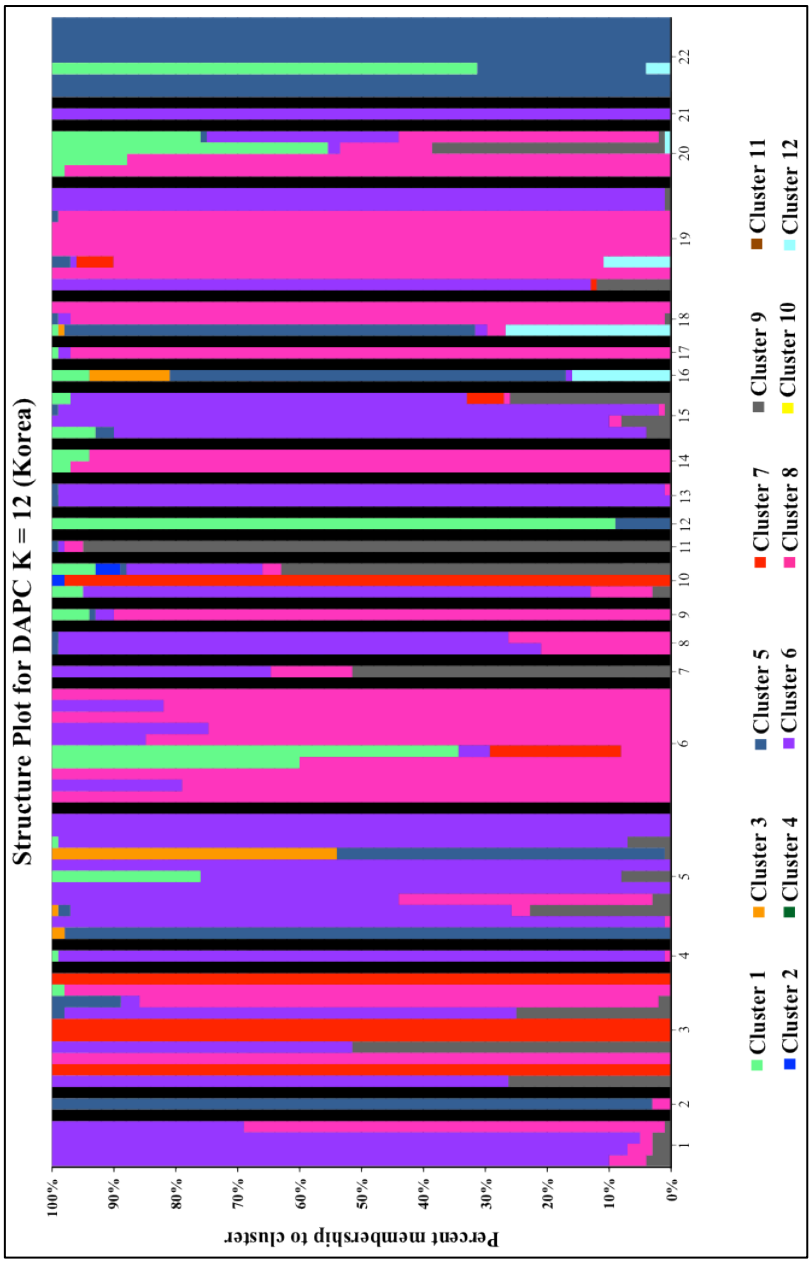


Figure 3.7. Structure plot of within-individual distribution of genetic clusters from a DAPC analysis with K of 12 in South Korea.

X-axis number	Population Code			
	China	Japan	Korea	US
1	C	KJP1	KKO1	AL3
2	H	KJP2	KKO2	AL5
3	KAN1	KJP3	KKO3	AL6
4	KAN2	KJP4	KKO4	AL7
5	KAN3	KJP5	KKO5	AL9
6	KAN4	KJP6	KKO6	AL10
7	KAN5	KJP7	KKO7	AL11
8	KAN6	KJP8	KKO8	AL12
9	KAN8	KJP9	KKO9	AL13
10	KAN9	KJP10	KKO10	AL14
11	KB1	KJP11	KKO11	AR1
12	KCQ1	KJP12	KKO12	AR2
13	KCQ2	KJP13	KKO13	FL1
14	KFU1	KJP14	KKO14	FL2
15	KFU2	KJP15	KKO15	FL3
16	KFU3	KJP16	KKO16	GA2
17	KFU4	KJP17	KKO17	GA3
18	KFU5	KJP18	KKO18	GA4
19	KFU6	KJP19	KKO19	GA6
20	KFU7	KJP20	KKO20	GA7
21	KFU8	KJP21	KKO21	GA9
22	KGD1	KJP22	KKO22	GA10
23	KGD2	KJP23		GA14
24	KGD3	KJP24		GA17
25	KGD4	KJP25		GA18
26	KGD5	KJP26		GA32
27	KGU1	KJP27		GA33
28	KGU3	KJP28		GA34
29	KGU4	KJP29		GA35
30	KGU5			GA36
31	KGZH1			GA37
32	KHAI1			GA39
33	KHE1			GA96
34	KHK1			KY7
35	KHN2			KY8
36	KHN3			KY9
37	KHN4			KY10
38	KHN5			KY11
39	KHN6			MD1
40	KHN7			MD2
41	KHN8			MO1
42	KHN9			MO2
43	KHN10			MS2
44	KHN11			MS3
45	KHN12			MS4
46	KHU2			MS5
47	KHU3			MS6
48	KJI1			NC1
49	KJI2			NC3
50	KJI3			NC4
51	KJI4			NC5
52	KJI5			NC6
53	KJIX1			NC7
54	KJIX2			NC13
55	KLN1			NC15
56	KLN2			NC17
57	KSA1			NC19
58	KSA2			NC20
59	KSA3			NC21
60	KSA4			NE1
61	KSC1			NY1

62	KSC2			NY4
63	KSC3			OK1
64	KSC4			SC1
65	KSD1			SC4
66	KSD2			SC6
67	KSD3			SC8
68	KSH1			TN2
69	KTW1			TN3
70	KTW4			TN6
71	KTW5			TN8
72	KYU1			TN9
73	KYU2			TN13
74	KYU3			TN14
75	KYU4			TN16
76	KYU7			TX2
77	KYU8			TX3
78	KYU9			TX4
79	KYU10			TX5
80	KZH1			TX6
81	KZH2			VA1
82	KZH3			VA4
83	KZH4			VA8
84	KZH5			VA9
85	KZH6			VA10
86	KZH7			VA11
87	KZH8			WV1
88	KZH9			
89	KZH10			
90	S			
91	X			

Figure 3.8 DAPC Legend. Population numbers in DAPC plots and corresponding population name for each DAPC plot, Figures 3.4, 3.5, 3.6 and 3.7 as well as Figures 3.18, 3.19, 3.20 and 3.21.

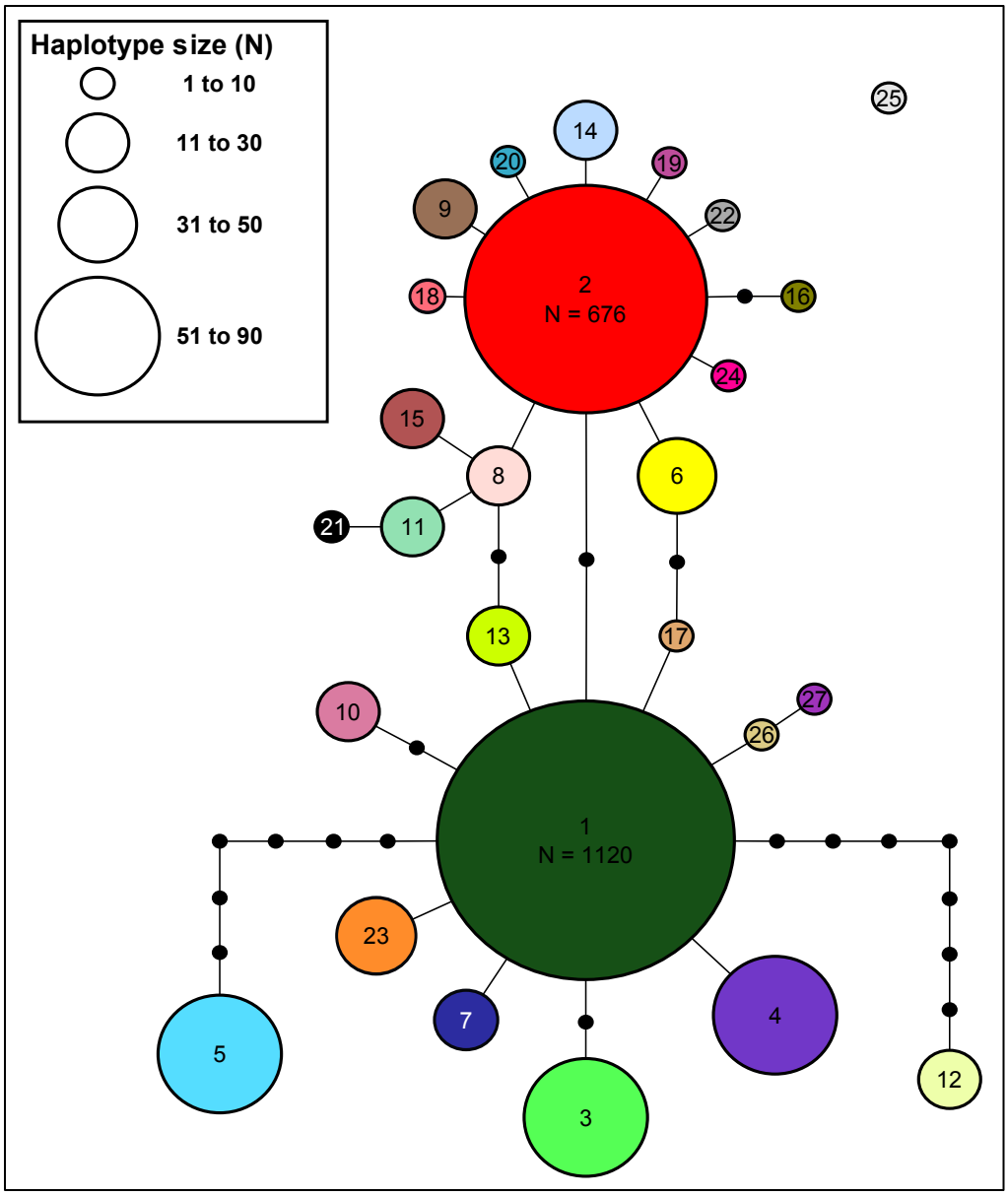


Figure 3.9. Haplotype network of 830 bp chloroplast region, rpoB-trnC for all 2316 sequenced invasive and native individuals.

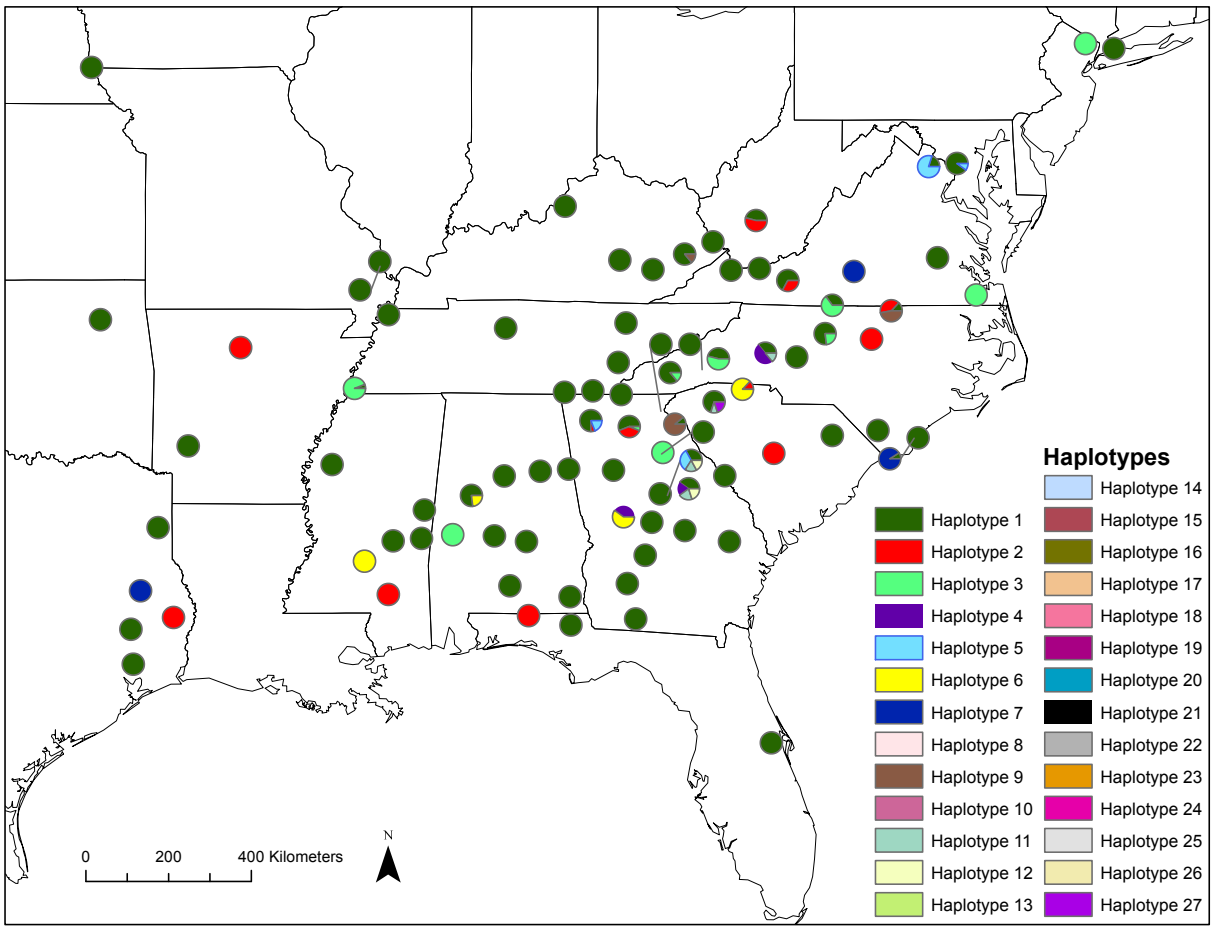


Figure 3.10. Haplotype map of the United States depicting the distribution of haplotypes across the invasive range based on results from the haplotype network.

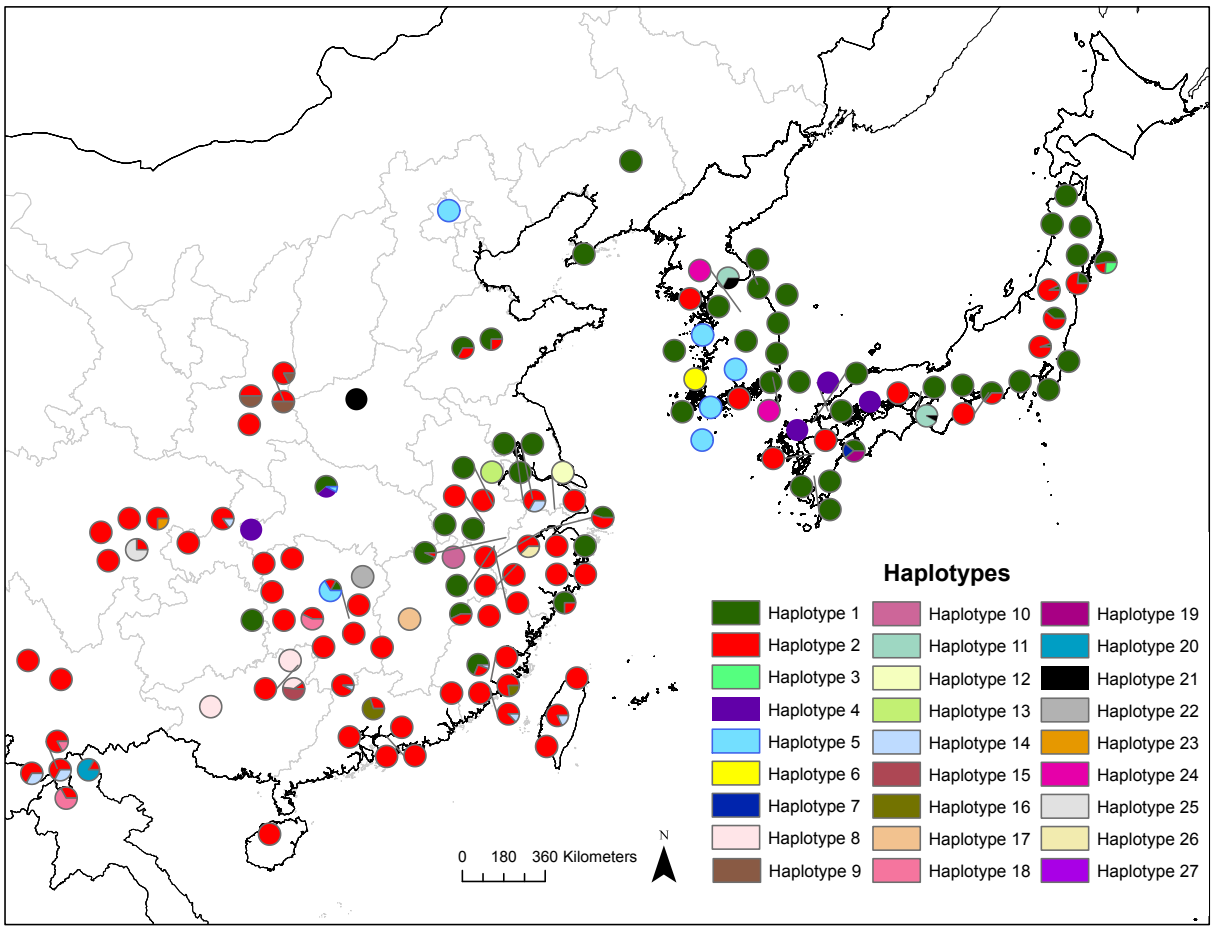


Figure 3.11. Haplotype map of Asia depicting the distribution of haplotypes across the invasive range based on results from the haplotype network.

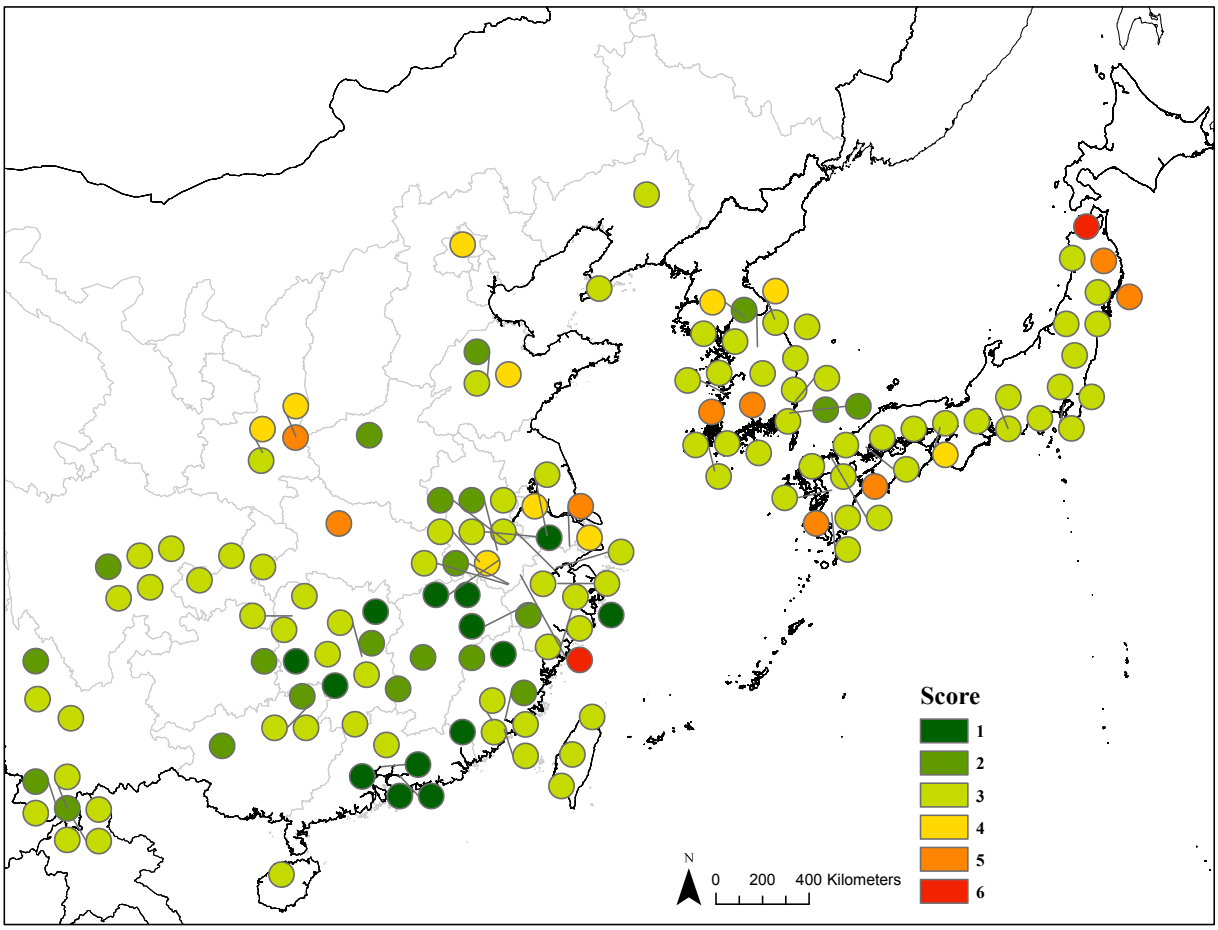


Figure 3.12. Heat map of native populations depicting the amount of evidence linking a given population to the United States based on DAPC K = 12 and K = 18 results, haplotype assignment and the analysis of shared clonal lineages

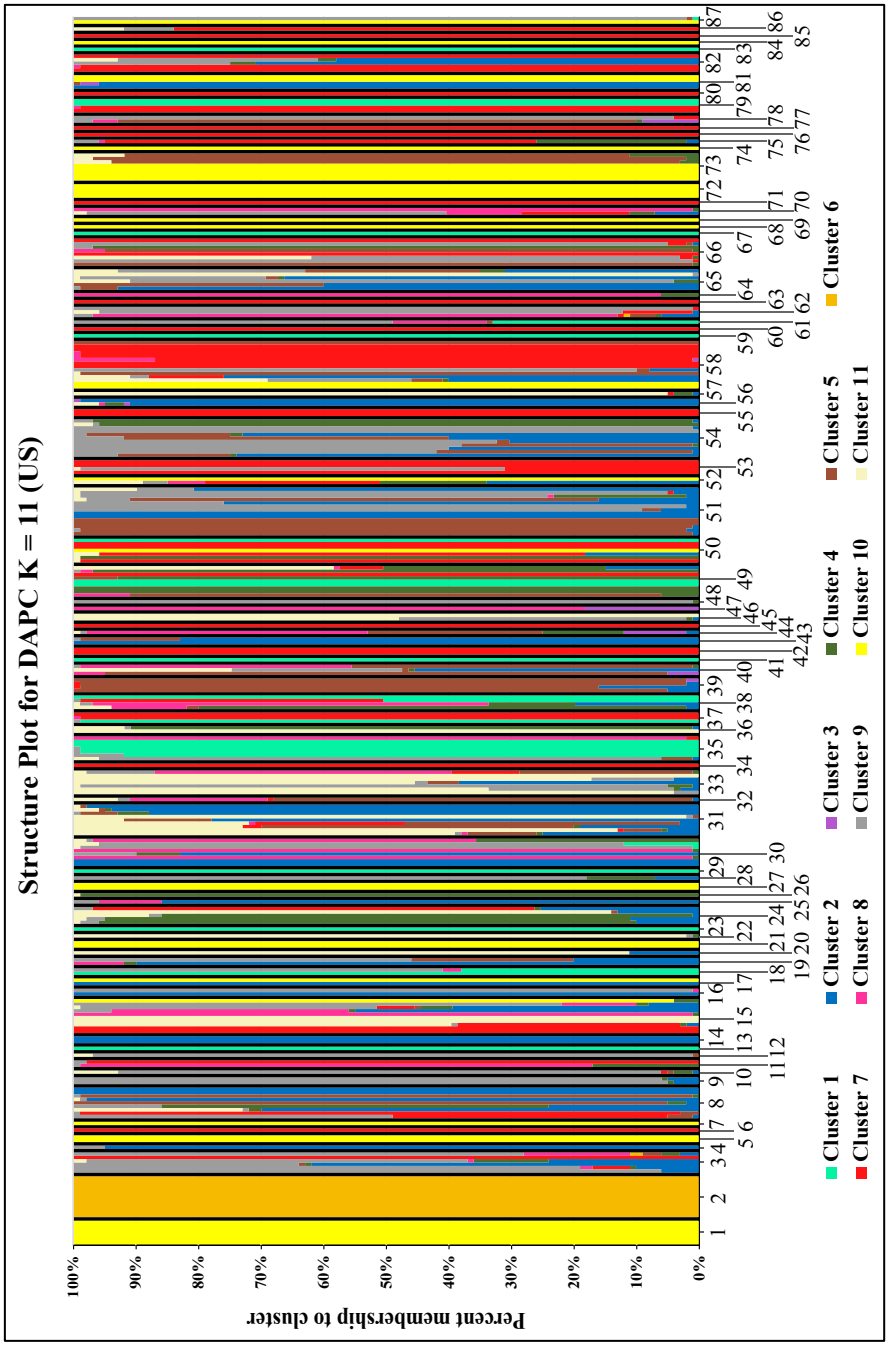


Figure 3.13. DAPC structure plot of individual assignment of invasive individuals to genetic clusters with putative source populations. Legend for population numbers is found after Figure 3.14.

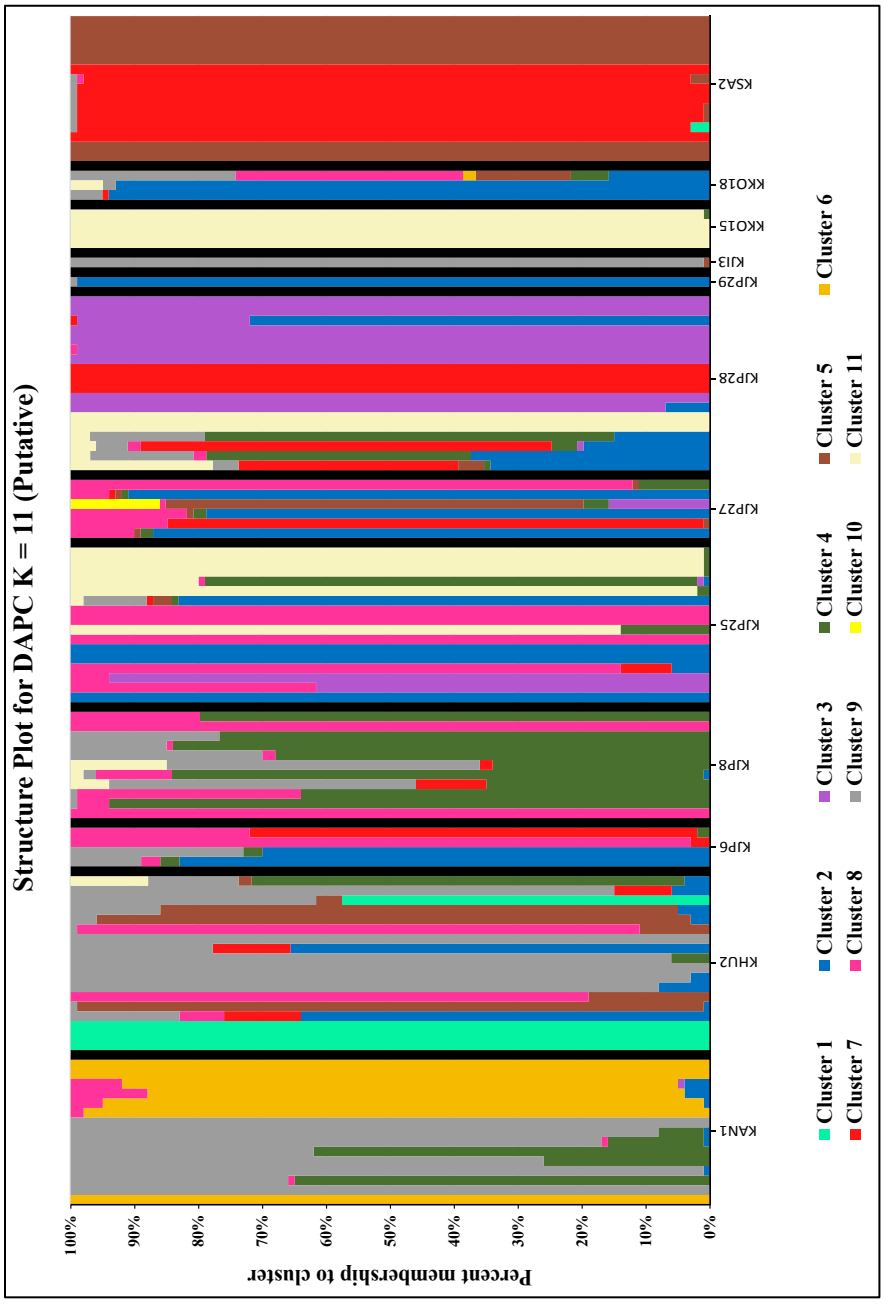


Figure 3.14. DAPC structure plot of individual assignment of individuals in putative source populations to genetic clusters with the invasive range.

X-axis Number	Population Code	
	Putative Sources	USA
1	KAN1	AL3
2	KHU2	AL5
3	KJP6	AL6
4	KJP8	AL7
5	KJP25	AL9
6	KJP27	AL10
7	KJP28	AL11
8	KJP29	AL12
9	KJI3	AL13
10	KKO15	AR1
11	KKO18	AR2
12	KSA2	FL1
13		FL2
14		FL3
15		GA2
16		GA3
17		GA4
18		AL14
19		GA6
20		GA7
21		GA9
22		GA10
23		GA14
24		GA17
25		GA18
26		GA32
27		GA33
28		GA34
29		GA35
30		GA36
31		GA37
32		GA39
33		GA96
34		KY7
35		KY8
36		KY9
37		KY10
38		KY11
39		MD1
40		MD2
41		MS2
42		MS3
43		MS4
44		MS5
45		MS6
46		MO1
47		MO2
48		NC1
49		NC3
50		NC4
51		NC5
52		NC6
53		NC7
54		NC13
55		NC15
56		NC17
57		NC19
58		NC20
59		NC21
60		NE1
61		NY1

62		NY4
63		OK1
64		SC1
65		SC4
66		SC6
67		SC8
68		TN2
69		TN3
70		TN6
71		TN8
72		TN9
73		TN13
74		TN14
75		TN16
76		TX2
77		TX3
78		TX4
79		TX5
80		TX6
81		VA1
82		VA4
83		VA8
84		VA9
85		VA10
86		VA11
87		WV1

Figure 3.15 Legend. Population numbers in DAPC plots and corresponding population name for each DAPC plot, Figures 3.13 and 3.14.

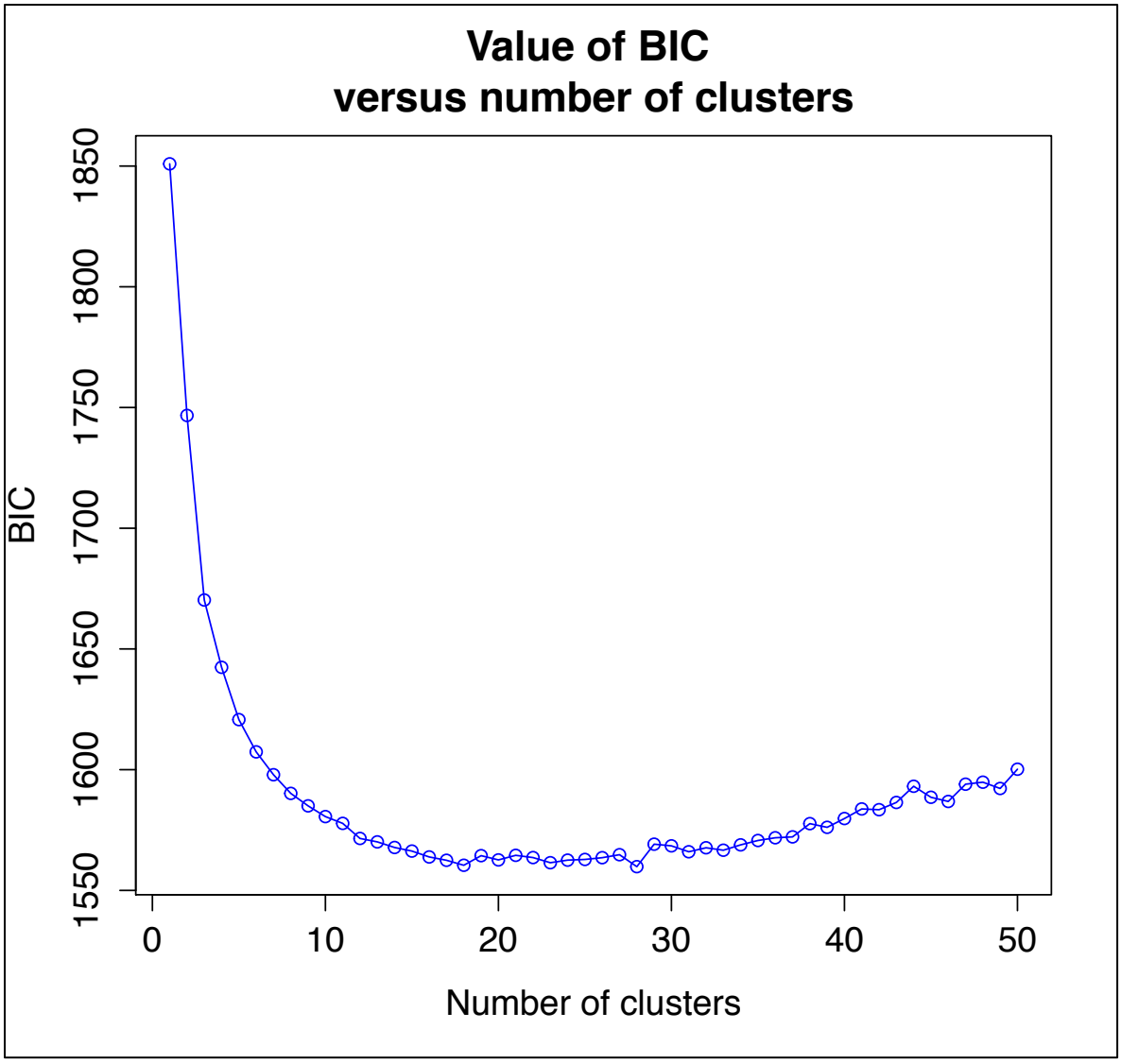


Figure 3.16. Regression of the Bayesian Information Criteria against the number of clusters (K) to identify the best K for Discriminant Analysis of Principal Components for the entire dataset.

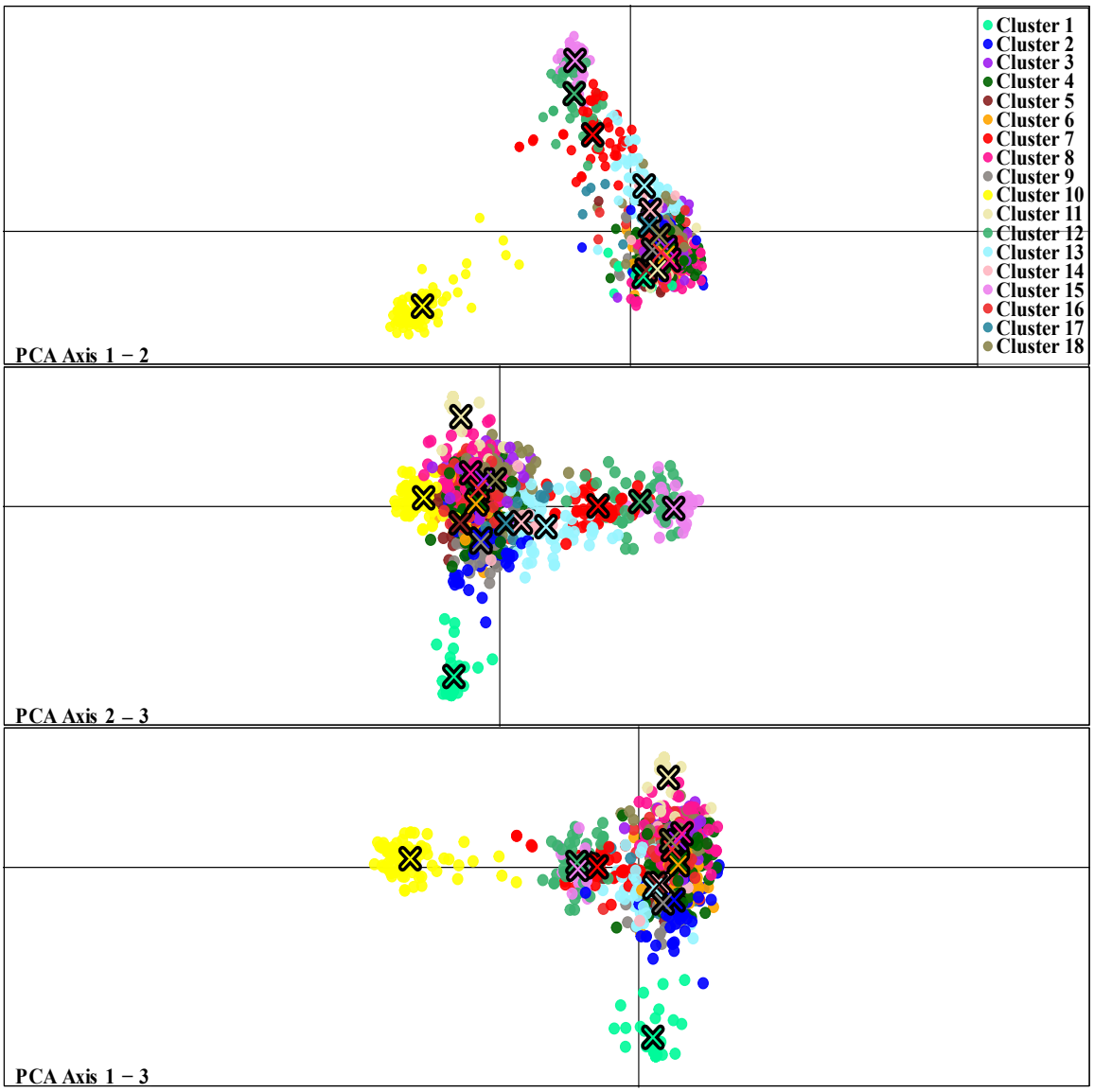


Figure 3.17. DAPC Plot with K = 18 for the entire dataset.

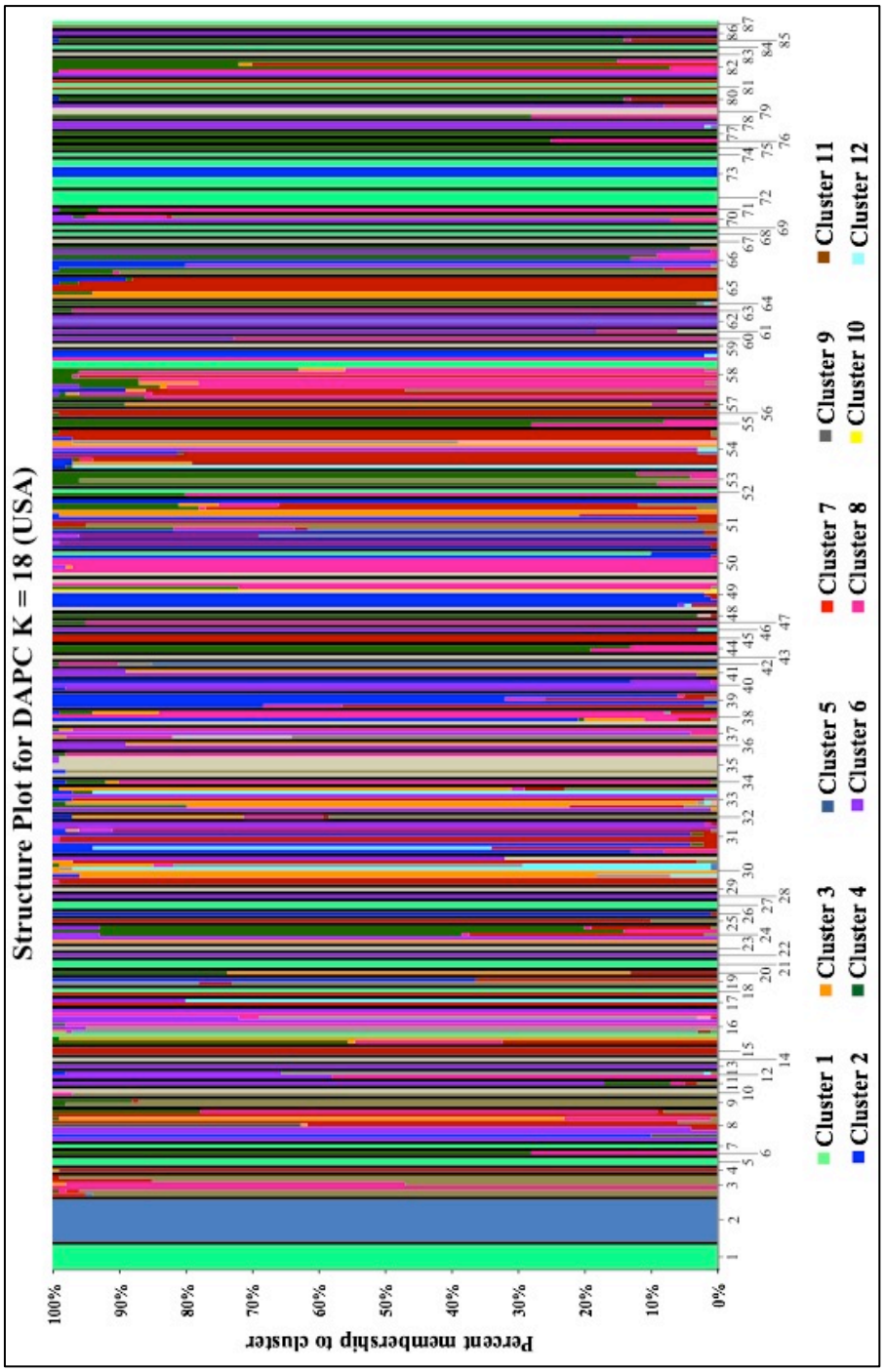


Figure 3.18. Structure plot for invasive individuals from the DAPC Plot for K = 18. Legend for population numbers found in Figure 3.8 DAPC Legend.

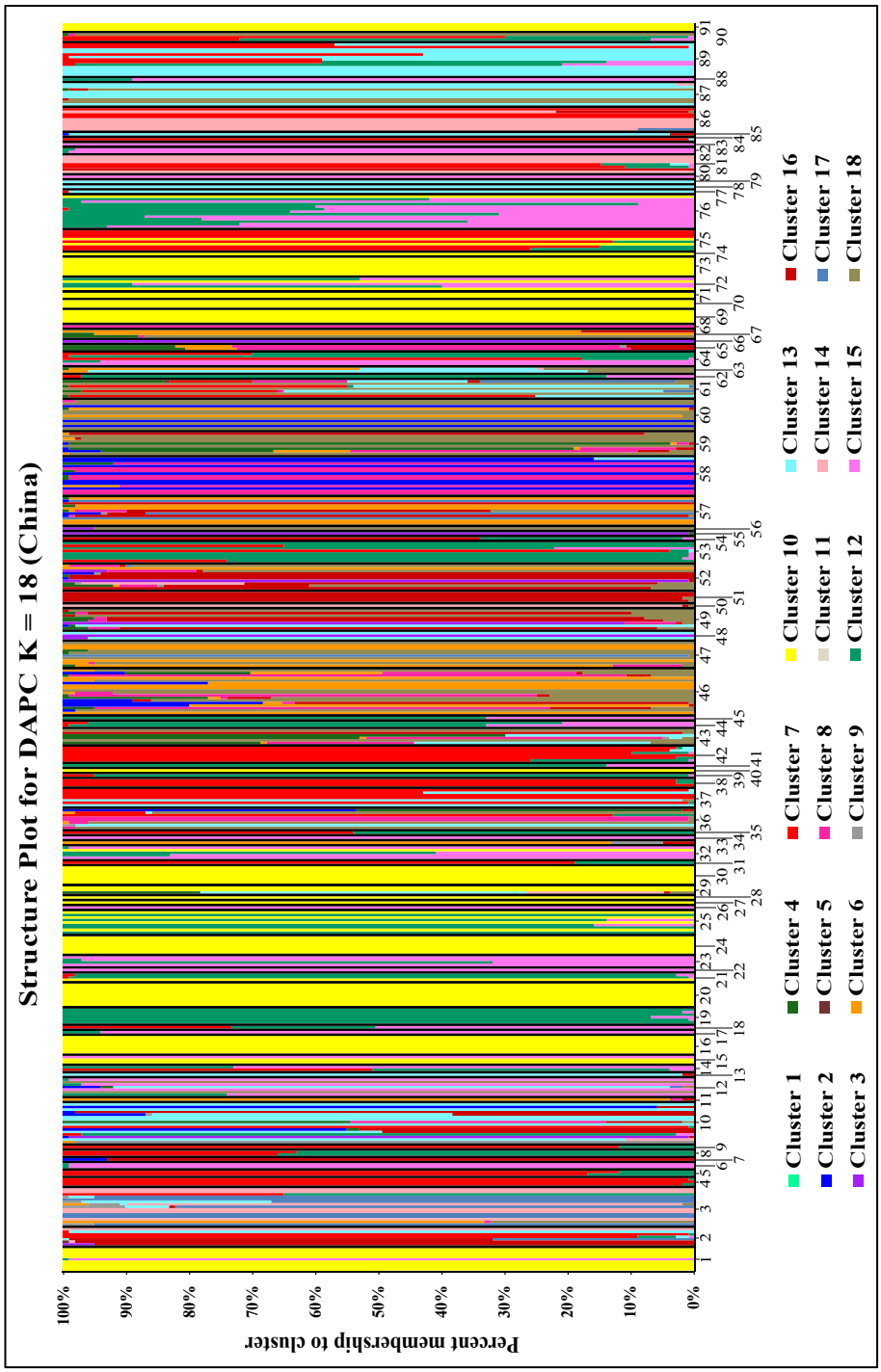


Figure 3.19. Structure plot for individuals from China from the DAPC Plot for K = 18. Legend for population numbers found in Figure 3.8 DAPC Legend.

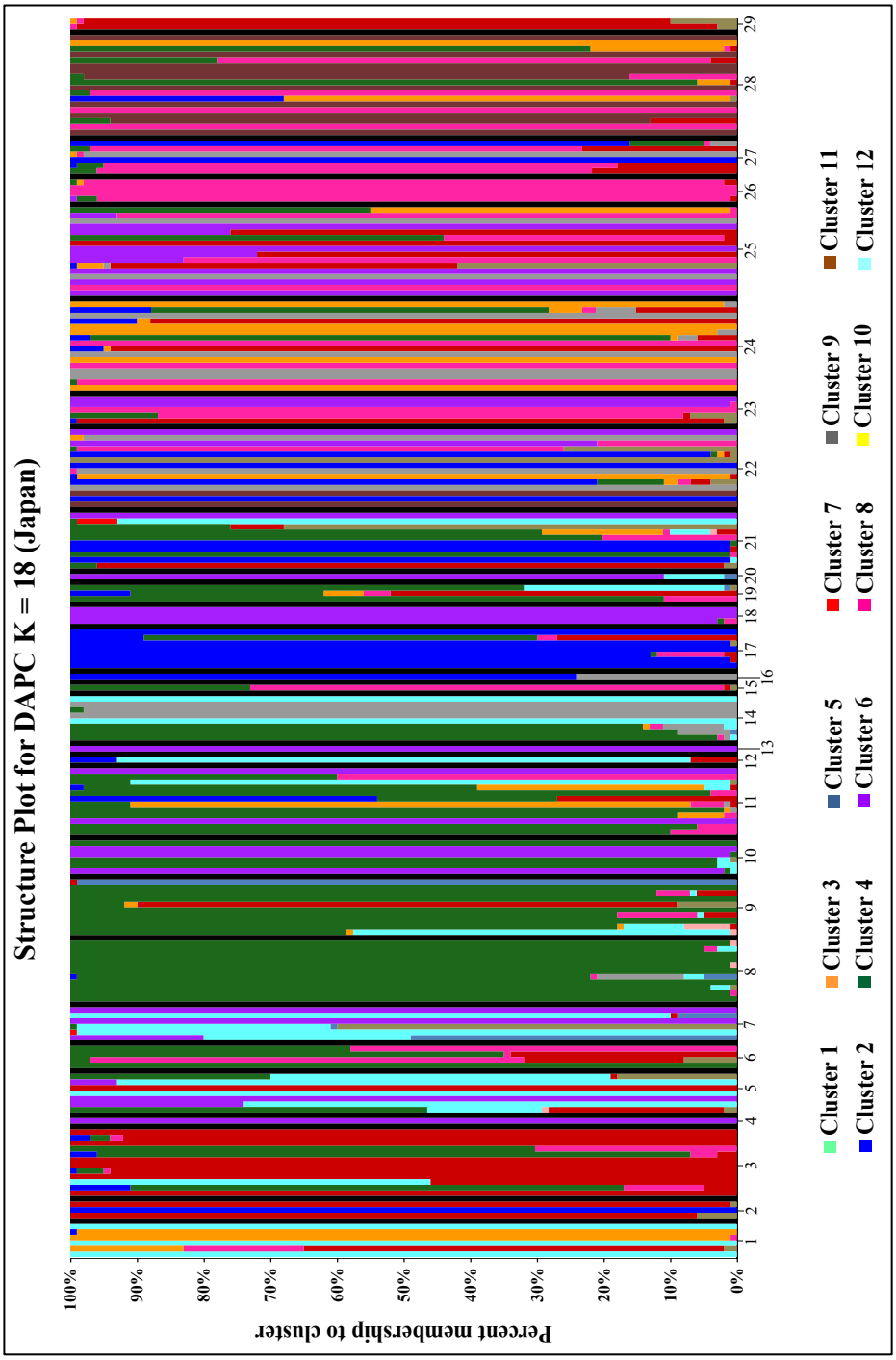


Figure 3.20. Structure plot for individuals from Japan from the DAPC Plot for K = 18. Legend for population numbers found in Figure 3.8 DAPC Legend.

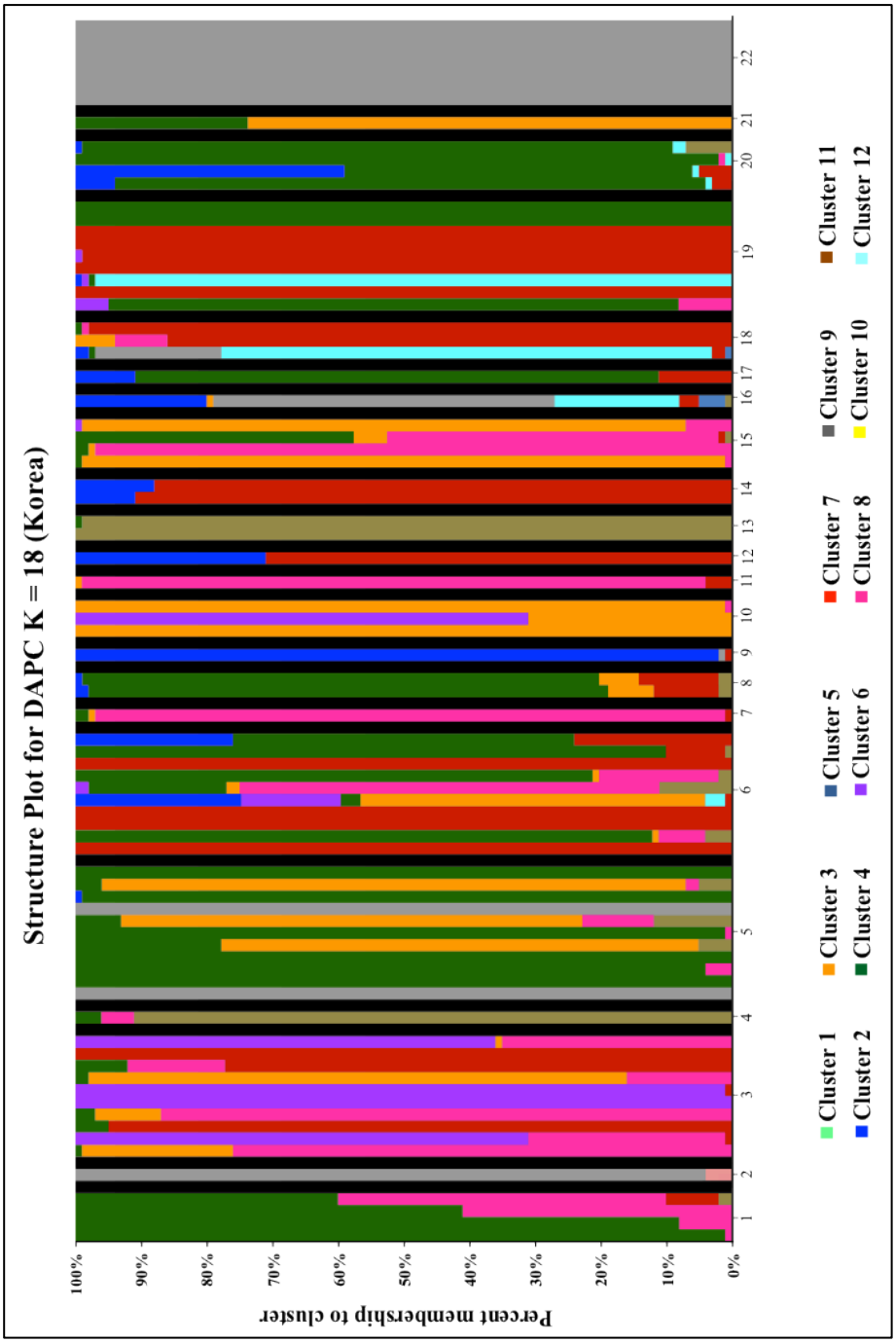


Figure 3.21. Structure plot for individuals from South Korea from the DAPC Plot for K = 18. Legend for population numbers found in Figure 3.8 DAPC Legend.

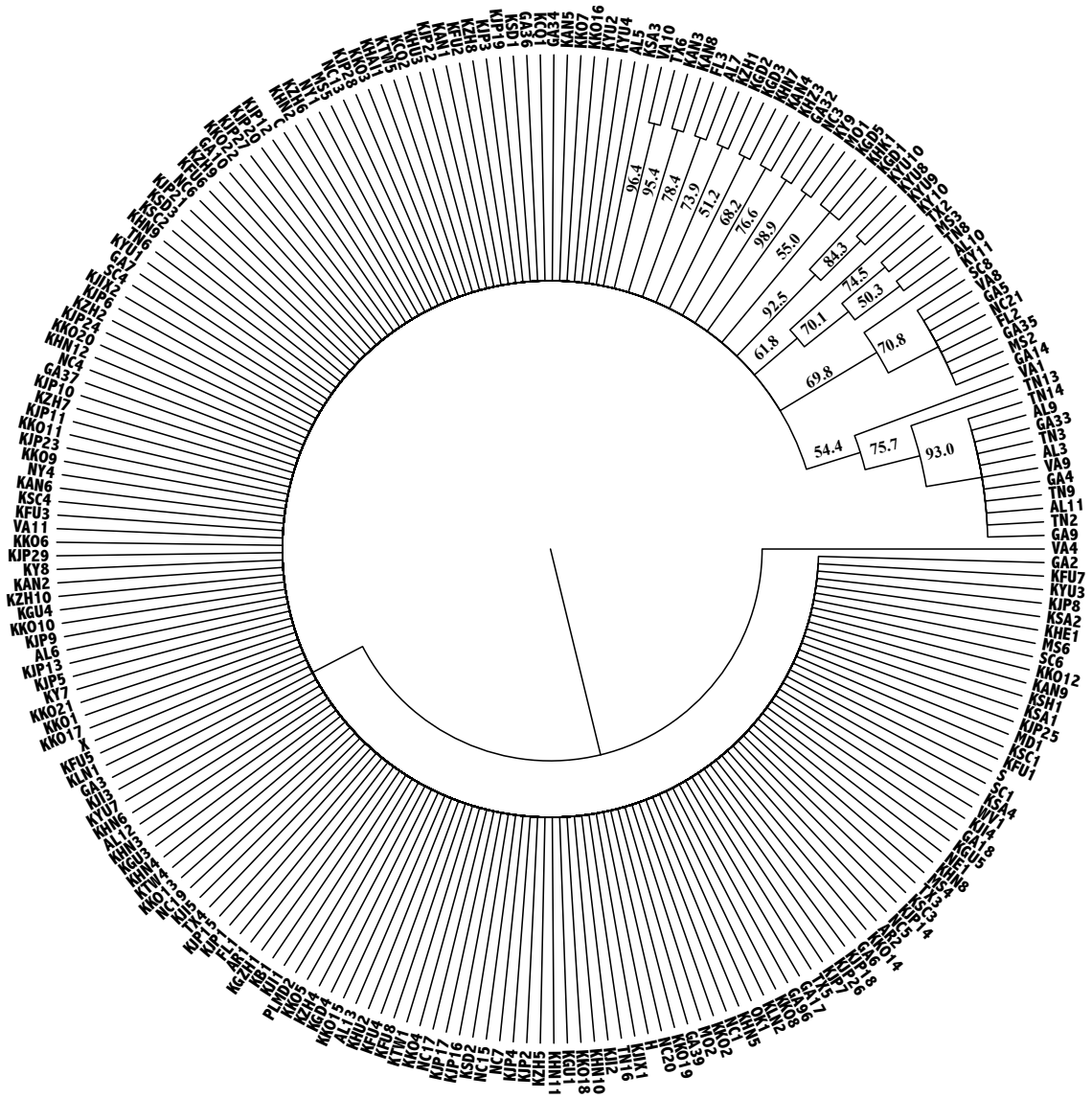


Figure 3.22. Neighbor-joining tree of microsatellite genetic distance.

References

Abramovitz, J. N. (1983) Pueraria lobata Willd. (Ohwi), kudzu: limitations to sexual reproduction. M.S. thesis. University of Maryland, College Park, Maryland, USA.

Agapow, P. Ä. and A. Burt. (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1:101-102.

Allendorf, F. W. (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo biology* 5:181-190.

Allendorf, F. W. and L. L. Lundquist. (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17:24-30.

Arnaud-Haond, S., C. M. Duarte, F. Alberto, and E. A. Serrão. (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology* 16:5115-5139.

Arnaud-Haond, S., C. M. Duarte, E. Diaz-Almela, N. Marbà, T. Sintes, and E. A. Serrão. (2012) Implications of extreme life span in clonal organisms: millenary clones in meadows of the threatened seagrass Posidonia oceanica. *PloS one* 7:e30454-e30454.

Baker, H. G. and G. L. Stebbins. (1965) The genetics of colonizing species. The genetics of colonizing species.

Balloux, F., L. Lehmann, and T. de Meeûs. (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635-1644.

Barrett, R. D. H. and D. Schluter. (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23:38-44.

Barrett, S. C. (2011) Why reproductive systems matter for the invasion biology of plants. 1st edition. Blackwell Publishing Ltd, Hoboken.

Barrett, S. C. H., R. I. Colautti, and C. G. Eckert. (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* 17:373-383.

Blaustein, R. (2001) Kudzu's invasion into Southern United States life and culture. *The Great Reshuffling: Human dimensions of invasive species*:55-62.

Clement, M., D. Posada, and K. A. Crandall. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9:1657-1659.

Dlugosch, K. and I. Parker. (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17:431-449.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32:1792-1797.

Ellstrand, N. C. and K. A. Schierenbeck. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences of the United States of America* 97:7043.

Everest, J. W., J. H. Miller, D. M. Ball, and M. Patterson. (1999) Kudzu in Alabama history, uses, and control.

Facon, B., B. J. Genton, J. Shykoff, P. Jarne, A. Estoup, and P. David. (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution* 21:130-135.

Forseth, I. N. and A. F. Innis. (2004) Kudzu (Pueraria montana): history, physiology, and ecology combine to make a major ecosystem threat. *Critical reviews in plant sciences* 23:401-413.

Forseth, I. N. and A. H. Teramura. (1986) Kudzu leaf energy budget and calculated transpiration: the influence of leaflet orientation. *Ecology*:564-571.

Gordon, D. R. (1998) Effects of invasive, non-indigenous plant species on ecosystem processes: lessons from Florida. *Ecological Applications* 8:975-989.

Grünwald, N. J., S. B. Goodwin, M. G. Milgroom, and W. E. Fry. (2003) Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* 93:738-746.

Hauswaldt, J. S. and T. C. Glenn. (2003) Microsatellite DNA loci from the Diamondback terrapin (Malaclemys terrapin). *Molecular Ecology Notes* 3:174-176.

Hejda, M., P. Pyšek, and V. Jarošík. (2009) Impact of invasive plants on the species richness, diversity and composition of invaded communities. *Journal of Ecology* 97:393-403.

Hickman, J. E., S. Wu, L. J. Mickley, and M. T. Lerdau. (2010) Kudzu (Pueraria montana) invasion doubles emissions of nitric oxide and increases ozone pollution. *Proceedings of the National Academy of Sciences* 107:10115.

Hoffberg, S. L., K. Bentley, J. B. Lee, K. E. Myhre, K. Iwao, T. C. Glenn, and R. Mauricio. (2014) Characterization of 15 microsatellite loci in kudzu (Pueraria montana var. lobata) from the native and introduced ranges. *Conservation Genetics Resources*:1-3.

Hurlbert, S. H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577-586.

Jombart, T. (2013) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1.4-0.

Jombart, T., S. Devillard, and F. Balloux. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* 11:94.

Kalinowski, S. T. (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics* 5:539-543.

Kalinowski, S. T. (2005) HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5:187-189.

Kalisz, S., D. W. Vogler, and K. M. Hanley. (2004) Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature* 430:884-887.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.

Kolbe, J. J., R. E. Glor, L. R. Schettino, A. C. Lara, A. Larson, and J. B. Losos. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177-181.

Kolbe, J. J., A. Larson, J. B. Losos, and K. De Queiroz. (2008) Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biology Letters* 4:434.

Li, Z., Q. Dong, T. P. Albright, and Q. Guo. (2011) Natural and human dimensions of a quasi-wild species: the case of kudzu. *Biological Invasions* 13:2167-2179.

Librado, P. and J. Rozas. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Lindgren, C. J., K. L. Castro, H. A. Coiner, R. E. Nurse, and S. J. Darbyshire. (2013) The Biology of Invasive Alien Plants in Canada. 12. Pueraria montana var. lobata (Willd.) Sanjappa & Predeep. *Canadian journal of plant science* 93:71-95.

Lowe, S., M. Browne, and S. Boudjelas. (2000) 100 of the world's worst invasive alien species: a selection from the global invasive species database. Invasive Species Specialist Group Auckland, New Zealand.

Maderspacher, F. (2011) The benefits of bottlenecks. *Current Biology* 21:R171-R173.

Meirmans, P. G. and P. H. Van Tienderen. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4:792-794.

Mitich, L. W. (2000) Kudzu [Pueraria lobata (Willd.) Ohwi]. *Weed Technology* 14:231-235.

Molnar, J. L., R. L. Gamboa, C. Revenga, and M. D. Spalding. (2008) Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6:485-492.

Mooney, H. and E. Cleland. (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the United States of America* 98:5446.

Peakall, R. and P. E. Smouse. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Peakall, R. and P. E. Smouse. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* 28:2537-2539.

Petit, R. J., A. El Mousadik, and O. Pons. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12:844-855.

Pimentel, D., R. Zuniga, and D. Morrison. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* 52:273-288.

Piry, S., G. Luikart, and J. M. Cornuet. (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. . *Journal of Heredity*:502–503.

Pompanon, F. o., A. l. Bonin, E. Bellemain, and P. Taberlet. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6:847-846.

Prentis, P. J., J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. (2008) Adaptive evolution in invasive species. *Trends in Plant Science* 13:288-294.

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

Rousset, F. (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103-106.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, and N. C. Ellstrand. (2001) The Population Biology of Invasive Species. *Annual Review of Ecology and Systematics*:305-332.

Sasek, T. W. and B. R. Strain. (1988) Effects of carbon dioxide enrichment on the growth and morphology of kudzu (Pueraria lobata). *Weed Science*:28-36.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. (2009) Evolution and ecology of species range limits.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92:142-166.

Templeton, A. R., K. A. Crandall, and C. F. Sing. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619-633.

Tsugawa, H., T. Sasek, M. Tange, and K. Nishikawa. (1988) The fate of buds of kudzu-vine (Pueraria lobata Ohwi). *Journal of Japanese Society of Grassland Science (Japan)*.

Van Oosterhout, C., W. F. Hutchinson, D. P. Wills, and P. Shipley. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4:535-538.

Ward, D. B. (1998) Pueraria montana: the correct scientific name of the kudzu. *Castanea* 63:76-77.

Wechsler, N. R. (1977) Growth and Physiological Characteristics of Kudzu: Pueraria Lobata (Willd.) Ohwi, in Relation to Its Competitive Success. University of Georgia.

CHAPTER 4

A GENOME SCAN FOR SELECTION IN KUDZU REVEALS POSSIBLE ADAPTIVE CHANGES WITHIN THE NATIVE RANGE PRIOR TO INTRODUCTION INTO THE UNITED STATES

Introduction

Invasive species represent powerful natural experiments to understand the role of evolutionary change in species' responses to swift biotic or abiotic changes (Lee 2002, Moran and Alexander 2014). In a biological invasion, individuals are introduced to an environment to which their species were never previously exposed, where they establish, reproduce, and after a brief lag phase, rapidly spread to outcompete native species that had been evolving in that environment for millions of years (Sakai et al. 2001). It is in that lag phase where genetic changes that drive invasion are believed to occur (Crooks et al. 1999, Siemann and Rogers 2001). The genetic basis of those changes and the mechanisms that produced them are still matters of uncertainty (Moran and Alexander 2014). Increasing evidence has shown that natural selection can play a pivotal role in adaptation during invasion (Vandepitte et al. 2014), but the frequency of this evolutionary response and the rate it occurs is still unclear due to a paucity of studies that evaluate the genetic basis of phenotypic trait evolution during invasion (Moran and Alexander 2014). Additional genetic studies of invasive species are necessary to characterize the evolutionary process of invasion, specifically how adaptation in new environments in relatively short time frames is achieved (Prentis et al. 2008).

Previous studies have demonstrated that adaptive evolution can occur in a wide variety of phenotypic traits such as growth rate, flowering time, size and fecundity in tens of generations or less (Huey et al. 2000, García-Ramos and Rodríguez 2002, Maron et al. 2004, Dlugosch and Parker 2008b). Furthermore, studies comparing patterns of neutral genetic diversity among the native and invasive ranges have shown that multiple introductions, which are common to invasion, maintain and perhaps increase genetic variation in the invasive range (Kolbe et al. 2004, Dlugosch and Parker 2008a). Increased diversity improves evolutionary potential for rapid adaptation (Prentis et al. 2008). However, due to the low number of studies investigating the genetic basis of adaptation during invasion, it is unclear how much selection on diversity confers invasion success across species. With the recent development of modern sequencing technologies and new genomic methods that no longer require a reference genome (Allendorf et al. 2010, Davey and Blaxter 2010), it is possible to apply genome scans for selection to non-model invasive species in order to assess if and how natural selection may affect the genetic makeup of an expanding range (Vandepitte et al. 2014).

Kudzu is an aggressively expanding invasive species, and identifying the mechanism responsible for its success is of high environmental and economic interest (Blaustein 2001). Within a couple decades of its first introduction to the United States, kudzu began a rapid geographic range expansion (Miller and Edwards 1983). The current estimate of kudzu's range is 5 million hectares with an additional 50,000 hectares added per year (Blaustein 2001). It is known to cost utilities and agricultural and forestry industries millions of dollars per year as well as pose a serious threat to the conservation of native habitats (Forseth and Innis 2004). Phenotypic divergence in patch growth exists between the invasive range and part of its native range (personal observation); therefore, it is possible that adaptive divergence has occurred in

traits related to this phenotype and, or genes associated with increased clonal reproduction as a shift in clonal rates was inferred in the invasive range. Population genetic methods have been used to characterize the introduction history of kudzu into the United States (Chapter 2), but a role of natural selection in the invasion has never been assessed.

Previous work that compared genetic diversity patterns between the native and invasive ranges (Chapter 2) implies that some common evolutionary mechanisms are more applicable to kudzu than others. While both sexual and vegetative reproduction is possible in this long-lived perennial (Forseth and Innis 2004), recombination of genetic variation was infrequent (Chapter 2). Even if seeds are produced from sexual reproduction, germination and survival of seeds are low (Abramovitz 1983). Consequently, a paucity of sexual reproduction that may exist in this species could put constraints on evolutionary mechanisms of change during the invasion process (Barrett et al. 2008). Evolution through selection on new mutation can lead to changes in phenotype or behavior (Prentis et al. 2008); however, a lack of rampant population admixture to spread adaptive alleles into other genetic backgrounds implies that this mechanism is improbable after introduction to the United States. On the other hand, it is possible that selection on standing genetic variation in the invasive range or selection on new or standing genetic variation in putative source populations prior to introduction could have occurred. Putative source populations identified in Chapter 2 were significantly more diverse than the rest of the native range, providing ample genetic diversity upon which natural selection could have acted before or after introduction. Furthermore, analysis of population structure indicated that many of the putative sources might share more standing genetic variation as compared to the rest of the native range. Thus, evaluations of adaptive evolution between the invasive range and its sources as well as between native sources and the rest of the native range are warranted.

In this study, we test for signatures of natural selection in kudzu using a genome scan for selection. Selection on a favorable allele rapidly increases its frequency, bringing genetic variation surrounding the locus with it in a process called genetic hitchhiking (Smith and Haigh 1974, Kaplan et al. 1989). This process produces population differentiation at that locus that is significantly higher than neutral expectation and thus becomes a candidate target of selection (Lewontin and Krakauer 1973, Storz 2005). To effectively test for selective sweeps in kudzu, we compare genome-wide single nucleotide polymorphisms between the invasive range and its sources as well as between populations that contributed to the invasion and those that did not, while accounting for population structure and introduction history to reduce the false positive rate (Excoffier et al. 2009). Subsequent comparisons of allele frequencies of outlier loci were made to determine whether selection on new or standing genetic variation may have occurred and whether this likely occurred before or after introduction to the United States. We also compared overall genetic diversity patterns among these groups to confirm patterns in variation observed in previous studies.

Materials and Methods

Study species

Pueraria montana (Lour.) Merr. var. *lobata* (Willd.) Maesen & Almeida (Ward 1998) is a diploid legume (2n = 22 or 24) (Van der Maesen 2002) with an estimated genome size of 1.1 gigabases (Plant DNA C-values database maintained by the Kew Royal Botanical Gardens; http://www.kew.org/cvalues/homepage.html). Kudzu is a perennial leguminous vine that originated from Asia and was deliberately planted in the US for soil erosion and fodder since the late 1800's (Mitich 2000). In the 1930s and 1940s, over 85 million kudzu seedlings from the native range were planted in the United States for soil erosion control and nutritional fodder, but

the onset of the boll weevil and soil leeching left kudzu patches uncontrolled from agricultural abandonment (Miller and Edwards 1983, Corley et al. 1997, Forseth and Innis 2004). Growing 30cm per day, kudzu now covers over 5 million hectares of the United States (Blaustein 2001).

3RAD library preparation

To reduce ascertainment bias we included populations that spanned the entire native and invasive ranges with 495 native individuals from 52 Asian populations spread across China, South Korea and Japan and 272 invasive individuals from 31 populations from the United States. We prepared genomic libraries that maximized coverage at homologous loci using a double digest RAD-seq method called 3RAD (Adapterama III methodology, In Prep). 3RAD combines combinatorial tagging for large-scale pooling of individuals and a double digest RADseq protocol that utilizes three restriction enzymes to produce higher quality genomic libraries.

Each sample was extracted using either a Qiagen DNeasy Kit (Qiagen, Germantown, MD) or Axyprep Multisource Genomic DNA Miniprep Kits with the centrifuge protocol for plant tissue (Axigen Biosciences, Union City, CA). Concentrations were standardized to approximately 20ng/uL using a BioTek Synergy HT Plate Reader (Biotek, Winooski, VT). We digested a 15uL reaction comprised of 10 units of XbaI (New England Biolabs), 10 units of EcoRI (New England Biolabs), 1.66 units of NheI (New England Biolabs), 1X Cutsmart Buffer (New England Biolabs), 0.5μM NheI inner adapter, 0.5μM EcoRI inner adapter, 4uL ddH_2O and 8.3ng of template DNA at 37ºC for 1 hour. A 5uL aliquot of ligation mix comprised of 3 mM ATP (New England Biolabs), 100 units of T4 DNA ligase (New England Biolabs), 1X T4 DNA ligase reaction buffer (New England Biolabs) and 2.75uL ddH_2O was added to the digest. This digest-ligation mix was incubated at 22ºC for 20 minutes, 37ºC for 10 minutes, 22ºC for 20 minutes, 37ºC for 10 minutes and 65ºC for 20 minutes. All 96 samples of a 96-well plate were

pooled at equal volume, 10uL per sample. This pool was split into two and cleaned with a Speedbead protocol. Specifically, each 480uL pool was incubated for 15 minutes with equal volume of 1X Sera-mag Magnetic Streptavidin-coated Speedbeads (GE Healthcare), cleaned with 100uL 80% Ethanol twice on a DynaMag-2 Magnet (ThermoFisher Scientific) and re-suspended in 60uL TLE. Polymerase chain reactions were then performed on 25uL reactions of 3.75uL ddH_2O, 1X KAPA buffer, 2.5 units of KAPA Taq HotStart DNA Polymerase, 0.3mM of each dNTP, 0.5µM iTru5 adapter, 0.5µM iTru7 adapter and 10uL of the cleaned DNA. The PCR protocol incubated the samples at 95ºC for 2 minutes, 11 cycles of 98ºC for 20 seconds, 61ºC for 15 seconds, and 72ºC for 5 minutes. Each individual was given a unique quadruple indexed barcode for accurate de-multiplexing. PCR reactions were then pooled and subjected to an additional Speedbead cleaning process, re-suspended in 30uL TLE. 3RAD PCR products were run on a 1.5% dye-free agarose gel with marker K (CDF1510) and size selected for 550 bp fragments on Tight Range (average range of 496-604 bp) with a Pippn Prep (Sage Science, Beverly, MA). Size selected 3RAD libraries were pooled at equal volume and sequenced on two runs of an Illumina NextSeq High Output Flow Cell platform for paired-end 150 bp reads.

Processing sequencing reads to identify loci and SNPs

Raw sequence reads were processed in the program Fastqc (Andrews 2010) to identify the appropriate trim length and check quality standards. These reads were prepared for the Stacks pipeline using the process_radtags program (Catchen et al. 2013) after trimming reads to 120 bp in length, rescuing barcodes and tags, removing reads with uncalled bases and discarding low quality scores with a phred of 10. R2 reads were reversed complemented in FastX (Gordon and Hannon 2010). R1, R2 and remainder reads were concatenated into a single sample file for the Stacks pipeline (Jombart et al. 2010, Catchen et al. 2013). The Stacks pipeline, specifically in the

order of 'ustacks', 'cstacks' and 'sstacks', was performed on all samples using 3 as the minimum depth of coverage to create a stack, 0 as the maximum distance allowed between stacks, 4 as the maximum distance allowed to align secondary reads to primary stacks, 3 as the maximum number of stacks at a single de novo locus, 0 as the number of mismatches allowed between sample tags when generating the catalog, and a freely varying SNP calling model. Over-merged stacks were resolved with the Deleveraging algorithm and highly repetitive stacks were removed. The chi square significance level required to call a heterozygote or homozygote was set at 0.05. Remaining samples were utilized to identify the 999 loci present in 75% of individuals in the two groups, native and invasive, using the Stacks script 'populations'. Of those samples, 19 individuals were excluded due to 80% or greater missing data, leaving 237 samples across 29 US populations and 448 samples across 52 native populations for analysis.

Detecting population structure and putative source populations

To determine which native populations grouped with the invasive range as putative source populations, the 999 loci for all 685 individuals were subjected to genetic clustering analysis using Discriminant Analysis of Principal Components. To identify the best number of genetic clusters (K), all eigenvalues and principal components from the alpha score method were retained (Jombart 2013). Genetic clustering of individuals from this analysis determined whether a native population using genome-wide single nucleotide polymorphism data grouped the population with or without invasive individuals in the same genetic cluster, in what will be referred to as the "full DAPC analysis". Any entire population that did not cluster with any invasive individuals was classified as a "non-putative source". The rest of the native range that clustered with the United States were then retained a finer-scale Discriminant Analysis of Principal Components with the invasive individuals to further determine if any populations were

not putative source populations hereafter called the "subset DAPC analysis". In this second analysis, the best number of genetic clusters (K) was found by retaining all eigenvalues and principal components from the alpha score method. Invasive individuals found in a genetic cluster comprised of only invasive populations were not used in the genome scan for selection comparing the putative source populations to the invasive range as no native source population for these individuals was not included in the dataset. Native populations that did not cluster with invasive individuals in the subset DAPC analysis were added to the list of non-putative source populations. Native populations that shared a genetic cluster with invasive individuals were classified as putative sources populations.

Comparing genetic diversity and identification of outlier loci

There were two comparisons for the genome scan for selection, a within-range comparison that compared F_{ST} values between putative and non-putative source populations and an among-range comparison between putative source populations and the invasive range. For each of the two comparisons, genetic clustering results from DAPC were utilized to maximize the number of loci compared among groups by re-running the Stacks 'populations' script that retained all loci found in 80% of all individuals in each DAPC-based group in a comparison. In the within-range comparison, there were 2 groups, putative and non-putative source populations. In the comparison across ranges, there were four groups, two groups per range delineated based on whether or not the individuals fell into Cluster 2 or 4 in the subset DAPC analysis.

The genotypes of each group for each comparison were subjected to genetic and clonal diversity analysis. We estimated the percentage of polymorphic loci (P%), allelic richness (N_A), effective number of alleles (N_E), observed heterozygosity (H_O), Nei's unbiased expected heterozygosity (uH_E), and the fixation index (F) using GenAlEx version 6.501 (Peakall and

Smouse 2006, 2012). An AMOVA was performed in this program for each comparison in addition to analysis of molecular variance within the invasive range using 999 permutations. Clonal diversity statistics were calculated after determining a clonal threshold of 1 in the within-range comparison and a threshold of 3 in the among-range comparison in Genodive (Meirmans and Van Tienderen 2004), including the number of distinct multilocus genotypes (MLG), the effective number of genotypes (eff), evenness (eve), Nei's genetic diversity (div), and the proportion distinguishable (R) defined as (MLG-1)/(N-1) (Ellstrand and Roose 1987). Genetic and clonal diversity measures were also compared within the United States to observe how diversity varied across invasive genetic clusters identified in the subset DAPC run.

Outlier loci demonstrating locus-specific genetic differentiation among groups with F_{ST} values higher or lower than the genomic average were detected based on a multinomial-Dirichlet model in Bayescan (Foll and Gaggiotti 2008). Bayescan has been shown to identify outlier loci with minimal type I and II error compared to other methods (Pérez-Figueroa et al. 2010, Narum and Hess 2011). Furthermore, Bayescan is based on a island model in which subpopulations share a common migrant pool by varying degrees , accommodating species with asymmetrical gene flow among populations (Nielsen et al. 2009) as was inferred by patterns of genetic diversity in invasive kudzu (Chapter 2). Locus population F_{ST} values were split into locus-specific (α) and population-specific (β) components. Posterior probabilities were estimated with and without the locus-specific components, and a locus was deemed to be under selection if the a parameter was necessary to explain the F_{ST} value. Outlier loci were detected based on a false discovery rate of 5% and the q-value, an multiple-testing analogue of a p-value that represents the minimum false discovery rate necessary to find the locus significant (Foll and Gaggiotti 2008). We implemented 20 pilot runs of 5000 iterations, with a burn-in of 50000 iterations and

an additional 5000 iterations with a thinning interval of 10. Consensus sequences were blasted against the NCBI BLASTN database. Allele frequencies were evaluated for both comparisons in GenAlEx version 6.501 (Peakall and Smouse 2006, 2012). Changes in allele frequencies within the native range were also compared to allele frequencies in the invasive range for all loci but one, which was found in less than 75% of US individuals. While these loci were not included in this within-range comparison, US allele frequencies were included to provide context as to changes in allele frequencies in the native range.

Results

Population structure within and among ranges

In the Discriminant Analysis of Principal Components with the entire dataset (full DAPC analysis), a K of 3 was determined as best based on the regression of Bayesian Information Criteria against a range of possible clusters (K) (Figure 1). Fifty-eight principal components and all 2 eigenvalues were retained, accounting for 79.7% of cumulative variance. With this K, 99% of all individuals were correctly assigned with individual clusters exhibiting correct assignment at 98% or higher. In the results, the proportion assignment of every individual was 100% to a given cluster (Figure 2). Individuals of ten populations did not cluster with any invasive individuals and were classified as "non-putative". When these native individuals were removed, remaining individuals were subjected to the finer-scale subset DAPC analysis. A K of 5 was determined as best (Figure 3), including 46 principal components and all 4 eigenvalues, accounting for 74.4% of the cumulative variance. With a K of 5, 99% of all individuals were correctly assigned with individual clusters exhibiting correct assignment at 99% or higher. Four native populations did not cluster with invasive individuals and were added to the list of non-putative source populations (Figure 4), leading to a total of 22 non-putative source populations.

The remaining 250 individuals from 31 native populations were considered putative source populations. One genetic cluster was comprised of 21 individuals from four invasive populations only (Figure 4), indicating that a putative source population of these populations was not included in this study. Thus, these individuals were excluded from the genome scan.

Genetic and clonal diversity within and among ranges

Within the invasive range, population structure between DAPC-based clusters was low with a ϕ_{ST} of 0.061. Moreover, genetic diversity measures did not vary greatly between invasive individuals assigned to Clusters 2 and 4, but there was differentiation between those clusters and Cluster 1, which did not contain native sources (Table 1). Cluster 1 exhibited slightly less allelic richness (N_A = 1.20) as well as a negative fixation index (F = -0.330) than both Cluster 2 (N_A = 1.75; F = 0.051) and Cluster 4 (N_A = 1.72; F = 0.056). Observed heterozygosity did not vary across the three clusters (Table 1), but there was less expected heterozygosity in US Cluster 1 (uH_E= 0.058) than Cluster 2 (uH_E= 0.110) and Cluster 4 (uH_E= 0.108). While the number and effective number of multilocus genotypes was higher in Cluster 4 (MLG = 44, eff = 24.5) than Cluster 1 (MLG = 4, eff = 2.22) and Cluster 2 (MLG = 28, eff = 14.8), clonal diversity (div) and the proportion distinguishable (R) were similar between Cluster 2 (div = 0.946; R = 0.620) and Cluster 4 (div = 0.966; R = 0.699) but lower in Cluster 1 (div = 0.576; R = 0.850). Evenness largely did not vary across groups (Table 2).

In the among-range comparison, 1039 loci were detected across the two clusters, Cluster 2 and 4, in both putative source and invasive populations. Genetic differentiation between the two groups was low with a ϕ_{ST} of 0.02. Genetic diversity measures between putative sources and invasive clusters did not show extensive differentiation either (Table 1). Allelic richness was slightly higher in Asian Cluster 2 than US Cluster 2 ($N_{A(ASIA)}$ = 1.84; $N_{A(US)}$ = 1.41), but similar

allelic richness in Cluster 4 ($N_{A(ASIA)}$ = 1.35; $N_{A(US)}$ = 1.38). The rest of the genetic diversity measures did not vary across clusters or ranges (Table 1). Observed heterozygosity largely mirrored expected heterozygosity for all groups. Fixation indices also did not vary, ranging from 0.049 to 0.083. On the other hand, clonal diversity measure did show some differences between ranges. Based on a clonal threshold of three, there were 124 distinct multilocus genotypes in Asia compared to the 69 multilocus genotypes in the United States (Table 2), and the effective number of genotypes was largely reduced by 50% in each group. Evenness was higher in Asia ($eve_{CLUSTER2}$ = 0.649; $eve_{CLUSTER4}$ = 0.688) than the invasive range ($eve_{CLUSTER2}$ = 0.482; $eve_{CLUSTER4}$ = 0.571). Clonal diversity estimates were similar across groups and close to 1, ranging from 0.886 to 0.991 (Table 2). Cluster 2 exhibited a higher proportion distinguishable in the native range (R_{ASIA} = 0.511) than the invasive range (R_{US} = 0.366), while the difference in R in Cluster 4 among ranges was less pronounced (R_{ASIA} = 0.333; R_{US} = 0.287).

In the within-native range comparison, 533 loci across the final list of putative and non-putative native source populations were detected with at least 80% of all individuals in each group. Population structure was higher than the other comparison but relatively close to 0 with a ϕ_{ST} of 0.079. Genetic diversity measures were similar between putative source populations and the rest of the native range (Table 1), however, the fixation index was lower in the putative sources (F = 0.071) than the rest of the native range (F = 0.173) by a difference of 10%. In terms of clonal diversity, there were more multilocus genotypes, a higher effective number of genotypes, higher clonal diversity and evenness in the putative sources compared to the rest of the native range (Table 2). The proportion distinguishable was higher in the putative sources as well at 0.518 as compared to 0.365, a difference of 35%.

When genetic diversity measures were compared among ranges overall (Grand means), genetic diversity measures were largely consistent (Table 1). There was a similar amount of allelic richness between the native and invasive range ($N_{A(US)}$ = 1.554, $N_{A(ASIA)}$ = 1.642) as was the effective number of alleles ($N_{E(US)}$ = 1.141; $N_{E(ASIA)}$ = 1.096). Neither range exhibited high observed heterozygosity ($H_{O(US)}$ = 0.097; $H_{O(ASIA)}$ = 0.047) or large differences from expected heterozygosity ($uH_{E(US)}$ = 0.092; $uH_{E(ASIA)}$ = 0.067). The fixation index was lower in the invasive range than in Asia ($F_{(US)}$ = 0.007; $F_{(ASIA)}$ = 0.131). Clonal diversity measures were also different across ranges (Table 2). The effective number of genotypes was higher in Asia (eff_{US} = 13.8; eff_{ASIA} = 59.1) with higher clonal diversity (div_{US} = 0.829; div_{ASIA} = 0.986) and evenness (eve_{US} = 0.547; eve_{ASIA} = 0.579). The average proportion distinguishable across invasive groups was 0.277, but the average in Asia was higher at 0.442, a difference of 45%.

F_{ST} Outliers

When putative source and non-putative source populations were compared in Bayescan defined as these two groups, no F_{ST} outlier loci were detected, but comparisons between native individuals defined by population and compared based on this definition found 27 F_{ST} outlier loci (Figure 5). Eight loci were significantly lower than the genomic F_{ST} average ($0.498 \pm 2.80 \times 10^{-3}$), ranging from 0.156 to 0.378 (Table 3). These outlier loci exhibited a decrease in $F_{ST,}$ indicative of modest changes in the minor allele frequency (Table 3). The 19 other loci were significantly higher than neutral expectation, ranging from 0.670 to 0.880 (Table 3). These outlier loci with an increase in F_{ST} were the result of an allele in the putative sources shifting to high frequency or fixation relative to non-putative sources. All outlier loci were highly degenerative and repetitive indicating that they were not likely genic sequences. There were no

significant blast hits for any of the consensus sequences for these loci suggesting they were also not likely sequences within transposable elements.

Distinct patterns in the allele frequencies of these outlier loci were observable between the putative and non-putative sources, and these patterns were supported by allele frequencies in the US. For example, seven loci (4425_7, 12130_16, 13333_77, 14595_0, 15058_6 and 16956_92, 18229_5) contained alleles segregating in the non-putative sources that were fixed in the putative sources. Those alleles were also fixed in the US for all seven cases. Three loci (3604_4, 4066_5 and 6556_16) exhibited an increase of the more common allele in putative source populations (frequency greater than 50%) and complete fixation in the US. Nine loci (2637_2, 2755_32, 5363_12, 6133_2, 9679_26, 15490_33, 16332_71, 17540_34 and 21565_35) exhibited a large increase of the minor allele in the putative sources (at least 30% change), four of which became fixed in the US. Locus 6789 exhibited an increase in the minor allele frequency in the putative source group (4% vs. 1%), but the US was fixed for the major allele. One locus (8960_13) exhibited an increase of the minor allele frequency from 0.03 to 0.25, but none of the two alleles in the native groups were present in the US, which was fixed for a third allele at this locus. The rest of the loci exhibited modest changes in allele frequency between the two native groups but there were no sweeps to fixation, including in the US.

When individuals from the invasive range and putative source populations were compared in Bayescan, no outlier loci were detected using DAPC-based groupings. However, when compared on a population level, Bayescan results indicated that there were four outlier loci between the native and invasive ranges (Figure 6). All were significantly lower than the genomic average (F_{ST} 0.401 ± 6.22 x 10^{-4}) (Table 4). Two of the four loci in this comparison involved a loss of a low to intermediate frequency allele in the United States as compared to its putative

sources (Table 4). The other two involved the emergence of an allele not found in the putative source populations. Specifically, locus 7536_22 was monomorphic across all groups except in Cluster 2 of the United States with the presence of a novel low frequency allele at 3%. It should be noted that locus 7526_22 was also detected as an outlier in the within-native range comparison as well but with different alleles. Locus 9306_4 was similar with the emergence of a low frequency allele at 6% in Cluster 4 of the United States not found in the putative sources. Blast results of these four outlier loci did not any produce significant hits, but they were also highly degenerative and repetitive in nature.

Discussion

This study is the first attempt to assess the role of natural selection in the invasion history of kudzu, using a genome scan for selection to identify candidate loci that could underlie adaptation or be linked to such loci. We sequenced and developed hundreds to thousands of homologous loci at high coverage *de novo* and utilized population structure from discriminant analyses of principal components across native and invasive populations to identify which native populations clustered with the invasive range as potential sources to the invasion. In addition to comparing levels of genetic and clonal diversity within and among the native and invasive ranges, we searched for outlier loci within the native range between the putative and non-putative source populations as well as among ranges because the putative source populations and the invasive range to assess the possible nature and timing of natural selection in the invasion history.

Genetic diversity results in this study of genome-wide SNPs were surprisingly different from other marker types. The previous study using microsatellites (Chapter 2) suggested that kudzu was highly clonal and more diverse than species with similar life histories (Hamrick and

Godt 1996) with the possibility that high clonal reproduction might be driving high heterozygosity levels (Chapter 2). However, our SNP diversity patterns imply that this species may be more sexually reproducing than previously thought. There was high genotypic diversity and no excess heterozygosity (Table 1), which does not meet the expectations of a highly clonal organism (Balloux et al. 2003). The large differences in heterozygosity among marker types has led to alternate conclusions about the amount of sexual reproduction in the species. The SNP study is likely more robust as it samples hundreds of loci across the genome, but it does not necessarily mean that the microsatellite study is incorrect. Discrepancies between microsatellite and SNPs could be due to increased mutation rates in microsatellites (Morin et al. 2004), higher number of allelic states that can accumulate at a locus (Vignal et al. 2002), background selection against slightly deleterious single nucleotide polymorphisms (Charlesworth et al. 1993), or amplification of other microsatellite loci during PCR (Pompanon et al. 2005), but it highlights the importance of studying an evolutionary question using multiple markers across the genome to determine genetic diversity across the genome.

Not all conclusions about the invasion history from the microsatellite study were negated by the SNP dataset. Similar to the microsatellite study, this study also demonstrated only mild evidence of bottlenecking after the introduction of multiple sources. Thirty-one putative source populations rather than eleven were identified, indicating that more introduction events from across the native range may have occurred. There were fewer multilocus genotypes in the United States as compared to these putative source populations, but genetic and clonal diversity measures were similar between these two groups. There was not an excess of heterozygosity or large decline in allelic richness typical of bottleneck events (Allendorf and Lundquist 2003).

The SNP dataset supported the hypothesis that the putative source populations were more sexually reproducing than the rest of the native range before shifting to higher clonal rates in the United States. All clonal diversity measures, including the number and effective number of multilocus genotypes, clonal diversity and evenness were higher in the putative source populations (Table 2), translating to a lower proportion distinguishable than the rest of the native range. Furthermore, while inferred vegetative rates were reduced overall, the invasive range was consistently more clonal with an inferred clonal rate at 72% than its putative sources at 52% and the rest of the native range at 64%. This was consistent within and among all genetic clusters compared. Moreover, the US genetic cluster that did not associate with any native populations, Cluster 1, exhibited even higher inferred clonal rates than the other two US clusters (85% vs. 62% and 69%, respectively), although it is unclear if there was a shift in clonal reproduction in that introduction event as well. Can you speculate briefly on what these results might mean? That the putative source populations were pre-adapted or there was selection for higher clonality in the ramets used in the introduction? You were telling me about it last week.

In the genome scan for selection, population differentiation was low within and among the invasive and native ranges, increasing the chances of detecting of real adaptive changes as opposed to outlier loci as an artifact of population structure (Hohenlohe et al. 2010). While these are only candidate loci that require further testing for natural selection, the majority of cases demonstrated significantly higher F_{ST} values indicative of possible directional selection, but there were also cases of decreased F_{ST} that imply possible balancing selection (Figures 1 and 2) (Foll and Gaggiotti 2008). It is more likely that these loci were linked to an allele under selection rather than the targets themselves as the regions are highly repetitive and not reflective of genetic variation common in genic regions. Alternatively, these loci may simply be the product of

genetic drift (Storz 2005). These loci were largely degenerate, repetitive and did not show hits to GenBank, suggesting they are not likely in genic regions or transposable elements.

If allelic frequency changes at any of these loci or closely linked regions were truly adaptive, then natural selection likely acted on standing genetic variation. All but one of the 27 loci in the native range involved changes in the frequencies of standing variation as opposed to the presence of new mutation. These changes in allele frequencies of previously segregating alleles were drastic among native groups with magnitudes of shifts as high as 75% with a mean of 35%. Selection on standing genetic variation is believed to a more effective means to achieve adaptation as time to fixation is less than that of new mutation (Prentis et al. 2008). However, it should also be noted that the chance of stochastic fixation due to drift from standing variation is also increased, depending on the allele frequency and the effective population size (Lande 1976). Frequency of changes in standing genetic variation versus the rise of new mutation is unclear in the invasive range as only four outliers were detected, two of which involved the emergence of an allele not in the putative source populations.

Furthermore, the majority of the outlier loci were found within the native range as opposed to among ranges, suggesting evolutionary changes likely occurred in the putative sources relative to the rest of the native range prior to introduction to the United States. In fact, the majority of outlier loci in the native range could represent hard selective sweeps in the putative sources (Storz 2005), which were then carried over to the United States. However, it is unclear in which direction selection would favor or disfavor an allele. It is easy to conclude that changes in allele frequencies represented positive selection on standing variation in the putative source populations for increased invasiveness in the introduced kudzu; however it may be

negative selection on a deleterious allele. The directionality of selection in these cases is unknown.

Fewer evolutionary changes were detected in the invasive range after introduction, and that changes that did occur were of a smaller magnitude. This is to be expected given the significantly smaller amount of evolutionary time that kudzu has existed in the United States. In addition to possible balancing selection or the beginnings of a selective sweep that has not yet gone to fixation, outlier loci in the United States could be the result of genetic drift from founder effects. Of the many outlier loci that exhibited fixation of alleles within the invasive range, fixation events could be also be cases of gene surfing, where alleles reach high frequencies or fixation due to population expansion (Excoffier and Ray 2008). It is also possible that the exclusion of important source populations or insubstantial selection gradients among the native and invasive ranges produced few outliers within the invasive range, though our sampled populations included a large proportion of the native and invasive range. Lastly, our ability to detect outlier loci is dependent on the genetic architecture of adaptive traits. If few genes of large effect were responsible for adaptive evolution in kudzu, then our sampling of the genome may be insufficient to include them. On the other hand, phenotypic differentiation from many genes of small effect could have produced small enough changes to allele frequencies to be largely undetectable using genome scans (Riquet et al. 2013).

Lastly, there is the possibility that natural selection is not the main driver of invasion in this species at all. Outliers were only present in the population-level comparison not based on the DAPC clustering of the native range, suggesting that patterns in allelic frequencies between the putative and non-putative sources may not be adaptive change between these two groups specifically but rather fixation events in certain populations or local adaptation of a few

populations which played little role in the invasion of the United States. An alternative hypothesis is that significant population differentiation at these loci are not differences in natural selection across the range but rather exogenous (e.g. temperature, precipitation) or endogenous (e.g. incompatible loci within the genome) barriers to gene flow (Bierne et al. 2011). In fact, previous work with microsatellites and chloroplast haplotypes demonstrated that genetic diversity in the native range was well structured between the northern and southern parts of the range (Chapter 2), suggesting that exogenous barriers to gene flow are possible. Furthermore, while it is tempting to attribute population differentiation at outlier loci to natural selection, it is possible that these changes represent non-adaptive events across the range (Hermisson 2009). For example, kudzu may be invasive due to phenotypic plasticity and outlers detected were simply due to genetic drift or rapid range expansion. Phenotypic plasticity is known to be an important mechanism for handling environment heterogeneity (Richards et al. 2006), particularly for species capable of clonal reproduction (Barrett et al. 2008).

Genome scans for selection are not the end point for understanding adaptation during invasion but rather the beginning. After the identification of outlier loci, further studies are needed to characterize the population differentiation surrounding those loci (Barrett and Hoekstra 2011). Genome scans enable detection of possible targets of selection without a priori knowledge of what phenotype traits have evolved. Consequently, genotypes needs to be connected with phenotypes as not all alleles that fix in a population will be within or linked to a functional gene. Secondly, an allele is only adaptive if it increases the fitness of those that possess it. Contributions of alleles at candidate loci require assessment through additional studies including additional observational studies in nature, common garden studies, and selection experiments. Lastly, other evolutionary mechanisms such as phenotypic plasticity should also be

tested to determine which evolutionary mechanisms most likely contributed to the invasive nature of a species.

Tables

Table 4.1. Genetic diversity measures within and among ranges based on genetic clusters identified in DAPC analysis. N_A = allelic richness; N_E = effective number of alleles; H_O = observed heterozygosity; uH_E = expected heterozygosity; F = fixation index.

Genetic Diversity Measure	Grand Mean				Among invasive groups (1367 loci)					
	US		Asia		US Cluster 1		US Cluster 2		US Cluster 4	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
N_A	1.554	0.008	1.642	0.015	1.202	0.011	1.747	0.012	1.715	0.012
N_E	1.141	0.004	1.096	0.006	1.103	0.008	1.159	0.007	1.160	0.007
H_O	0.097	0.003	0.047	0.003	0.090	0.007	0.101	0.004	0.100	0.004
uH_E	0.092	0.002	0.067	0.004	0.058	0.004	0.110	0.004	0.108	0.004
F	0.007	0.005	0.131	0.008	-0.330	0.012	0.051	0.007	0.056	0.006

Genetic Diversity Measure	Among native sources and invasive groups (1039)								Among native groups (553 loci)			
	Asia Cluster 2		US Cluster 2		Asia Cluster 4		US Cluster 4		Non-putative sources		Putative sources	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
N_A	1.844	0.011	1.411	0.015	1.353	0.015	1.388	0.015	1.762	0.018	1.522	0.022
N_E	1.089	0.006	1.083	0.006	1.096	0.006	1.085	0.006	1.114	0.009	1.079	0.009
H_O	0.058	0.003	0.053	0.003	0.058	0.004	0.054	0.004	0.048	0.003	0.046	0.005
uH_E	0.065	0.003	0.058	0.003	0.068	0.004	0.058	0.004	0.081	0.005	0.052	0.005
F	0.049	0.005	0.055	0.008	0.083	0.011	0.050	0.007	0.173	0.013	0.071	0.007

Table 4.2. Clonal diversity estimates within and among ranges based on the full and subset DAPC runs. All values represent group averages except N and MLG, which represent total number. T indicates the clonal threshold set for each dataset. N = sample size; MLG = number of multilocus genotypes in a population; eff = effective number of genotypes; div = Nei's genetic diversity; eve = evenness; R = proportion distinguishable; 1-R = proportion indistinguishable also referred to as the inferred rate of clonal reproduction

Clonal Diversity Measures	Grand Mean		Among invasive clusters (T=6)			Among putative sources & invasive clusters (T=3)				Among native clusters (T=1)	
	Asia	US	US Cluster 1	US Cluster 2	US Cluster 4	Asia Cluster 2	US Cluster 2	Asia Cluster 4	US Cluster 4	Non-putative sources	Putative sources
N	448	237	21	72	144	222	72	28	144	198	250
MLG	203	76	4	28	44	114	27	10	42	73	130
eff	59.1	13.8	2.216	14.811	24.511	74.0	13.0	6.88	24.0	41.5	76.6
div	0.986	0.829	0.576	0.946	0.966	0.991	0.936	0.886	0.965	0.981	0.991
eve	0.579	0.547	0.554	0.529	0.557	0.649	0.482	0.688	0.571	0.569	0.589
R	0.442	0.277	0.150	0.380	0.301	0.511	0.366	0.333	0.287	0.365	0.518
1-R	0.559	0.723	0.850	0.620	0.699	0.489	0.634	0.667	0.713	0.635	0.482

Table 4.3. Allele frequencies of outlier loci in the native putative source and non-putative source populations with US allele frequencies included providing context as to the changes in the native range.

Locus	Allele/n	Non-putative sources	Putative sources	US
108_44	**N**	168	207	200
	1	0.235	0.181	0.063
	4	0.765	0.819	0.938
935_2	**N**	198	219	226
	1	0.955	0.874	0.876
	3	0.045	0.126	0.124
2637_2	**N**	202	213	196
	1	0.443	0.756	0.798
	3	0.557	0.244	0.202
2755_32	**N**	173	232	225
	3	0.717	0.032	0.000
	4	0.283	0.968	1.000
3358_29	**N**	193	235	218
	2	0.233	0.130	0.174
	4	0.767	0.870	0.826
3604_4	**N**	171	235	197
	1	0.412	0.006	0.000
	3	0.588	0.994	1.000
4066_5	**N**	191	237	229
	2	0.416	0.027	0.000
	4	0.584	0.973	1.000
4425_7	**N**	199	231	222
	2	0.847	1.000	1.000
	4	0.153	0.000	0.000
5363_12	**N**	167	229	226
	1	0.314	0.969	1.000
	3	0.686	0.031	0.000
6133_22	**N**	201	235	235
	1	0.373	0.883	0.943
	3	0.627	0.117	0.057
6556_16	**N**	184	209	191
	2	0.440	0.017	0.000
	4	0.560	0.983	1.000
6789_114	**N**	173	197	189
	1	0.000	0.000	0.003
	2	0.991	0.964	0.997
	4	0.009	0.036	0.000
7536_22	**N**	190	228	207
	1	0.016	0.125	0.169
	4	0.984	0.875	0.831
7968_7	**N**	158	228	215
	1	0.155	0.261	0.247
	4	0.845	0.739	0.753
8960_13	**N**	154	181	209
	1	0.055	0.254	0.000
	3	0.000	0.000	1.000

	4	0.945	0.746	0.000
9679_26	**N**	191	223	210
	2	0.510	0.009	0.000
	4	0.490	0.991	1.000
12130_16	**N**	165	206	220
	1	0.555	1.000	1.000
	4	0.445	0.000	0.000
13333_77	**N**	202	212	222
	2	0.616	1.000	1.000
	4	0.384	0.000	0.000
14595_0	**N**	199	217	223
	1	0.769	1.000	1.000
	2	0.231	0.000	0.000
15058_6	**N**	192	239	220
	2	0.339	0.000	0.000
	4	0.661	1.000	1.000
15490_33	**N**	188	215	206
	2	0.279	0.953	1.000
	4	0.721	0.047	0.000
16332_71	**N**	200	226	224
	1	0.783	0.077	0.092
	2	0.218	0.923	0.908
16956_92	**N**	202	183	215
	1	0.351	0.000	0.000
	3	0.649	1.000	1.000
17540_34	**N**	190	208	224
	1	0.755	0.091	0.029
	3	0.245	0.909	0.971
17909_33	**N**	157	198	0
	2	0.025	0.010	0.000
	4	0.975	0.990	0.000
18229_5	**N**	163	212	196
	1	0.647	0.998	1.000
	4	0.353	0.002	0.000
21565_35	**N**	183	205	230
	2	0.770	0.022	0.007
	4	0.230	0.978	0.993

Table 4.4. Allele frequencies of outlier loci in the putative source populations and invasive populations. Changes in allele frequencies were compared between both ranges based on the genetic clustering determined in the subset DAPC run. The Allele/n column includes the SNP number or N to indicate sample size for that row.

		Asia		US	
Locus	**Allele/n**	**Cluster2**	**Cluster4**	**Cluster2**	**Cluster4**
7536_22	N	217	26	60	127
	1	1.000	1.000	0.975	1.000
	4	0.000	0.000	0.025	0.000
9306_4	N	221	26	67	141
	1	0.000	0.000	0.000	0.057
	3	1.000	1.000	1.000	0.943
9392_11	N	208	23	69	140
	1	0.995	0.870	1.000	1.000
	4	0.005	0.130	0.000	0.000
9684_55	N	206	24	58	131
	2	0.012	0.000	0.000	0.000
	4	0.988	1.000	1.000	1.000

Table 4.5. This table is the key for population numbers on the x-axis in both the full and subset DAPC analysis, Figures 4.2 and 4.4, respectively.

X-axis number	**Population Code**	
	Full DAPC Analysis (K = 3)	**Subset DAPC Analysis (K = 5)**
1	AL5	KY11
2	AL12	KY8
3	AR2	NC21
4	FL3	NC3
5	GA3	TX5
6	GA34	AL12
7	GA36	GA3
8	GA96	GA34
9	KY11	MS4
10	KY7	NC13
11	KY8	SC4
12	MD1	GA36
13	MO1	KY11
14	MS4	GA96
15	MS6	AR2
16	NC13	NC6
17	NC20	NC20
18	NC21	NY4
19	NC3	TN6
20	NC6	WV1
21	NE1	KY7

22	NY4	AL5
23	OK1	FL3
24	SC4	KY8
25	TN6	MD1
26	TN8	MO1
27	TX4	MS6
28	TX5	NC3
29	WV1	NE1
30		OK1
31		TN8
32		TX4
33		TX5

Figures

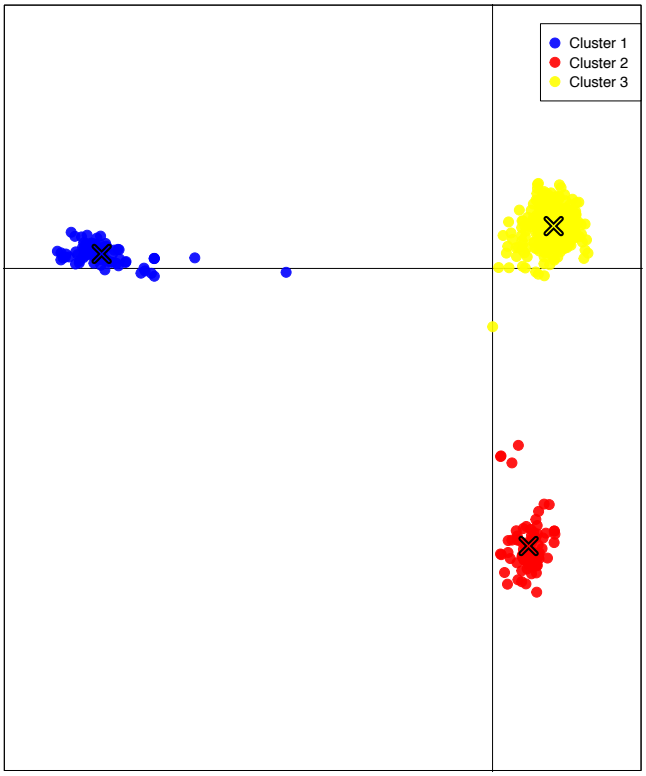


Figure 4.1. DAPC plot of the full DAPC analysis that included all native and invasive populations, demonstrating a best K of 3.

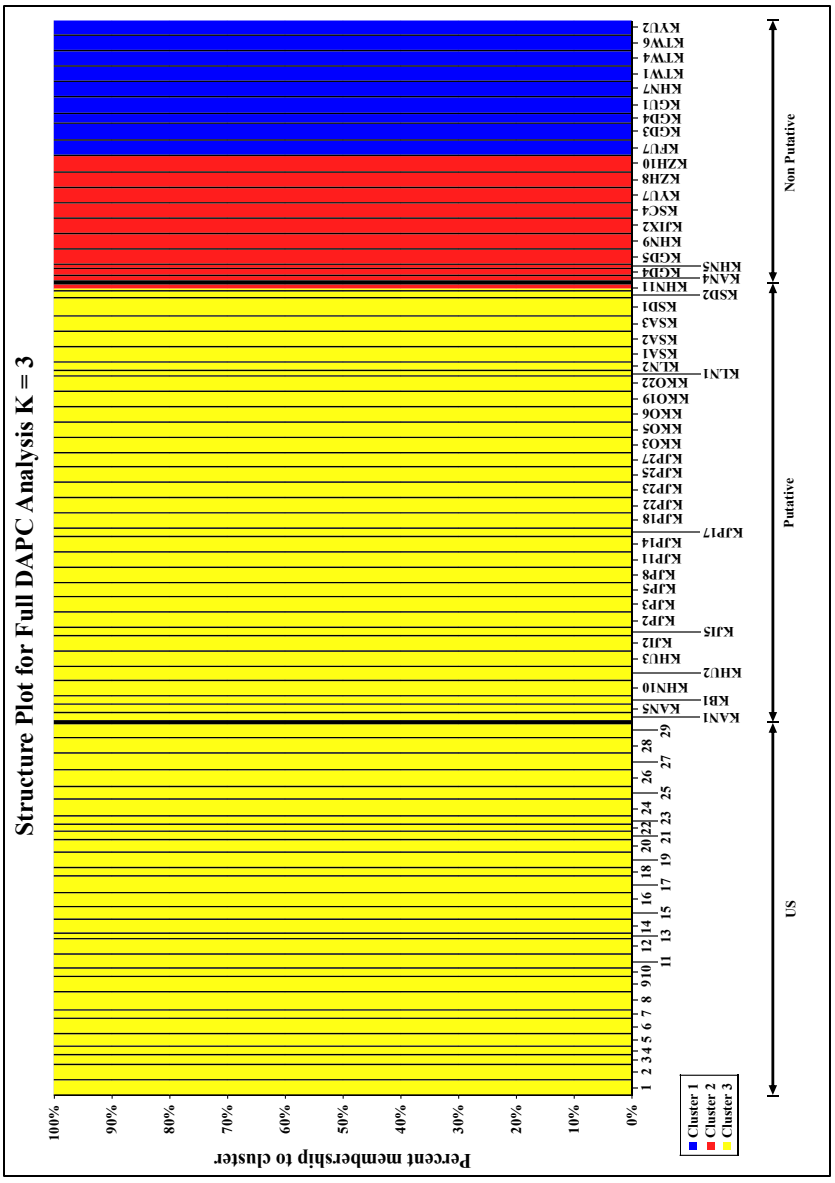


Figure 4.2. Structure plot for full DAPC analysis that includes all native and invasive populations and demonstrated a best K of 3. This plot reflects the percent membership of a given individual to a cluster. The population names of invasive populations can be found in Table 4.5.

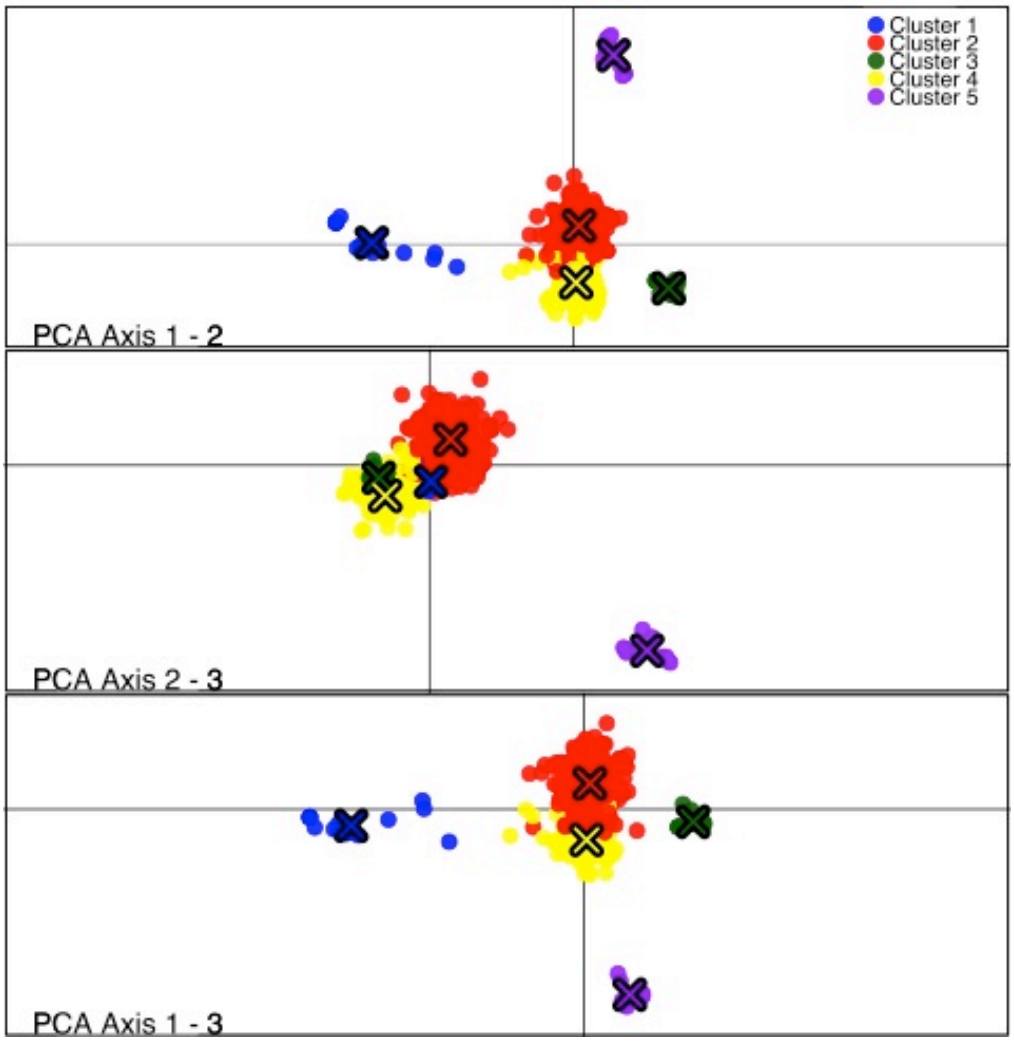


Figure 4.3. DAPC plot of the subset DAPC analysis with a best K of 5 after excluding native populations that did not cluster with invasive individuals in the full DAPC analysis.

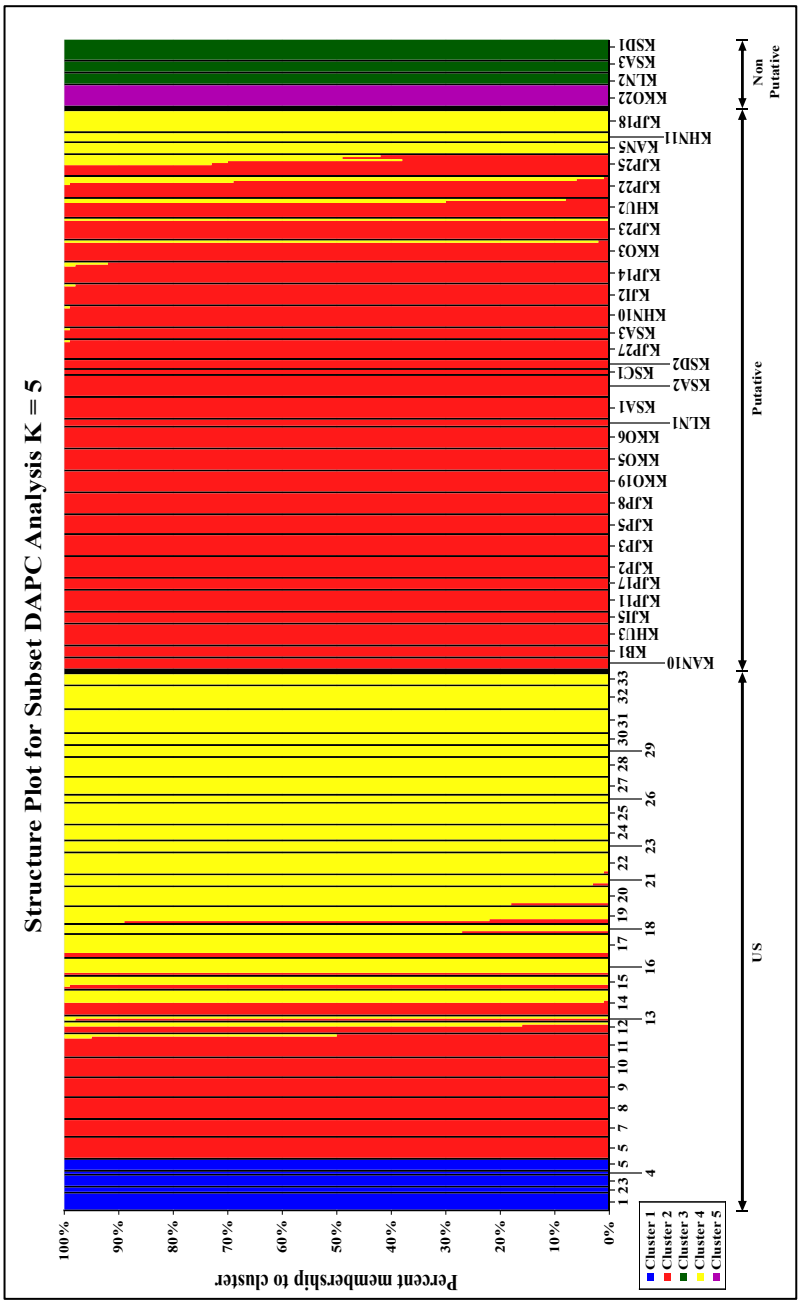


Figure 4.4. Structure plot of the subset DAPC analysis demonstrating a best K of 5 after the exclusion of native populations that did not cluster with invasive individuals in the full DAPC analysis. This plot reflects the percent membership of a given individual to a cluster. The population names of invasive populations can be found in Table 4.5.

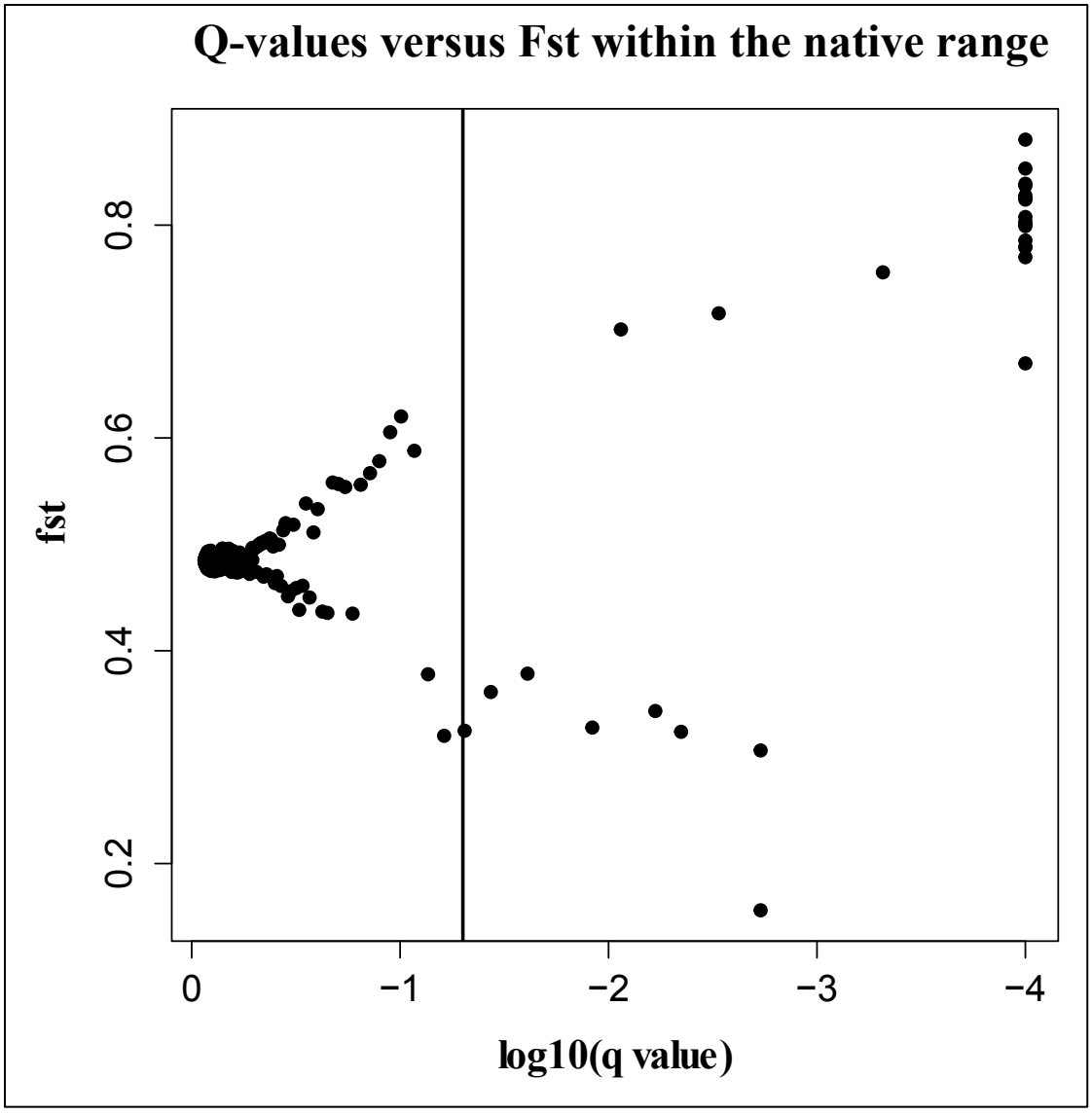


Figure 4.5. Q-values of loci versus F_{ST} in the within-range comparison between putative and non-putative source populations. The vertical line between -1 and -2 likelihood values on the x-axis indicates the threshold of significance with a 5% False Discovery Rate. All values to the right of the vertical bar represent outlier loci.

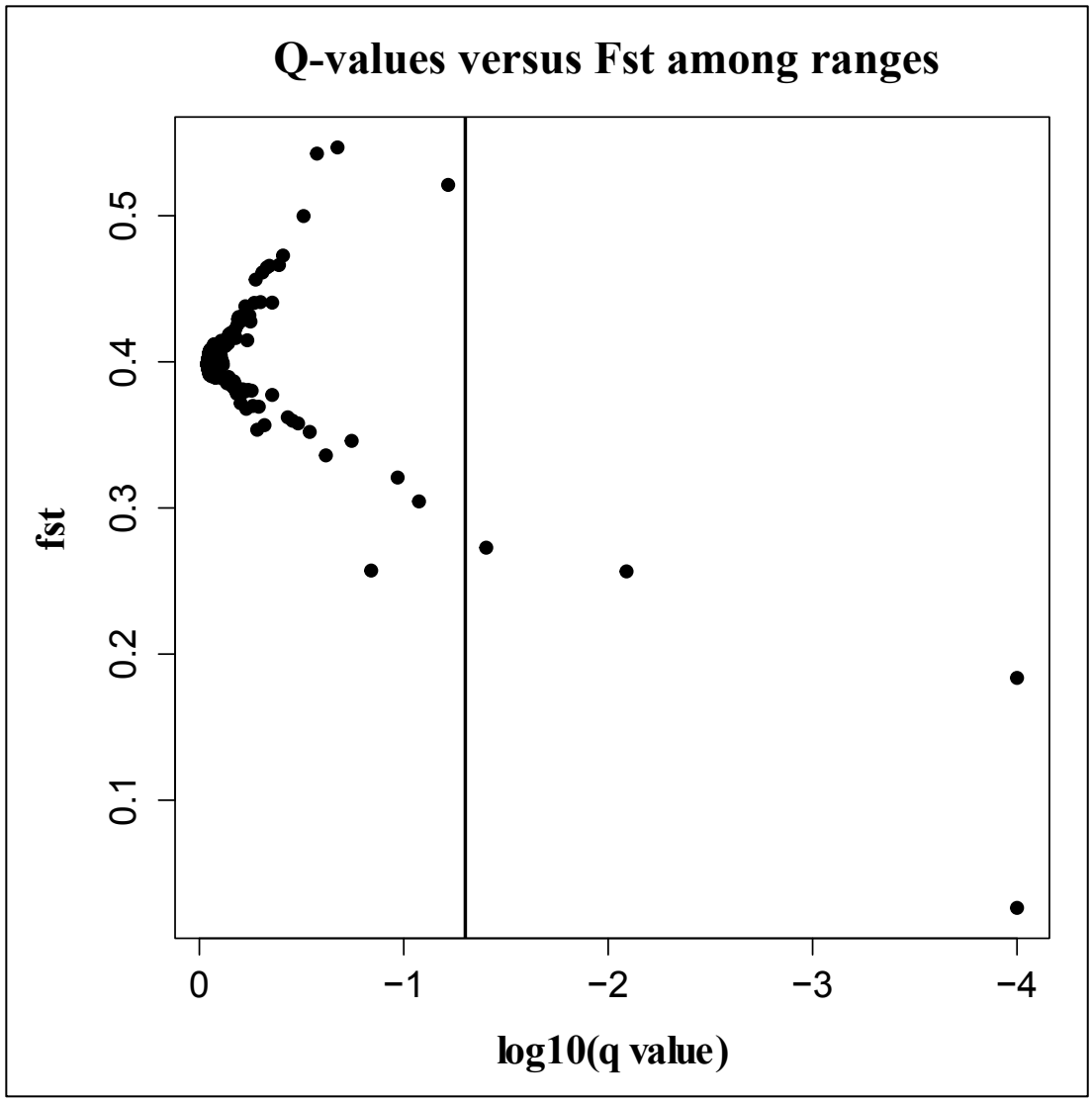


Figure 4.6. Q-values of loci versus F_{ST} in the among-range comparison between putative source populations and invasive populations. The vertical line between -1 and -2 likelihood values on the x-axis indicates the threshold of significance with a 5% False Discovery Rate. All values to the right of the vertical bar represent outlier loci.

References

Abramovitz, J. N. (1983) Pueraria lobata Willd. (Ohwi), kudzu: limitations to sexual reproduction. M.S. thesis. University of Maryland, College Park, Maryland, USA.

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11:697-709.

Allendorf, F. W. and L. L. Lundquist. (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17:24-30.

Andrews, S. (2010) FastQC. A quality control tool for high throughput sequence data. [http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc%5D.

Balloux, F., L. Lehmann, and T. de Meeûs. (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635-1644.

Barrett, R. D. H. and H. E. Hoekstra. (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* 12:767-780.

Barrett, S. C. H., R. I. Colautti, and C. G. Eckert. (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* 17:373-383.

Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* 20:2044-2072.

Blaustein, R. (2001) Kudzu's invasion into Southern United States life and culture. The Great Reshuffling: Human dimensions of invasive species:55-62.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124-3140.

Charlesworth, B., M. Morgan, and D. Charlesworth. (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289-1303.

Corley, R., A. Woldeghebriel, and M. Murphy. (1997) Evaluation of the nutritive value of kudzu (Pueraria lobata) as a feed for ruminants. *Animal feed science and technology* 68:183-188.

Crooks, J. A., M. Soulé, O. Sandlund, P. Schei, and A. Viken. (1999) Lag times in population explosions of invasive species: causes and implications. *Invasive species and biodiversity management*:103-125.

Davey, J. L. and M. W. Blaxter. (2010) RADSeq: next-generation population genetics. *Briefings in functional genomics* 9:416.

Dlugosch, K. and I. Parker. (2008a) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17:431-449.

Dlugosch, K. M. and I. M. Parker. (2008b) Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters* 11:701-709.

Ellstrand, N. C. and M. L. Roose. (1987) Patterns of genotypic diversity in clonal plant species. *American Journal of Botany*:123-131.

Excoffier, L., T. Hofer, and M. Foll. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285-298.

Excoffier, L. and N. Ray. (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* 23:347-351.

Foll, M. and O. Gaggiotti. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977.

Forseth, I. N. and A. F. Innis. (2004) Kudzu (Pueraria montana): history, physiology, and ecology combine to make a major ecosystem threat. *Critical reviews in plant sciences* 23:401-413.

García-Ramos, G. and D. Rodríguez. (2002) Evolutionary speed of species invasions. *Evolution* 56:661-668.

Gordon, A. and G. Hannon. (2010) Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) [http://hannonlab.cshl.edu/fastx_toolkit/index.html%5D.

Hamrick, J. and M. Godt. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical transactions: biological sciences* 351:1291-1298.

Hermisson, J. (2009) Who believes in whole-genome scans for selection? *Heredity* 103:283-284.

Hohenlohe, P. A., P. C. Phillips, and W. A. Cresko. (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences* 171:1059.

Huey, R. B., G. W. Gilchrist, M. L. Carlson, D. Berrigan, and L. Serra. (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science* 287:308.

Jombart, T. (2013) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1.4-0.

Jombart, T., S. Devillard, and F. Balloux. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* 11:94.

Kaplan, N. L., R. Hudson, and C. Langley. (1989) The" hitchhiking effect" revisited. *Genetics* 123:887-899.

Kolbe, J. J., R. E. Glor, L. R. Schettino, A. C. Lara, A. Larson, and J. B. Losos. (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431:177-181.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*:314-334.

Lee, C. E. (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* 17:386-391.

Lewontin, R. and J. Krakauer. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175.

Maron, J. L., M. Vil‡, R. Bommarco, S. Elmendorf, and P. Beardsley. (2004) Rapid evolution of an invasive plant. *Ecological Monographs* 74:261-280.

Meirmans, P. G. and P. H. Van Tienderen. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4:792-794.

Miller, J. H. and B. Edwards. (1983) Kudzu: Where did it come from? And how can we stop it? *Southern Journal of Applied Forestry* 7:165-169.

Mitich, L. W. (2000) Kudzu [Pueraria lobata (Willd.) Ohwi]. *Weed Technology* 14:231-235.

Moran, E. V. and J. M. Alexander. (2014) Evolutionary responses to global change: lessons from invasive species. *Ecology Letters* 17:637-649.

Morin, P. A., G. Luikart, and R. K. Wayne. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* 19:208-216.

Narum, S. R. and J. E. Hess. (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources* 11:184-194.

Nielsen, E. E., J. Hemmer-Hansen, N. A. Poulsen, V. Loeschcke, T. Moen, T. Johansen, C. Mittelholzer, G.-L. Taranger, R. Ogden, and G. R. Carvalho. (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (Gadus morhua). *BMC Evolutionary Biology* 9:276.

Peakall, R. and P. E. Smouse. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288-295.

Peakall, R. and P. E. Smouse. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* 28:2537-2539.

Pérez-Figueroa, A., M. García-Pereira, M. Saura, E. Rolán-Alvarez, and A. Caballero. (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* 23:2267-2276.

Pompanon, F. o., A. l. Bonin, E. Bellemain, and P. Taberlet. (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6:847-846.

Prentis, P. J., J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. (2008) Adaptive evolution in invasive species. *Trends in Plant Science* 13:288-294.

Richards, C. L., O. Bossdorf, N. Z. Muth, J. Gurevitch, and M. Pigliucci. (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters* 9:981-993.

Riquet, F., C. Daguin-Thiébaut, M. Ballenghien, N. Bierne, and F. Viard. (2013) Contrasting patterns of genome,Äêwide polymorphism in the native and invasive range of the marine mollusc Crepidula fornicata. *Molecular Ecology* 22:1003-1018.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, and N. C. Ellstrand. (2001) The Population Biology of Invasive Species. *Annual Review of Ecology and Systematics*:305-332.

Siemann, E. and W. E. Rogers. (2001) Genetic differences in growth of an invasive tree species. *Ecology Letters* 4:514-518.

Smith, J. M. and J. Haigh. (1974) The hitch-hiking effect of a favourable gene. *Genetical research* 23:23-35.

Storz, J. F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14:671-688.

Van der Maesen, L. (2002) Pueraria: botanical characteristics. Pages 1-28 *in* W. Keung, editor. Pueraria: The genus Pueraria. Medicinal and aromatic plants - industrial profiles. Taylor & Francis, New York, NY.

Vandepitte, K., T. Meyer, K. Helsen, K. Acker, I. Roldán-Ruiz, J. Mergeay, and O. Honnay. (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology* 23:2157-2164.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34:275-306.

Ward, D. B. (1998) Pueraria montana: the correct scientific name of the kudzu. *Castanea* 63:76-77.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

A major goal of evolutionary biology is to understand how species are able to adapt to their surroundings, particularly when environments vary (Darwin 1859). Invasive species are excellent models for addressing this goal by examining the genetic and evolutionary mechanisms that lead to rapid adaptation associated with invasion (Lee 2002). As adaptation is dependent on the introduction of beneficial genetic diversity, whether from pre-existing alleles in the native range that gained beneficial function in the new environment or new mutation (Prentis et al. 2008), characterization of the amount and distribution of genetic variation across the genome both within and among the native and invasive ranges is essential (Sakai et al. 2001). Patterns in genetic diversity can generate inferences about the amount of diversity introduced, the origin of those introduction events, gene flow and reproduction as well as generate hypotheses about possible evolutionary mechanisms during invasion (Tsutsui et al. 2000). Subsequent genome scans for selection can then be informed by population structure and expectations of what could be targets of selection or means by which alleles may have been fixed (Excoffier et al. 2009), whether due to adaptive or non-adaptive mechanisms (Storz 2005), thereby making genome scans for selection and interpretation of results more accurate (Hohenlohe et al. 2010).

This comprehensive approach was taken to understand the invasion history of kudzu, *Pueraria montana var. lobata*. The invasion of kudzu represents a highly successful invasion in that it has spread across 5 million hectares in less than 150 years (Blaustein 2001) and has abiotic and biotic effects on the local and regional habitats (Forseth and Innis 2004). To reconstruct the

invasion history, I compared patterns of genetic and clonal diversity within and among ranges to infer the amount and distribution of genetic variation as well as identify the number of introduction events and putative sources. I compared findings among multiple marker types to see if conclusions were corroborated. In both datasets with different molecular markers, I found consistent evidence of multiple introductions and that the invasive range exhibited an increase in clonal reproduction as compared to the more sexually reproducing putative sources as well as the rest of the native range. However, microsatellites demonstrated much higher heterozygosity levels compared to SNPs, calling into question the true nature of gene diversity in the genome and the invasion. Lastly, I conducted a genome scan for selection after accounting for population structure and introduction history. I identified outlier loci between invasive populations and putative sources as well as outlier loci between the putative and non-putative sources to the invasion. Thirty-one outlier loci detected did not show similarity to known functional sequences and could represent loci linked to adaptive loci if population differentiation was not due to non-adaptive processes such as genetic drift. Outlier loci represent candidate targets of selection and require further study to determine whether population differentiation at these loci are artifacts of stochastic or demographic processes or actual cases of natural selection (Lande 1976, Barrett and Hoekstra 2011).

Now that next-generation sequencing technologies and methodologies are available to non-model species without previous genetic resources (Allendorf et al. 2010), it can be expected that more studies investigating the role of genetic variation and natural selection will be conducted in the near future. This study of kudzu's invasion history in addition to the plethora of others that are hopefully generated quickly will enable scientists to characterize the frequency of natural selection, the common targets of selection in phenotypic trait evolution, whether natural

selection more commonly acts on standing versus novel genetic variation, and the time to adaptation after beneficial alleles arise in a new environment (Moran and Alexander 2014). The continuous generation of big data will facilitate our understanding of widespread patterns of adaptation across space, time and changing environments to hopefully reach consensus about natural selection's role during invasion.

References

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* 11:697-709.

Barrett, R. D. H. and H. E. Hoekstra. (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* 12:767-780.

Blaustein, R. (2001) Kudzu's invasion into Southern United States life and culture. *The Great Reshuffling: Human dimensions of invasive species*:55-62.

Darwin, C. (1859) The Origins of Species by Means of Natural Selection, Or the Preservation of Favoured Races in the Struggle for Life. John Murray, London, England.

Excoffier, L., T. Hofer, and M. Foll. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285-298.

Forseth, I. N. and A. F. Innis. (2004) Kudzu (*Pueraria montana*): history, physiology, and ecology combine to make a major ecosystem threat. *Critical reviews in plant sciences* 23:401-413.

Hohenlohe, P. A., P. C. Phillips, and W. A. Cresko. (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences* 171:1059.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*:314-334.

Lee, C. E. (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* 17:386-391.

Moran, E. V. and J. M. Alexander. (2014) Evolutionary responses to global change: lessons from invasive species. *Ecology Letters* 17:637-649.

Prentis, P. J., J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. (2008) Adaptive evolution in invasive species. *Trends in Plant Science* 13:288-294.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, and N. C. Ellstrand. (2001) The Population Biology of Invasive Species. *Annual Review of Ecology and Systematics*:305-332.

Storz, J. F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14:671-688.

Tsutsui, N. D., A. V. Suarez, D. A. Holway, and T. J. Case. (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America* 97:5948.

APPENDICES

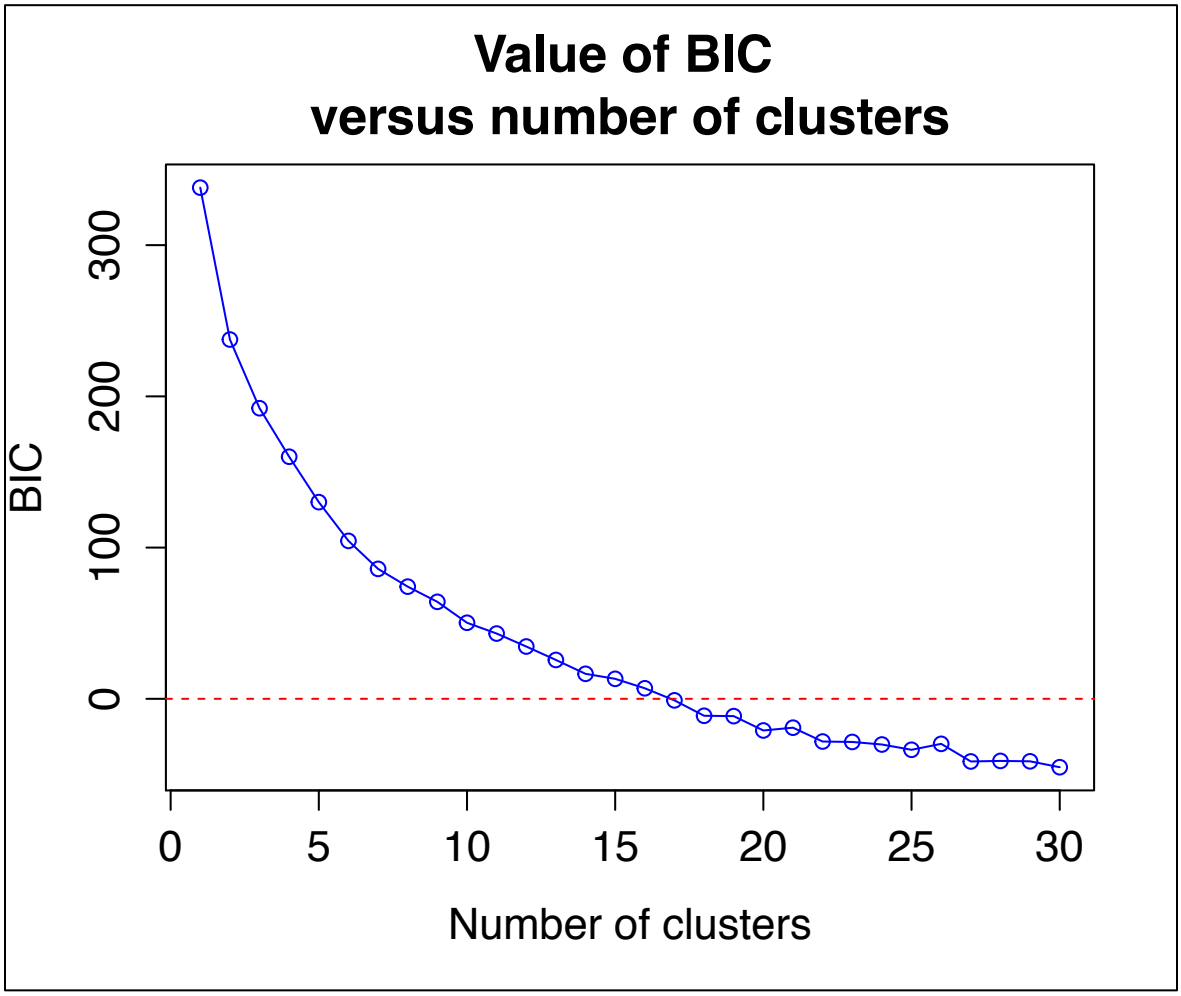
Appendix 2.1. Genetic diversity measures for the dataset representing the genet-only dataset, including GPS coordinates for all populations in this study. A_R, allelic richness; F, fixation index; H_O, observed heterozygosity; N, sample size; N_E, average effective number of alleles; MLG, distinct multilocus genotype; R, "proportion distinguishable"; uH_E, Nei's unbiased expected heterozygosity. E/I designates whether or not this population was considered at the "Interior" or "Edge" of the range.

Population	**Latitude**	**Longitude**	**E/I**	**N**	**MLG**	**R**	**A_R**	**N_E**	**H_O**	**uH_E**	**F**
AL10	31.512783	-86.702148	I	7	2	0.167	1.46	1.64	0.633	0.456	-0.933
AL11	31.222782	-85.394867	I	23	2	0.045	1.34	1.51	0.500	0.344	-0.917
AL12	33.17025	-87.53775	I	20	10	0.474	1.54	2.47	0.533	0.540	-0.057
AL13	32.584816	-88.185951	I	15	4	0.214	1.33	1.55	0.489	0.333	-0.704
AL14	33.648048	-85.426285	I	19	2	0.056	1.32	1.46	0.300	0.322	-0.489
AL3	33.524685	-86.825119	I	22	9	0.381	1.31	1.60	0.354	0.313	-0.173
AL5	33.610401	-85.76815	I	22	14	0.619	1.37	1.79	0.438	0.365	-0.271
AL6	32.454117	-87.944	I	15	9	0.571	1.38	1.63	0.412	0.376	-0.166
AL7	32.431782	-87.037987	I	22	2	0.048	1.34	1.51	0.500	0.344	-0.917
AL9	32.328934	-86.341347	I	24	3	0.087	1.31	1.51	0.489	0.311	-0.875
AR1	35.81686	-92.550194	E	23	3	0.091	1.41	1.69	0.644	0.409	-0.914
AR2	34.066254	-93.689255	E	25	4	0.125	1.52	2.13	0.533	0.524	-0.169
FL1	30.958817	-86.296432	I	23	2	0.045	1.30	1.44	0.433	0.300	-0.905
FL2	30.782818	-85.37545	I	22	1	0.000	1.27	1.27	0.267	0.267	-1.000
FL3	28.560625	-81.02325	I	18	2	0.059	1.26	1.37	0.367	0.256	-0.889
GA10	33.62965	-84.451385	I	24	3	0.087	1.16	1.22	0.156	0.156	-0.180
GA14	34.687099	-83.422363	I	18	1	0.000	1.33	1.33	0.333	0.333	-1.000
GA17	34.446316	-83.119232	I	26	6	0.200	1.47	1.94	0.547	0.474	-0.247
GA18	33.921417	-83.406235	I	26	1	0.000	1.47	1.47	0.467	0.467	-1.000
GA2	33.169716	-83.2808	I	19	10	0.500	1.57	2.64	0.464	0.566	0.119

GA3	32.329033	-81.933052	I	22	5	0.190	1.44	1.88	0.470	0.441	-0.030
GA32	30.896784	-83.969666	I	24	1	0.000	1.53	1.53	0.533	0.533	-1.000
GA33	31.586634	-84.151535	I	23	4	0.136	1.29	1.48	0.467	0.288	-0.833
GA34	32.077068	-83.763084	I	24	1	0.000	1.47	1.47	0.467	0.467	-1.000
GA35	32.8517	-83.619781	I	25	1	0.000	1.27	1.27	0.267	0.267	-1.000
GA36	33.292648	-83.376297	I	24	10	0.391	1.53	2.54	0.423	0.531	0.123
GA37	34.666569	-84.94297	I	21	11	0.500	1.53	2.41	0.521	0.526	-0.049
GA39	33.94738333	-83.3801333	I	24	1	0.000	1.40	1.40	0.400	0.400	-1.000
GA4	32.530434	-82.901436	I	13	2	0.083	1.62	2.02	0.533	0.622	-0.162
GA6	32.78175	-84.234581	I	16	3	0.133	1.55	2.09	0.522	0.547	-0.116
GA7	33.528767	-82.034782	I	21	1	0.000	1.27	1.27	0.267	0.267	-1.000
GA9	33.199219	-83.437714	I	23	4	0.136	1.32	1.53	0.450	0.319	-0.559
GA96	34.413933	-84.109383	I	25	11	0.417	1.52	2.30	0.555	0.521	-0.151
KY10	37.343876	-84.313889	I	6	3	0.400	1.44	1.75	0.489	0.440	-0.561
KY11	37.179298	-83.596069	I	22	5	0.190	1.47	1.90	0.521	0.471	-0.292
KY7	38.268234	-85.502586	E	19	1	0.000	1.20	1.20	0.200	0.200	-1.000
KY8	37.38015	-82.432785	I	24	8	0.304	1.35	1.62	0.339	0.351	0.043
KY9	37.446583	-82.906631	I	25	2	0.042	1.48	1.74	0.400	0.478	-0.127
MD1	38.940334	-77.121368	E	23	5	0.182	1.43	1.97	0.460	0.430	-0.186
MO1	37.234722	-89.521111	E	18	2	0.059	1.32	1.47	0.333	0.322	-0.367
MO2	37.318611	-89.527222	E	9	1	0.000	1.47	1.47	0.467	0.467	-1.000
MS2	33.730831	-90.563904	I	22	1	0.000	1.27	1.27	0.267	0.267	-1.000
MS3	32.341217	-89.136818	I	23	4	0.136	1.34	1.59	0.533	0.335	-0.807
MS4	31.967251	-89.859482	I	13	5	0.333	1.35	1.64	0.497	0.347	-0.555
MS5	31.352716	-89.342384	I	8	3	0.286	1.43	1.68	0.667	0.427	-0.882
MS6	32.386185	-88.621902	I	25	5	0.167	1.29	1.52	0.453	0.287	-0.726
NC1	34.345032	-78.708649	I	21	1	0.000	1.13	1.13	0.133	0.133	-1.000
NC13	36.341019	-78.419914	I	24	13	0.522	1.54	2.39	0.512	0.540	0.020
NC15	35.966133	-78.847336	I	26	2	0.040	1.17	1.24	0.233	0.167	-0.833
NC17	35.426731	-82.528015	I	18	2	0.059	1.40	1.62	0.567	0.400	-0.881
NC19	34.205597	-77.922783	I	17	1	0.000	1.53	1.53	0.533	0.533	-1.000
NC20	36.39465	-79.698265	I	33	15	0.438	1.51	2.25	0.502	0.509	-0.002
NC21	34.212399	-77.835297	I	10	1	0.000	1.20	1.20	0.200	0.200	-1.000
NC3	35.375084	-83.221382	I	19	9	0.444	1.56	2.67	0.317	0.557	0.340
NC4	35.619469	-82.173416	I	16	8	0.467	1.48	2.21	0.333	0.476	0.200

NC5	35.719601	-81.153816	I	21	14	0.650	1.52	2.66	0.520	0.518	-0.066
NC6	35.650284	-80.469467	I	22	3	0.095	1.52	2.07	0.489	0.524	-0.097
NC7	36.062199	-79.856621	I	22	5	0.190	1.39	1.81	0.476	0.394	-0.320
NE1	40.595171	-95.787825	E	8	1	0.000	1.33	1.33	0.333	0.333	-1.000
NY1	40.987316	-73.805351	E	23	1	0.000	1.40	1.40	0.400	0.400	-1.000
NY4	40.906551	-73.584579	E	15	5	0.286	1.34	1.68	0.520	0.344	-0.669
OK1	36.306821	-95.595824	E	24	1	0.000	1.40	1.40	0.400	0.400	-1.000
MD2	38.99283333	-76.991667	E	22	4	0.143	1.46	2.02	0.500	0.462	-0.223
SC1	33.933933	-80.97023	I	18	2	0.059	1.30	1.44	0.433	0.300	-0.905
SC4	35.078365	-81.717468	I	22	6	0.238	1.50	2.04	0.502	0.497	-0.109
SC6	34.855518	-82.26638	I	11	8	0.700	1.54	2.66	0.479	0.541	0.025
SC8	34.255383	-79.699432	I	24	2	0.043	1.21	1.31	0.300	0.211	-0.867
TN13	34.98735	-84.370117	I	24	11	0.435	1.43	1.80	0.565	0.427	-0.375
TN14	35.527134	-84.348236	I	24	4	0.130	1.34	1.56	0.550	0.336	-0.789
TN16	36.395649	-89.339897	I	25	1	0.000	1.47	1.47	0.467	0.467	-1.000
TN2	35.026348	-85.312897	I	7	2	0.167	1.43	1.51	0.500	0.433	-0.917
TN3	36.161068	-86.794365	I	14	2	0.077	1.34	1.51	0.500	0.344	-0.917
TN6	35.095901	-90.076668	E	17	2	0.063	1.53	2.01	0.567	0.533	-0.430
TN8	36.249367	-84.18145	I	25	1	0.000	1.67	1.67	0.667	0.667	-1.000
TN9	35.051533	-84.900452	I	17	4	0.188	1.26	1.46	0.417	0.264	-0.800
TX2	30.704468	-94.934166	E	18	2	0.059	1.40	1.55	0.533	0.400	-0.852
TX3	32.583782	-94.342934	E	24	2	0.043	1.39	1.57	0.567	0.389	-0.926
TX4	30.924683	-94.007835	E	20	4	0.158	1.37	1.57	0.450	0.374	-0.285
TX5	30.049217	-94.881516	E	25	5	0.167	1.44	1.99	0.460	0.437	-0.151
TX6	31.420532	-94.723633	E	20	1	0.000	1.27	1.27	0.267	0.267	-1.000
VA1	37.38335	-77.414047	I	23	5	0.182	1.53	2.04	0.600	0.526	-0.258
VA10	37.147148	-79.23037	I	25	1	0.000	1.27	1.27	0.267	0.267	-1.000
VA11	36.73859	-76.574051	I	24	2	0.043	1.30	1.44	0.433	0.300	-0.905
VA4	36.990284	-80.790535	I	18	5	0.235	1.50	2.06	0.573	0.501	-0.256
VA8	37.198532	-81.392899	I	14	1	0.000	2.67	1.20	0.267	0.267	-1.000
VA9	37.168633	-81.898033	I	9	2	0.125	1.34	1.51	0.500	0.344	-0.917
WV1	38.023884	-81.354553	I	17	2	0.063	1.60	2.13	0.433	0.600	0.036
Mean				**20**	**4**	**0.165**	**1.41**	**1.70**	**0.444**	**0.398**	**-0.575**

Appendix 2.2. DAPC plot of the regression between Bayesian Information Criterion (BIC) values and number of clusters.



Appendix 2.3. Clonal diversity estimates of all populations. div, Nei's genetic diversity index; eff, the effective number of MLGs based on rarefaction; eve, evenness; Ia, index of association; rbarD, the standardized index of association; shc, the Shannon-Weiner index corrected for sample size; N/A, not applicable.

Population	eff	div	shc	eve	Ia	rbarD
AL10	1.32	0.286	0.285	0.662	0.000	N/A
AL11	1.29	0.237	0.173	0.647	0.000	N/A
AL12	6.06	0.879	1.080	0.606	3.94	0.390
AL13	2.92	0.705	0.572	0.731	0.126	0.065

AL14	1.11	0.105	0.147	0.555	1.49	0.829
AL3	5.04	0.840	0.968	0.560	0.339	0.078
AL5	10.5	0.948	1.352	0.752	1.865	0.257
AL6	3.95	0.800	1.141	0.439	2.759	0.213
AL7	1.10	0.091	0.132	0.548	0.000	N/A
AL9	1.19	0.163	0.242	0.395	-0.043	-0.043
AR1	1.31	0.245	0.263	0.435	-0.060	-0.062
AR2	2.00	0.520	0.443	0.499	9.213	0.872
FL1	1.73	0.443	0.267	0.867	0.000	N/A
FL2	1.00	0.000	0.000	1.000	-0.075	-0.085
FL3	1.12	0.111	0.152	0.559	0.000	N/A
GA10	1.29	0.236	0.255	0.430	1.045	0.296
GA14	1.00	0.000	0.000	1.000	N/A	N/A
GA17	2.43	0.612	0.634	0.405	4.634	0.499
GA18	1.00	0.000	0.000	1.000	0.000	N/A
GA2	4.57	0.825	1.094	0.457	3.527	0.288
GA3	2.22	0.576	0.564	0.444	6.670	0.664
GA32	1.00	0.000	0.000	1.000	0.000	N/A
GA33	2.14	0.557	0.480	0.535	0.019	0.019
GA34	1.00	0.000	0.000	1.000	-0.073	-0.027
GA35	1.00	0.000	0.000	1.000	0.000	N/A
GA36	3.51	0.746	0.975	0.351	3.888	0.361
GA37	6.04	0.876	1.137	0.549	4.431	0.363
GA39	1.00	0.000	0.000	1.000	N/A	N/A
GA4	1.17	0.154	0.192	0.583	7.374	0.766
GA6	1.47	0.342	0.333	0.490	8.028	0.760
GA7	1.00	0.000	0.000	1.000	N/A	N/A
GA9	1.44	0.320	0.379	0.360	0.238	0.053
GA96	4.77	0.823	1.053	0.434	4.325	0.383
KY10	2.57	0.733	0.546	0.857	3.037	0.571
KY11	1.48	0.338	0.493	0.295	2.935	0.289
KY7	1.00	0.000	0.000	1.000	0.000	N/A
KY8	1.95	0.507	0.767	0.243	1.963	0.179
KY9	1.08	0.080	0.120	0.542	4.169	0.429
MD1	2.61	0.644	0.594	0.521	5.719	0.709

MD2	1.33	0.260	0.378	0.332	6.774	0.752
MO1	1.12	0.111	0.152	0.559	3.800	0.847
MO2	1.00	0.000	0.000	1.000	N/A	N/A
MS2	1.00	0.000	0.000	1.000	N/A	N/A
MS3	1.58	0.383	0.404	0.395	0.854	0.432
MS4	3.93	0.808	0.723	0.786	2.345	0.356
MS5	2.91	0.750	0.494	0.970	0.042	0.021
MS6	2.62	0.643	0.583	0.523	0.012	0.007
NC1	1.00	0.000	0.000	1.000	N/A	N/A
NC13	6.00	0.870	1.219	0.462	3.419	0.298
NC15	1.08	0.077	0.116	0.540	0.000	N/A
NC17	1.12	0.111	0.152	0.559	0.428	0.218
NC19	1.00	0.000	0.000	1.000	N/A	N/A
NC20	5.70	0.850	1.186	0.380	4.451	0.360
NC21	1.00	0.000	0.000	1.000	N/A	N/A
NC3	2.80	0.678	0.991	0.311	4.060	0.366
NC4	6.40	0.900	0.996	0.800	4.554	0.401
NC5	5.73	0.867	1.413	0.409	4.395	0.368
NC6	1.45	0.325	0.300	0.483	9.136	0.843
NC7	3.14	0.714	0.630	0.629	4.887	0.680
NE1	1.00	0.000	0.000	1.000	N/A	N/A
NY1	1.00	0.000	0.000	1.000	0.000	N/A
NY4	2.10	0.562	0.636	0.421	2.957	0.594
OK1	1.00	0.000	0.000	1.000	N/A	N/A
SC1	1.12	0.111	0.152	0.559	0.000	N/A
SC4	2.88	0.684	0.692	0.480	3.909	0.370
SC6	5.26	0.891	1.256	0.658	2.906	0.239
SC8	1.18	0.159	0.137	0.590	0.000	N/A
TN13	4.30	0.801	1.063	0.391	4.728	0.407
TN14	2.01	0.525	0.456	0.503	0.079	0.041
TN16	1.00	0.000	0.000	1.000	N/A	N/A
TN2	1.32	0.286	0.285	0.662	-0.250	-0.258
TN3	1.32	0.264	0.194	0.662	-0.008	-0.008
TN6	1.12	0.118	0.159	0.562	5.296	0.592
TN8	1.00	0.000	0.000	1.000	0.000	N/A

TN9	1.45	0.331	0.450	0.363	-0.136	-0.073
TX2	1.25	0.209	0.166	0.623	0.639	0.665
TX3	1.09	0.083	0.123	0.543	0.000	N/A
TX4	2.11	0.553	0.482	0.526	5.443	0.702
TX5	3.42	0.737	0.633	0.683	5.390	0.609
TX6	1.00	0.000	0.000	1.000	N/A	N/A
VA1	2.61	0.644	0.594	0.521	8.764	0.852
VA10	1.00	0.000	0.000	1.000	N/A	N/A
VA11	1.09	0.083	0.123	0.543	0.000	N/A
VA4	2.08	0.549	0.592	0.415	6.388	0.720
VA8	1.00	0.000	0.000	1.000	0.000	N/A
VA9	1.25	0.222	0.244	0.623	0.000	N/A
WV1	1.99	0.529	0.300	0.997	7.099	0.678
Mean	**2.20**	**0.361**	**0.406**	**0.658**	**N/A**	**N/A**

Appendix 2.4. Description of chloroplast haplotypes for the rpoB-trnC chloroplast region. Each mutation is in comparison to Haplotype 1. bp, the base pair position; N, the number of individuals belonging to that haplotype.

Haplotype	**N**	**Mutation**
Haplotype 1	171	N/A
Haplotype 2	27	TT for AA at position 549 bp
Haplotype 3	25	G for T, at position 346 bp
Haplotype 4	12	C for A at position 126 bp
Haplotype 5	11	7bp deletion at positions 403-410 bp
Haplotype 6	11	Single deletion at position 737 bp
Haplotype 7	4	T for G at position 333 bp
Haplotype 8	3	C for T at position 517 bp; TCTCC insertion at position 519 bp; CC for AA at position 524 bp
Haplotype 9	2	T for C at position 27 bp
Haplotype 10	2	G for T at position 138 bp

Haplotype 11	1	G insertion at position 634 bp
Haplotype 12	1	T for A at position 759 bp
Haplotype 13	1	G for A at position 584 bp
Total	**271**	

Appendix 2.5. Neighbor-joining population tree with bootstrap values, based on Nei's genetic distance calculated from microsatellite gene frequencies. Nodes with bootstrap values below 50% were collapsed.