



Population genomics of the invasive *Anoplophora glabripennis* for the purpose of biosurveillance

Thèse

Mingming Cui

Doctorat en sciences forestières
Philosophiæ doctor (Ph. D.)

Québec, Canada

Population genomics of the invasive *Anoplophora glabripennis* for the purpose of biosurveillance

Thèse

Mingming Cui

Sous la direction de:

Ilga Porth, directrice de recherche
Richard Hamelin, codirecteur de recherche

Résumé

Le longicorne asiatique, *Anoplophora glabripennis*, originaire de Chine et de la péninsule coréenne, est devenu un insecte nuisible qui représente une menace significative pour les forêts et les économies du monde entier. La gestion efficace de ses invasions et de ses dommages potentiels repose sur une détection précoce grâce aux méthodes de biosurveillance. Dans cette thèse, j'utilise des techniques de séquençage de nouvelle génération (NGS) et des méthodes de génomique des populations pour 1) caractériser la similarité et la différenciation des populations de ce coléoptère, 2) identifier les signatures génomiques qui déterminent la différenciation entre les lignées génétiques et les signatures génomiques associées à l'adaptation environnementale (*i.e.* variables climatiques), en particulier l'adaptation au froid, 3) éclairer l'histoire de son invasion en Amérique du Nord.

Dans le premier chapitre, j'étudie la structure de population des ALB indigènes de l'Asie pour informer la conception des outils de biosurveillance afin d'identifier les sources d'invasion. En utilisant des individus de populations indigènes, en séquençant leurs génomes grâce au séquençage par génotypage (GBS), et en identifiant des marqueurs de polymorphisme nucléotidique (SNP) à l'échelle du génome, j'ai comparé la diversité génétique, mesuré le flux de gènes, et effectué des tests d'assignation de population. J'ai identifié six groupes génétiques avec une division claire entre les groupes du nord et du sud de l'aire de répartition native de cette espèce. Nos résultats démontrent qu'un petit nombre de SNPs peut assigner précisément des individus à des régions géographiques, jetant les bases pour de nouveaux outils de biosurveillance pour l'ALB.

Dans le deuxième chapitre, j'examine les signatures génomiques sous sélection associées aux variables climatiques, en particulier la température. J'ai effectué un séquençage complet du génome sur dix échantillons multiplexés, chacun avec 20 individus, pour produire une forte densité de marqueurs SNP à l'échelle du génome. Le criblage du génome et l'analyse gène-environnement (GEA) ont été utilisés pour identifier plus de 5 000 gènes potentiellement impliqués dans l'adaptation locale chez l'ALB. Alors que des gènes communs de tolérance au froid, tels que la glycérol kinase, les cytochromes P450, la protéine antigel et les protéines de choc thermique ont été trouvés à travers des analyses de sélection, ceux-ci n'étaient pas significatifs dans l'analyse GEA, indiquant une relation complexe entre les facteurs environ-

nementaux et l'évolution génétique de la tolérance au froid. Ce profilage génomique complet offre de nouvelles perspectives sur les mécanismes génétiques sous-jacents à l'adaptation locale chez cette espèce.

Dans le troisième chapitre, j'utilise des données génomiques obtenues grâce à la technologie GBS pour reconstruire l'histoire invasive du longicorne asiatique en Amérique du Nord (NA) et mieux comprendre les invasions biologiques. Les résultats révèlent que la plupart des populations invasives de NA proviennent de plusieurs introductions indépendantes du nord de la Chine. Les origines spécifiques incluent la Région Nord Deux pour les populations de l'Illinois, de Toronto et de l'Ohio, la Région Sud pour les populations du Massachusetts, et la Région Nord Un pour les populations de Farmingdale, New York. Après leur introduction, toutes les populations ont connu des goulots d'étranglement, certaines connaissant également une seconde expansion, comme New York. Les résultats de notre étude sont cohérents avec les données historiques, offrant de nouvelles perspectives sur la dynamique d'invasion de cette espèce.

En conclusion, cette thèse présente une étude approfondie sur la structure populationnelle, l'adaptation et l'histoire invasive du longicorne asiatique en utilisant des techniques génomiques avancées. Nos découvertes améliorent non seulement notre compréhension de la biologie de ce coléoptère et de ces mécanismes adaptatifs, mais fournissent également des perspectives précieuses pour le développement de stratégies de biosurveillance efficaces pour gérer et atténuer les impacts de ses invasions. Alors que la menace des espèces invasives continue d'augmenter à l'échelle mondiale, les méthodes et les connaissances acquises grâce à cette étude peuvent être appliquées à d'autres nuisibles invasifs, contribuant finalement à la protection de nos forêts, écosystèmes et économies.

Abstract

The Asian longhorned beetle (ALB) *Anoplophora glabripennis*, native to China and Korea Peninsula, has become a global insect pest that poses a significant threat to forests and economies worldwide. Effective management of its invasions and potential damage relies on early detection through biosurveillance methods. In this thesis, I utilize next-generation sequencing (NGS) techniques and population genomics methods to 1) characterize the beetle's population similarity and differentiation, 2) identify genomic signatures that determine the differentiation between genetic lineages in its native range and signatures related to environmental adaptation associated with climate variables especially cold adaptation, and 3) provide insight on its invasion history in North America.

In the first chapter, I investigate the population structure of native ALB populations to inform the design of biosurveillance tools for identifying invasion sources. By collecting native population samples, sequencing their genomes using genotyping-by-sequencing (GBS), and obtaining genome-wide single nucleotide polymorphism (SNP) markers, I compared genetic diversity, measured gene flow, and conducted population assignment tests. I identified six genetic clusters with a clear division between northern and southern groups. Our results demonstrate that a small number of SNPs can accurately assign individuals to geographic regions, laying the foundation for new ALB biosurveillance tools.

In the second chapter, I examine genomic signatures under selection and associated with climate variables, specifically temperature. I performed whole genome sequencing on ten pooled samples, each with 20 individuals, to produce high-density genome-wide SNP markers. Genome scans and gene-environment analysis (GEA) were used to identify over 5,000 genes potentially involved in local adaptation in ALB. Notably, while common cold tolerance genes, such as glycerol kinase, cytochrome P450s, antifreeze protein, and heat shock proteins, were found through selection scans, these were not significant in GEA analysis, indicating a complex relationship between environmental factors and genetic evolution of cold tolerance. This comprehensive genomic profiling provides new insights into the genetic mechanisms underlying local adaptation in this species.

In the third chapter, I use genomic data obtained through GBS technology to reconstruct the invasion history of the Asian longhorned beetle in North America (NA) to better understand

the biological invasions in the invasive region. Findings reveal that most NA invasive populations originated from multiple independent introductions from northern China. Specific origins include North Region Two for populations in Illinois, Toronto, and Ohio, South Region for Massachusetts populations, and North Region One for Farmingdale, New York populations. Post introduction, all populations experienced bottleneck events, with some also experiencing secondary spread, such as New York. The results of our study are consistent with historical records, offering further insights into the invasion dynamics of this species.

In conclusion, this thesis presents a detailed study of the Asian longhorned beetle's population structure, adaptation, and invasion history using advanced genomic techniques. Our findings not only enhance our understanding of the beetle's biology and adaptive mechanisms but also provide valuable insights for developing effective biosurveillance strategies to manage and mitigate the impacts of its invasions. As the threat of invasive species continues to rise globally, the methods and knowledge gained from this study can be applied to other invasive pests, ultimately contributing to the protection of our forests, ecosystems, and economies.

Contents

Résumé	ii
Abstract	iv
Contents	vi
List of Tables	viii
List of Figures	ix
Abbreviations	xi
Acknowledgements	xv
Preface	xvii
Introduction	1
1 Genome-scale phylogeography resolves the native population structure of the Asian longhorned beetle, <i>Anoplophora glabripennis</i> (Motschulsky)	17
1.1 Résumé	17
1.2 Abstract	18
1.3 Introduction	18
1.4 Materials and Methods	21
1.5 Results	29
1.6 Discussion	38
1.7 Conclusion	44
1.8 Data availability statement	44
1.9 Supplementary materials	44
1.10 Acknowledgements	54
1.11 References	55
2 Pool-Seq reveals genes selective for cold tolerance across Asian longhorned beetle's native range	66
2.1 Résumé	66
2.2 Abstract	67
2.3 Introduction	67
2.4 Materials and Methods	69
2.5 Results	73

2.6	Discussion	80
2.7	Conclusion	84
2.8	Supplementary materials	84
2.9	References	91
3	Tracking the Asian longhorned beetle's invasion history within North America with genomic evidence	100
3.1	Résumé	100
3.2	Abstract	101
3.3	Introduction	101
3.4	Materials and Methods	103
3.5	Results	108
3.6	Discussion	115
3.7	Conclusion	120
3.8	Supplementary materials	120
3.9	References	130
	Conclusion	142
	Bibliography	145

List of Tables

1	Life cycle and characteristics of <i>A. glabripennis</i>	10
1.1	Asian longhorned beetle sampling and basic population genetic analyses using 6,102 SNPs, 16 locations in China and four locations in Korea.	30
2.1	Variant detection in ALB using two different SNP callers.	74
2.2	Population genetics parameters from NPStats for the 10 ALB populations. . . .	75
2.3	39 shared genes identified with the two GEA analyses BayPASS and LFMM. .	79
3.1	Sampling location and genetic diversity of <i>A. glabripennis</i> in North America. .	104

List of Figures

1	Biological invasion and global trade over time	2
2	Approaches to identify adaptive genetic variation.	9
3	ALB infestations throughout its invasive range	13
4	Flowchart of research	16
1.1	Larva and adult of <i>Anoplophora glabripennis</i>	19
1.2	Asian longhorned beetle sampling and PCA.....	23
1.3	Spatial PCA, DAPC and Admixture.....	32
1.4	Maximum likelihood phylogenetic tree for Chinese Asian longhorned beetles analyzed in RAxML.....	33
1.5	Migration rates and population history model.	34
1.6	Exploring the potential functional relevance of a missense mutation under positive selection ($F_{ST} = 0.335$) within the <i>Anoplophora glabripennis</i> glycerol kinase gene <i>AGLA000593</i>	36
1.7	Discriminant analysis of principal components (DAPC) of the Asian longhorned beetle using microsatellites.....	37
1.8	Prediction accuracy of Asian longhorned beetle individual assignment with an increasing number of SNPs selected (ranging from 20 to 500).....	38
S1.1	Nine demographic models tested in Migrate.	45
S1.2	Grouping map	45
S1.3	Observed heterozygosity plot for ALB individuals.....	46
S1.4	IBD and Relationships between heterozygosity and latitude/longitude.....	47
S1.5	Relationships between heterozygosity and latitude/longitude without CIX	48
S1.6	PCA analysis of the ALB populations in China.	49
S1.7	A Bayesian Information Criterion (BIC) for the SNP dataset used in the DAPC analysis.....	50
S1.8	Trace file visualization of BayesAss results.	50
S1.9	Heterozygosity (H_T) - F_{ST} relationship with smoothed quantiles in <i>fsthet</i>	51
S1.10	The sPCA analysis for the microsatellite dataset.....	52
S1.11	Assignment accuracy for each population assignment class in Mycorrhiza	53
S1.12	Population genetic analyses of native ALBs in China and Korea	54
2.1	Sampling map and phylogenetic relationship of the 10 pools.....	70
2.2	Correlations between genetic (Gen), geographical (Geo), and environmental (MTCM: mean temperature of the coldest month; ISO: isothermality) matrices.	76
2.3	SNP outliers identified via genome-wide selection scan analysis (GWSS) and gene-environment analysis (GEA).....	77

2.4	Venn diagram showing shared numbers and percentages of genes found associated with MTCM and ISO using LFMM and BayPASS.....	78
S2.1	Screeplot of LFMM analysis showing the percentage of variance explained by each component in a principal component analysis.	85
S2.2	Annotation values distribution of all SNPs called in GATK.	86
S2.3	Allele frequency distribution of shared SNPs between the two SNP callers....	87
S2.4	Poolseq and individual-genotyping comparison.....	88
S2.5	GO enrichment analysis of all the genes putatively under selection.	89
S2.6	KEGG enrichment analysis of all genes identified as under selection.	90
S2.7	GO enrichment analysis of the 87 shared genes.	91
3.1	Sampling map	105
3.2	Genetic diversity of ALB in native and invasive ranges.....	108
3.3	Population structure of global ALBs.....	110
3.4	Population assignment in DAPC.....	111
3.5	Maximum likelihood phylogenetic tree	112
3.6	Invasion history of North American ALB populations.....	113
3.7	Gene flow between and within ranges.....	115
S3.1	Schematic representation of DIYABC analysis and model selection.....	122
S3.2	Ten replicate runs in BayesAss.	123
S3.3	Principal component analysis for native ALBs.	124
S3.4	Principal component analysis for global ALBs.	125
S3.5	Principal component analysis for global ALBs without Toronto samples.....	126
S3.6	Principal component analysis for each invasive sampling locality projecting onto the native ALBs.	127
S3.7	Admixture analysis.	127
S3.8	DAPC assignment plot for each invasive population.	128
S3.9	PCA for model checking of each step (1-8) in the DIYABC.....	129

Abbreviations

ALB: Asian longhorned beetle

Biosurveillance: Biosecurity surveillance

BF: Bayes factor

CP: Cryoprotectant molecules

COI: Mitochondrial cytochrome oxidase subunit I

DAPC: Discriminant analysis of principal component

ddRADseq: Double digest restriction-site associated DNA sequencing

FAO: Food and Agriculture Organization

GBS: Genotyping-by-sequencing

GDP: Gross domestic product

GEA: Gene-environment association analysis

GO: Gene Ontology

GWAS: Genome-wide association studies

GWSS: Genome-wide selection scans

HWE: Hardy-Weinberg equilibrium

IAS: Invasive alien species

IBD: Isolation-By-Distance

IBE: Isolation-By-Environment

IBIS: Institut de Biologie Intégrative et des Systèmes

ISO: Isothermality

ISPMs: International Standards for Phytosanitary Measures
IUCN: International Union for Conservation of Nature
KEGG: Kyoto Encyclopedia of Genes and Genomes
LAMP: Loop-mediated isothermal amplification
LD: Linkage disequilibrium
LFMM: Latent factor mixed models
ML: Maximum likelihood
MTCM: Mean temperature of the coldest month
mtDNA: Mitochondrial DNA
NGS: Next-generation sequencing
PCA: Principal component analysis
Pool-Seq: Pooled whole genome sequencing
QTL: Quantitative trait locus
RAPD: Random amplified polymorphic DNA
RRLs: Reduced representation genomic libraries
SCP: Supercooling point
SCAR: Sequence Characterized Amplified Region
SFS: Site frequency spectrum
SNP: Single nucleotide polymorphism
WGS: Whole genome sequencing
WPM: Wood packing material

To all my grandparents

All science begins with a leap of intuition, but we can only ever find objective truths by knowing when to let evidence take over from emotion.

Richard Dawkins

Acknowledgements

During the journey of completing the thesis, I have been very lucky to receive a lot of help from numerous individuals. It is difficult to fully express the immense gratitude I feel for the knowledge and experiences I have gained from each one of them. Nonetheless, I will attempt to express my appreciation.

Firstly, I would like to thank my director, Ilga Porth, for her guidance and for giving me the chance to gain experience and find my interests. I am grateful for every moment. She consistently provides quick and helpful feedback on my work and is always proactive in helping come up with solutions for the challenges encountered in the research. Her hardworking attitude shows me that complaining doesn't help, and that focusing on practical problem-solving is the way to go.

I would like to say thank you to my co-directors who I feel lucky to know. Amanda Roe, who has provided me in-depth guide on where to focus more during data analysis and paper writing – invaluable lessons that will benefit me throughout my academic career. I truly appreciate it. Richard Hamelin, who has been supportive and always provide valuable insights from a broader perspective. Your positive energy is infectious and motivates everyone around you. I would also like to thank you for inviting me to visit University of British Columbia. Those two months have left a memorable mark on my journey.

I want to thank Isabelle Giguère for being such a great help with her efficient and skilled handling of lab work and paperwork. I have gained a lot and learned a great deal from her strong work ethic.

Thank you to Gwylim Blackburn for your guidance and mentorship. I always remember our weekly discussions and thank you for the online meetings during the Covid times. Your meticulous attitude towards science and the hundreds of pages of notes you take have been a great inspiration to me.

Thank you to Julien Prunier for introducing me to bioinformatics and providing me with valuable assistance with population genomics. Your insightful opinions and positive attitude are treasures.

I would like to thank Donna Mazerolle for expertly managing the financial paperwork and for always greeting me with a kind smile. Thank you to Jean-Guy Catford and André Gagné for maintaining the lab, to Brian Boyle for providing access to the sequencing platform, and to Jérôme Laroche for answering bioinformatics requests. And thank you to all the staff at IBIS who work tirelessly to ensure the smooth functioning of our beloved institute.

Thank you to all my friends at IBIS. You are the ones that keep me company for the most part and who I spent most of the time with.

Thank you to Ying Zhang for the two years company and support during the covid times. Thank you to Xiaojun Zhu for the delicious meals you prepared from time to time. Thank you to Hong Chen and Xuande Chen for your encouragement and support.

Thanks to my old pals back home. I miss you a lot. The thought of having you there always backing me up gives me the strength to keep going.

Thank you to my family, my ultimate harbour. It is important to know no matter how far I fly, there is always a place full of unconditional love for me. Thank you my parents. I will give you my most patience. A special thanks to my big brother for taking care of them, and to my sister-in-law for joining our family. We are blessed to have you. Thank all my extended family who have always been low-maintenance and loving. You remind me to be kind.

One last shout-out to the China Scholarship Council for awarding me a four-year scholarship. I am truly grateful for this opportunity.

Preface

The first chapter, titled “Genome-scale Phylogeography Resolves the Native Population Structure of the Asian Longhorned Beetle, *Anoplophora glabripennis* (Motschulsky)” has been published. The article was submitted to Evolutionary Applications on August 23, 2021, accepted on March 21, 2022, and made open access on April 11, 2022. There are no changes between the article included in the thesis and the published version. I am the first author of the article. I participated in the experimental design, performed the SNP markers-related DNA preparation, conducted all data analysis, and contributed to most of the writing and all revisions for the acceptance of the article. Yunke Wu, Jose A. Andres, Baode Wang, and Scott E. Pfister conducted microsatellite marker-related work, including sample collection, microsatellite data analysis, and writing the method and result sections for the microsatellites. Marion Javal, Géraldine Roux, Melody Keena, Juan Shi, and Evan Braswell provided ALB samples. Isabelle Giguère assisted with sample collection and SNP marker-related DNA preparation. Ilga Porth and Amanda Roe participated equally in the experimental design and the revision of the manuscript, with close discussion and revision from Richard Hamelin.

The second chapter, titled “Pool-Seq reveals genes selective for cold tolerance across Asian longhorned beetle’s native range” is currently unpublished. In this work, I contributed to the experimental design alongside I. Porth, who participated in meetings and revisions. With help from I. Giguère, I conducted the DNA preparation. I performed all the data analysis. Valuable scripts for SNP calling/filtering using bcftools were provided by Anna Fijarczyk, who also participated in multiple meetings. A. Roe made contributions to the revision process.

The third chapter, titled “Tracking the Asian longhorned beetle’s invasion history within North America with genomic evidence” is currently unpublished. For this chapter, DNA preparation was conducted with assistance from I. Giguère. Samples were provided by M. Keena. I performed all data analysis and drafted the methods and results section. A. Roe contributed to the introduction and discussion section. I. Porth participated in experiment design and paper revision.

Introduction

Due to globalization, invasive species have continuously increased worldwide. Invasive forest pests have greatly threatened forest ecosystems, and some have caused enormous economic losses, such as my focal species, the Asian longhorned beetle (ALB, *Anoplophora glabripennis*, Motschulsky, Coleoptera: Cerambycidae). Early intervention to prevent the introduction and spread of invasions is one of the most efficient ways to control invasive insects (Leung et al., 2002). This research aims to examine the genomics of ALB populations and leverage this knowledge for enhanced biosurveillance. The introduction will an overview of invasion biology and population genomics, followed by a discussion on the applications to invasive species and the case of ALB.

Invasion biology with an emphasis on forest insects

The grim reality of biological invasion

With the rapid pace of international trade and globalization, the rates of biological invasions have grown exponentially (Figure 1). Most successful invasions occur through unintentional introduction, and it can lead to massive ecological and economic losses (Hulme et al., 2009; Pyšek & Richardson, 2010; Seebens et al., 2017).

Invasions can cause changes in landscape structure, ecosystem functions, and redefine biogeography (Capinha et al., 2015). Invasive alien species (IAS) have significantly reduced biodiversity and are the second biggest threat to biodiversity following habitat loss (Lowe et al., 2000). For example, the invasion of *Miconia calvescens* has led to habitat loss for the endemic species, resulting in a change in the biodiversity of the invaded area (Burnett et al., 2007). According to a study by Cuthbert (2022), using records from the InvaCost database (Diagne et al., 2020), the 60 worst taxa caused an economic impact of US\$148.9 billion, while the impacts from other taxa amounted to US\$163.2 billion. To raise general awareness of invasive species, the International Union for Conservation of Nature (IUCN) initiated the list of the "100 worst invasive species" in 2000 (Lowe et al., 2000), which was updated in 2013 following the successful eradication of the rinderpest virus, leaving 37 plants, 30 vertebrates, 26 invertebrates, five fungi, and two micro-organisms on the list (Luque et al., 2014).

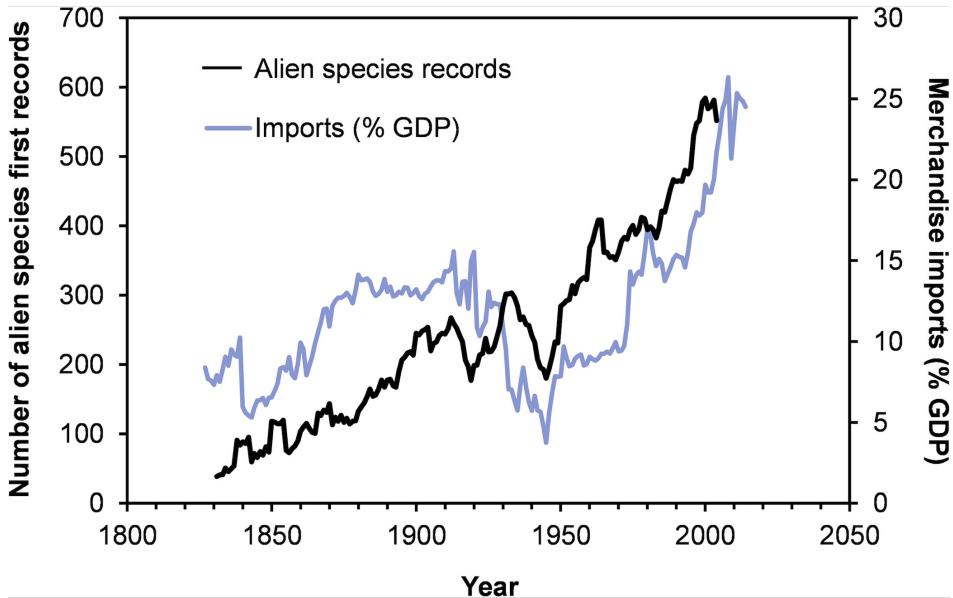


Figure 1: Biological invasion and global trade over time. GDP: Gross domestic product. Image source: Hulme (2021), licensed under CC BY 4.0.

Insects represent the most species-rich class in the animal kingdom, with many adapted to endure extreme environments. Consequently, it is not surprising that they are frequently involved in biological invasions. Among the "100 worst invasive species" list, 14 are insect species, and thousands of other forest insect species have successfully colonized non-native areas. Invasive insects contribute significantly to the global economic burden of invasive species, accounting for an estimated US\$70 billion ¹ per year globally (Bradshaw, 2016).

How can we effectively combat forest IAS?

Understanding the pathways of invasive insects is crucial. Generally, the invasion process comprises four stages: transport, introduction, establishment, and spread (Blackburn et al., 2011). An invasion can fail at any stage if the IAS cannot overcome barriers, or if appropriate measures are taken at the right time. Intervening during the early stages of invasion is the most cost-effective strategy, as managing invasive species becomes significantly more challenging and expensive once they are established and reach high densities (Leung et al., 2002).

Biosurveillance is key to this effort. It addresses questions of when, where, and how to search for new IAS while gathering as much information as possible about them (Roe et al., 2019). Biosurveillance involves collecting, integrating, and communicating information and risks associated with IAS (Roe et al., 2019). This process enables quick action to control and eradicate

¹The numbers may not be directly comparable to Cuthbert's study in 2022, as they utilize different data sources and methods.

invasions, thereby reducing the risks they pose to the environment, economy, and human health.

Biosurveillance efforts should be tailored to the stage of invasion (Poland & Rassati, 2019):

Pre-border (transport stage)

Human-mediated dispersal is a major driver of biological invasions (Holzapfel & Harrell, 1968; Hulme et al., 2009; Pyšek & Richardson, 2010; Seebens et al., 2017). In the case of forest insect invasions, key pathways include the transport of living plants and wood-packing materials (WPM), as well as trade in other wood products, hitchhiking insects on vehicles, mail, and passengers (Meurisse et al., 2019). To mitigate invasion risks associated with WPM, policies must be adopted to ensure the safe importation of commodities. Measures such as International Standards for Phytosanitary Measures (ISPMs) #15 have been developed and maintained by the Food and Agriculture Organization (FAO) to address these risks (FAO, 2002). However, despite these efforts, invasion rates continue to remain high due to factors such as increasing trade volumes, variations in quarantine procedures, and rapidly evolving distribution networks (Meurisse et al., 2019).

Border and post-border surveillance

Introduction stage At ports of entry, biosurveillance can detect new IAS using visual surveys, baited traps, biosurveillance animals, and genomic tools (Poland & Rassati, 2019). Swift actions, such as rejecting cargo or destroying infested materials, can prevent the establishment of new IAS. Prevention is the most efficient strategy, with eradication being the second-best option (Brockerhoff et al., 2010; Simberloff, 2003).

Early establishment stage There is generally a lag phase between introduction, with the duration of this phase varying among taxonomic groups (Memmott et al., 2005). Early intervention during this phase can effectively prevent further establishment of IAS. For instance, the Asian longhorned beetle (ALB) was successfully eradicated from Boston, Massachusetts, by removing only six trees, due to early detection and a coordinated response (Stefan et al., 2014).

Spread stage The spread stage of IAS marks the expansion beyond their initial introduction point. At this point, the IAS has successfully adapted to the local environment and starts to grow in number. Key management strategies during this stage include population reduction, aiming to lower the species numbers to non-detrimental levels; eradication, a challenging and resource-intensive goal that involves completely removing the IAS from the affected area; and long-term monitoring to evaluate the effectiveness of these strategies and promptly respond to population changes. These approaches need to be adaptable and are often applied in combination for effective management (Poland & Rassati, 2019).

Genomic Biosurveillance in IAS Management

In the context of globalization, innovative methods for early detection and rapid response can play a significant role in mitigating the risks associated with IAS throughout the invasion process. Approaches range from visual inspection to advanced techniques, such as genetic and genomic methods for diagnostic purposes. These methods provide rapid and accurate identification of forest IAS regardless of life stage, integrity of specimens, or even the detection from their frass (Poland & Rassati, 2019; Rizzo et al., 2020). Over the past decade, the application of genetic and genomic approaches to forest IAS biosurveillance has grown significantly (Roe et al., 2019).

Species identification

Genomic biosurveillance tools can help address questions related to species identification, invasion source, pathway and history, and traits aiding forest IAS invasion (Roe et al., 2019). This information can inform risk assessments and management strategies. For instance, DNA barcoding, a widely used molecular identification technique, has been applied extensively to identify invasive insect species, track their spread, and determine the effectiveness of control measures (Hebert et al., 2003), thus aiding in early detection at all stages of the invasion process.

Understanding invasion sources and history

Understanding the sources of invasive populations, their pathways of entry, and invasion history is important for pest risk assessments, improving our understanding of the invasion process, and informing management strategies (Roe et al., 2019). For example, genome scans have been used to identify the origins and routes of spread for the mountain pine beetle (MPB, *Dendroctonus ponderosae*), a destructive pest that has spread far beyond its native range (Samarasekera et al., 2012; Batista et al., 2016).

Identifying genomic markers for relevant traits

Certain traits of an IAS, such as invasion history, body size, generation time, availability or quality of food resources, physiological responses to novel environments, and other fitness traits are relevant to the establishment and success of invasions (Colautti, Grigorovich & MacIsaac, 2006). If these traits have a genetic basis (i.e., are heritable traits), genomic markers can be used to predict potential risk even without the availability of live material.

Approaches used to establish potential links between genomes and traits include quantitative trait locus (QTL) analysis (Mackay, Stone & Ayroles, 2009) and genome-wide association studies (GWAS) (Visscher et al., 2017), gene-environment association analysis (GEA) (de Villemerueil et al., 2014) and genome-wide selection scans (GWSS) (Nielsen et al., 2007). Various “-omics” techniques, such as transcriptomics, proteomics, and metabolomics are additional

approaches that target functional subsets of the genome to better understand the phenotype of invasive species. These techniques can shed light on the number of genes and their response in the process of invasion or adaptation (Wang, Gerstein & Snyder, 2009; Aebersold & Mann, 2016; Johnson, Ivanisevic & Siuzdak, 2016).

Genomic biosurveillance has the capacity to improve our understanding of the invasion process and contribute to protecting forests from invasive species and reducing their potential negative impacts. However, the field is still in its infancy and requires computational resources and bioinformatics expertise. Integrating genomic tools into biosurveillance presents challenges, including their implementation within plant health regulatory agencies (Roe et al., 2019; Bilodeau et al., 2019). However, as genomic resources for IAS advance, their use in biosurveillance will expand.

Population genomics application in invasive species

Population genetics studies focus on the source, scale, and distribution of genetic variance and how genetic variation changes over time and space. Population genetics approaches have been used to address essential questions about the invasion history of invasive species, such as the source(s) of introduction, whether single or multiple introduction events are involved, hybridization and introgression (gene flow between species), the extent of bottleneck effects, and local adaptation to the invasive range. The advent of next-generation sequencing (NGS) has greatly advanced population genetic studies to address more complex questions. It has facilitated the use of genome-wide molecular markers and the application of emerging bioinformatics and computational methodologies. Consequently, the term 'population genomics' has gained increased prominence. Charlesworth (2010) defines population genomics as the study of the amount and causes of genome-wide variability in natural populations.

The two terms, population genetics and population genomics, can be used interchangeably when discussing the overall goals of understanding genetic variation in populations. Both fields aim to understand genetic variation, population structure, adaptation, and evolution. However, they differ in the scope, scale, and methodologies employed. Population genomics is an extension of population genetics that leverages recent advances in genomic techniques and computational approaches (Balkenhol et al., 2017).

Characterizing population delimitation

Tracking introduction pathways and identifying the geographic sources of alien pests is an essential component of an invasive species biosurveillance pipeline (Cristescu, 2015; Bilodeau et al., 2019). Knowledge of invasion sources allows management responses to focus on high-risk points of entry and routes of spread, which facilitates invasive species monitoring, trade negotiations, and future risk assessments (Bilodeau et al., 2019). Isolated populations usually

form different spatial distributions of genetic variation due to genetic drift, limited gene flow and local adaptation. Such variations can be analyzed to trace back to the source of the invasion. Through the examination of genetic differentiation in both native and invasive populations, we can better understand the genetic structure and interrelationships among populations. Utilizing this knowledge can guide the development of targeted and effective biosurveillance and management strategies for high-risk pests.

In the process of understanding these dynamics, firstly, it is crucial to accurately characterize the population structure and genetic diversity within the native range of an invasive species. A clear delimitation of an invasive species' population structure in its native range is a prerequisite, so the intercepted individuals can be genetically assigned to a source population (Hamelin & Roe, 2020; Manel et al., 2005; Roe et al., 2019). Understanding the population differentiation characteristics of invasive populations is also imperative. It provides valuable insights into the invasion dynamics, enabling us to determine the source(s) of introduction and detect admixture among separate introductions.

A set of hypothetically neutral markers is commonly used to investigate population structure. Common methods include principal component analysis (PCA) (Abdi & Williams, 2010), a multivariate analysis which reduces dimensions down to several principal components to illustrate the population structure, and model-based methods. Model-based clustering methods, such as STRUCTURE, on the other hand, rely on assumptions of Hardy-Weinberg equilibrium (HWE) and Linkage disequilibrium (LD). These methods assume that each individual originates from one of the populations in accordance with HWE (Pritchard, Stephens & Donnelly, 2000; Alexander, Novembre & Lange, 2009). They assign individuals to different subpopulations while minimizing the deviation from HWE and LD within each subpopulation. These approaches provide valuable tools for understanding and characterizing population structure in invasive species studies.

Invasion history inference

Studies on invasive species frequently utilize genetic methods to reconstruct their invasion history. Historical records of introductions that can be revisited to test different hypotheses around the invasion processes using genome-wide markers. The common use of molecular markers in invasion biology studies has enabled us to reconstruct the evolutionary history of biological invasions. This includes determining the routes of introduction and spread, estimating the founding population size, and detecting bottleneck events, sometimes with high precision (Estoup & Guillemaud, 2010; Bock et al., 2016).

Clustering methods can offer valuable insights into the history of species invasions (Estoup & Guillemaud, 2010). However, when the invasion history is more complex, such as multiple introductions, gene flow between separate introductions, and evolution due to genetic drift,

genetic admixture in divergent populations can cause clustering methods to incorrectly assign admixture ancestry (Kalinowski, 2011; Wang, 2016; Lawson, Van & Falush, 2018). There are now exist methods that can estimate various invasion scenarios while considering complex evolutionary events, such as bottlenecks, multiple introductions, and gene flow among populations. Commonly used methods include coalescent simulations in the Approximate Bayesian Computation (ABC) framework (Csilléry et al., 2010), such as DIYABC (Cornuet et al., 2014). In addition to coalescent methods, the $\delta\alpha\delta$ i method uses a diffusion approximation to model the evolution of the joint allele frequency spectrum of populations under alternative models, with faster computation relative to ABC simulations (Gutenkunst et al., 2009). Not only do these methods account for more intricate evolutionary events, but they also offer computational efficiency. This becomes especially relevant when genome-wide markers from whole genome resequencing need to be considered.

Identification of genomic signatures reflective of adaptation

Based on the principles of Kimura's neutral theory, it can be assumed that most mutations are either neutral or deleterious (Kimura, 1968). Deleterious mutations continually occur and are eventually eliminated due to natural selection by means of purifying selection, as a consequence, they are detected at low frequency within the genomes across all populations. Neutral mutations account for a substantial portion of all the nucleotide changes that can be observed.

Though it is difficult to detect beneficial mutations because of their small effect sizes, due to linkage disequilibrium in the genome surrounding these mutations, they are often accompanied by a selective sweep phenomenon. So it is possible to detect such selective sweeps because of the three features they leave on the genome: (1) a decrease in regional polymorphism; (2) a skewed site frequency spectrum (SFS); and (3) a unique pattern of linkage disequilibrium (LD) (Alachiotis & Pavlidis, 2018).

The common population genetics approach to detect loci underlying selection is to identify loci with divergent allele frequencies among populations. Common methods are population differentiation-based, such as F_{ST} ; SFS-based methods and LD-based methods. Thus only through the acquisition of genetic data, it is possible to obtain valuable insights into the genes that are responsible for adaptation (Figure 2) (Weigand & Leese, 2018).

Understanding the phenotypes or quantitative genetic characteristics that contribute to the successful establishment and adaptation of an invasive species in a new environment can be useful for assessing the risks posed by such species. By identifying the genomic signatures underlying these phenotypes, it becomes possible to predict the likelihood of an alien insect establishing and expanding in specific regions based on such characteristics (for example, cold hardiness), serving as a primary objective in genomic-based monitoring and risk assessment

strategies. Association methods integrating various elements, such as genetic data, geospatial data such as climate-related data, and phenotypic data, can then identify adaptive genetic variation (Figure 2). By analyzing the associations between genetic variation and environmental variables of interest using GEA analysis, one can identify genomic signatures that reflect selection influenced by these environmental factors. This approach is an aspect of landscape genomics (de Villemereuil et al., 2014); On the other hand, when associating genetic data with specific adaptive phenotypic traits, one can pinpoint genes related to these adaptive traits. Commonly employed methods for this include GWAS and QTL mapping (Mackay, Stone & Ayroles, 2009; Visscher et al., 2017). Ideally, if SNPs found to be subject to selection are also identified underlying these adaptive traits, such findings help in the interpretation of genetic signals of local adaptation (Sork et al., 2013; Porth et al., 2015). The successful application of such methods allow us to apply them to invasive insects with regards to determining signatures of adaptive evolution in their genomes.

The field of invasion biology is witnessing significant progress thanks to advancements in techniques and methodologies. This progression has elevated the level of sophistication in research, enhancing accuracy in understanding biological invasions. As a result, we are able to deepen our scientific knowledge in this field, and develop more effective strategies for managing and mitigating the impacts caused by invasive species.

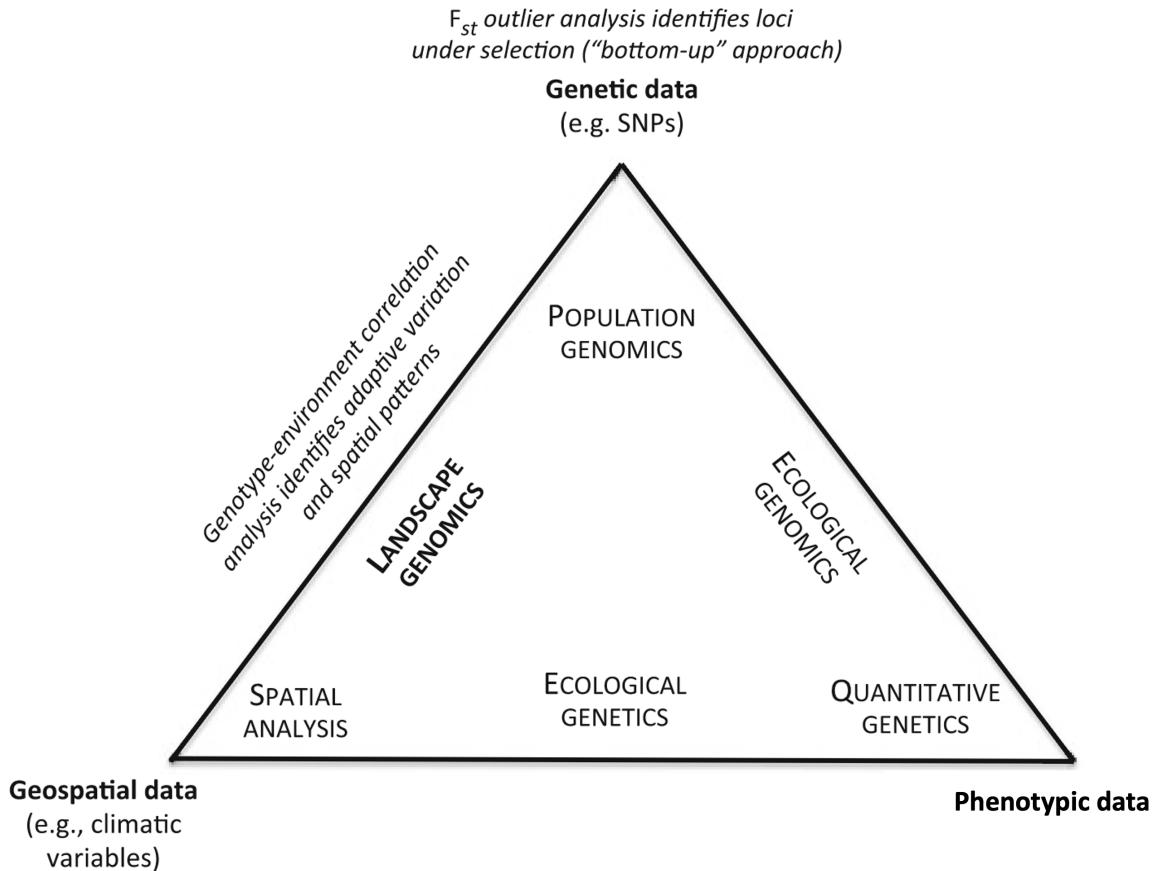


Figure 2: Approaches to identify adaptive genetic variation. The “top-down” approach analyzes a combination of genotypic and phenotypic data to identify genetic loci that underlie phenotypic variation for adaptation. The “bottom-up” approach requires solely genotypes but does not need information on phenotypic trait variation to identify genetic variation related to adaptation. Ultimately, a combination of both approaches will provide the most comprehensive picture of genetic variation linked to environmental cues. Figure (Sork et al., 2013) edited with permission of Springer [License Number: 5300841022950].

Invasion of the Asian Longhorned Beetle

The Asian longhorned beetle (ALB), *Anoplophora glabripennis* Motschulsky (Coleoptera: Cerambycidae), native to China and the Korean peninsula, is a highly destructive wood-boring pest that poses a significant threat to broadleaved trees (Lingafelter & Hoebeke, 2002). The characteristics of its various life stages are provided in Table 1.

Table 1: Life cycle and characteristics of *A. glabripennis*. If not specified otherwise, the information are taken from Yan, Qin & Xiao (1992). The measurements provided may be obtained under specific conditions, such as host species, temperature, and other rearing conditions. Images are kindly provided by Dr. Lixiang Wang.

Life Stage	Size	Position and Feeding	Lifespan	Image
Eggs	5.5-7 mm in length	Laid by female adults in the bark	8-12 days to mature (Li & Wu, 1993); 15.0 ± 0.6 days at 25°C (Keena, 2006)	
Larvae	Up to 50 mm in length, with head capsule up to 5 mm (Cavey, 1998)	Five or more instars; tunnel inside the tree and feed on phloem, cambium, and xylem	1-2 years	
Pupae	30-37 mm in length and 11 mm in width	Found in the tree within the larval tunnels	12-50 days (Keena & Moore, 2010); 19.6 days (Hu et al., 2009)	
Adults	20-35 mm in length and 7-12 mm in width	Emerge during summer to early fall; feed on leaf petioles and cambium of twigs	<50 days (♂) (Yan, Qin & Xiao, 1992); <66 days (♀); <158 days (♀); <202 days (♂) (Keena, 2006)	

Infestations of ALB can lead to the illness or death of healthy trees, causing considerable economic and ecological losses. Nowak et al. (2001) evaluated that the feeding activity of ALB larvae can kill a large host tree within just four years. They suggested that a widespread ALB infestation could destroy nearly 30% of the urban trees in the US, leading to \$669 billion in economic damage. Consequently, the invasion of the Asian longhorned beetle presents an ongoing challenge for the management and preservation of a number of broadleaved tree species, necessitating effective monitoring, prevention, and control strategies to mitigate its impact on ecosystems and economies.

ALB's adaptation and invasiveness potential

For a species to become invasive, it must possess certain characteristics that enable it to adapt readily to new environments. Many invasive insects share common features such as a wide

host range, high reproductive ability, high tolerance to extreme temperatures, insecticides, and pathogenic infections, and a lack of parasitoids (Wan & Yang, 2015). The Asian longhorned beetle (ALB), one of the 14 insect species on the list of the "100 worst invasive species", exhibits many of these traits.

Like other insects, ALB can produce a large number of offspring rapidly. Under laboratory conditions, females have been observed to lay an average of 50-75 eggs (Keena, 2002). A single female can copulate with one male up to 27 times and produce 1,366 progeny (Keena & Sánchez, 2006). Smith, Bancroft & Tropp (2001) observed even higher numbers of eggs, emphasizing the importance of host species and host quality. However, ALB generally shows higher fecundity and longevity in the laboratory than generally observed (Keena, 2002).

ALB has been recorded to attack more than 100 tree species (Meng, Hoover & Keena, 2015). However, it is important to note that not all host tree species have been confirmed to support the complete development of ALB. In China, the major hosts of ALB are tree species from four genera, *Populus* (poplar), *Salix* (willow), *Ulmus* (elm), and *Acer* (maple) (Yan, Qin & Xiao, 1992). In North America, *Acer* is strongly preferred.

ALBs primarily overwinter as larvae, although eggs and pupae also show some level of overwintering capacity on rare occasions (Haack et al., 2006). Feng et al. (2016) observed that ALB larvae are capable of bearing within-body ice formation and it was also observed surviving freezing by Roden et al. (2009), suggesting freeze-tolerance as cold tolerance strategy of ALB larvae. Torson et al. (2021) also confirmed larvae as freeze-tolerant and they observed more than 92% of larvae were able to survive at -25.8 °C or lower for 24 h under laboratory condition; eggs were freeze-avoidant with an average supercooling point (SCP) of -25.8 °C and larvae have a lethal temperature limit of -25 °C; but other life stages, such as pupae and adults, cannot survive exposure to freezing temperature.

It is known that insects accumulate natural cryoprotectant molecules (CPs), such as sugar alcohol (e.g. glycerol), and sugar (e.g. trehalose), and amino acids (e.g. proline) when acclimated to low temperature (Lee, 1991; Storey & Storey, 2012). It has been found that these CPs help maintain the integrity of the cell membrane to prevent freezing injury in freeze-tolerant insects (Grgac et al., 2022). Other than the low-molecular-weight CPs, large-molecular cryoprotectants such as lipids (e.g. antifreeze glycolipids), ice-binding proteins (e.g. ice-nucleating agents (INAs), antifreeze proteins (AFPs), and antifreeze glycoproteins (AFGPs)), and transport proteins (e.g. aquaporins (AQPs) and cryoprotectant transporters) are assumed to be associated with insect freeze tolerance. Other proteins such as antioxidants, cell adhesion proteins, chaperonones (e.g. heat shock proteins (HSPs)), chelators, cytochrome P450s, cytoskeletal protein isoforms, disordered proteins (e.g. dehydrins), and sirtuins are predicted to facilitate freeze tolerance (Toxopeus & Sinclair, 2018). Although CPs can be an indicator of cold tolerance capacity, little still we know about the cold tolerance mechanisms

regulated in freeze-tolerant insects. No single CP or group of CPs is accumulated by all freeze-tolerant insects (Toxopeus et al., 2019; Toxopeus & Sinclair, 2018). For example, glycerol, which is the earliest known cryoprotectant (CP), can be found accumulated in most freeze-tolerant insects, but not in orthopterans such as crickets or the drosophilid fly *Chymomyza costata* (Grgac et al., 2022; Toxopeus et al., 2016). Interestingly, even though glycerol is not accumulated by *C. costata*, it has been shown to be more effective in protecting the membranes of freezing fat body cells *in vitro* than proline and trehalose, both of which are naturally accumulated by *C. costata* larvae (Grgac et al., 2022).

Feng et al. (2014, 2016) investigated the geographical and seasonal variation in supercooling point (SCP) and cryoprotectant concentrations, including glycerol, glucose, and trehalose, in ALB populations. Their findings revealed that these variables differ among different geographical populations and among seasons. Torson et al. (2021) also observed an increase in hemolymph glycerol content during chilling in laboratory-reared ALB larvae.

Distribution and invasion history

The Asian Longhorned Beetle (ALB) first drew attention due to its severe damage it caused in the 1980s and has since spread broadly. In China, it was first recorded in Beijing in 1853 (Motschulsky, 1853). It has been reported in 24 provinces and municipalities in China initially (Yan, Qin & Xiao, 1992). The highest damage has been observed in Liaoning, Hebei, Jilin, Heilongjiang, Inner Mongolia, Beijing, Tianjin, Ningxia, Gansu, Henan, Anhui, Shaanxi, Shanxi, Shandong, and Zhejiang (Gao & Li, 2001).

ALB has been introduced and established in North America and Europe due to the expansion of global trade. In North America, it was first reported in New York in 1996, followed by Illinois (1998), New Jersey (2002), Ontario (2003), Massachusetts (2008), and Ohio (2011), and most recently, it was detected in South Carolina in 2020 (Haack et al., 1996; Hopkin, Scarr & Marie, 2003; Shatz et al., 2013; Dodds et al., 2013; Turgeon et al., 2015; Coyle et al., 2021); In Europe, ALB was first reported in Austria in 2001, followed by France (2003), Germany (2005), Italy (2007), the Netherlands (2010), Switzerland (2012), Finland (2015), and Montenegro (2015) (Hérard et al., 2006; Schröder et al., 2006; Tomiczek and Hoyer-Tomiczek, 2007; Forster and Wermelinger, 2012). The timeline of invasion is given in Figure 3.

As a major destructive invasive insect pest, details of ALB's invasion process, such as its source of invasion, invasive history and pathway, are needed to better regulate its management. The damage caused by ALB and the current control situation must be thoroughly evaluated to mitigate the negative impact of this invasive species.

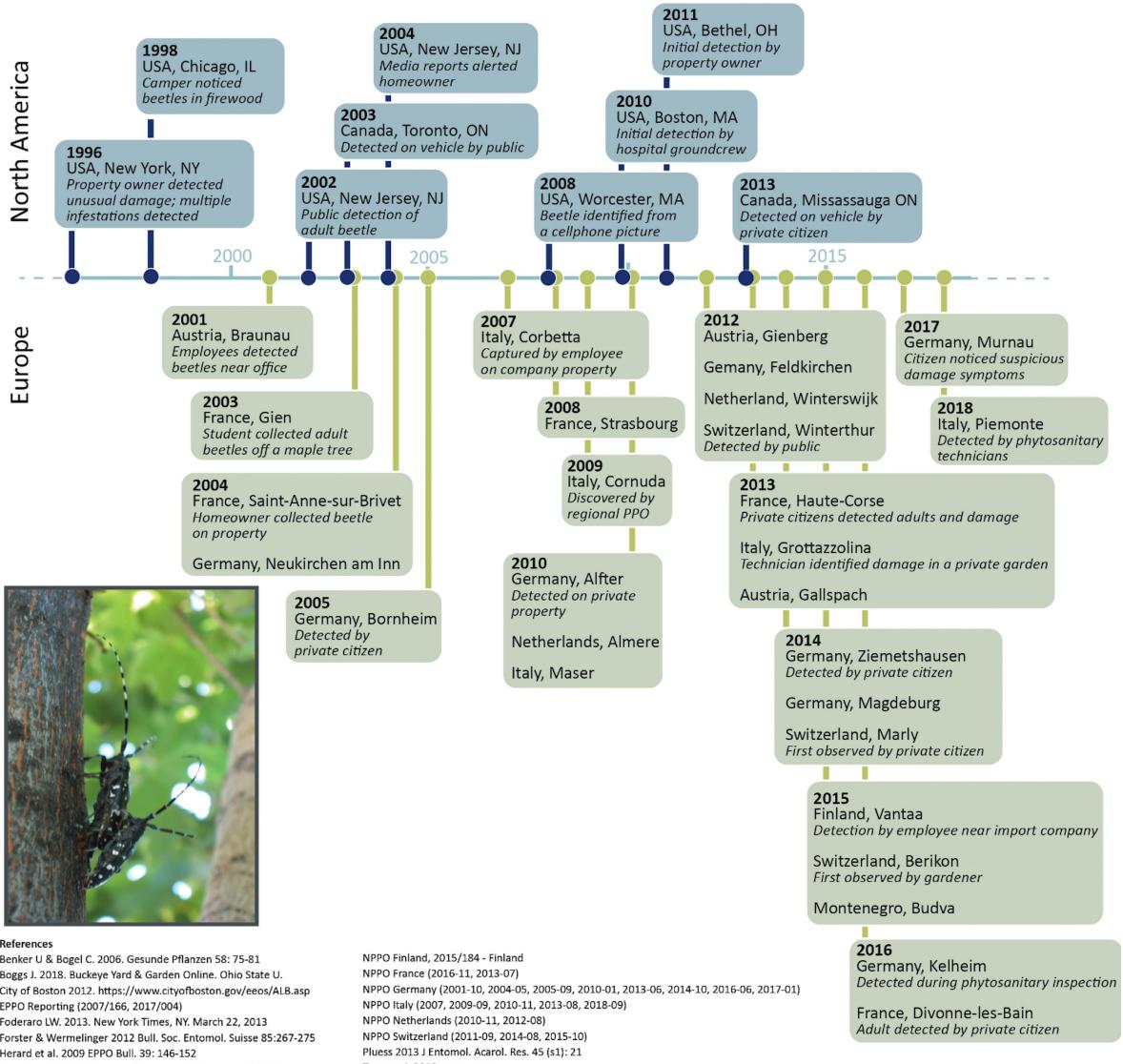


Figure 3: ALB infestations throughout its invasive range over 23 years of detections. Figure (Blackburn et al., 2020) used with permission of Oxford University Press [License Number: 5300940449579].

ALB's invasion genetics

Genetic approaches have been employed to identify Asian longhorned beetle (ALB) species and gain insights into its invasion history and pathways. These genetic methods involve analyzing the genetic material of ALB specimens to determine their species and shedding light on their invasion history and pathways, contributing to more effective management and control strategies for this invasive pest.

Kethidi et al. (2003) successfully used Sequence Characterized Amplified Region (SCAR)

markers to distinguish ALB from closely related species such as *A. chinensis* and *A. malasiaca*. Wu et al. (2017) combined DNA barcoding and morphological identification and successfully identified 275 Cerambycidae specimens including ALB. Rizzo et al. (2020) developed a rapid diagnostic protocol based on loop-mediated isothermal amplification (LAMP) to identify ALB using frass samples.

A phylogeny analysis (An et al., 2004) based on random amplified polymorphic DNA (RAPD) markers showed geographical genetic variation among six Chinese ALB populations, indicating genetic differentiation among native populations. Carter et al. (2009) utilized mitochondrial DNA (mtDNA) and microsatellites to study native ALB populations and suggested admixture among different Asian ALB populations due to natural dispersal abilities and reforestation efforts in China. They also used mtDNA to study the invasive history and pathway of ALB populations through genetic analysis of both native and invasive ranges. Javal et al. (2019) using the mtDNA COI gene on populations showed that no clear population structure was observed across the native range of ALB in China, and suggested multiple introduction events and secondary spread within the invasive range following the introduction of ALB in both Europe and North America. Lee, Lee & Lee (2020) conducted a study on the population structure of native ALB in South Korea and identified three distinct subgroups. They also proposed the occurrence of ALB invasion within its own native range, specifically from northeastern and northwestern China to urban areas in Korean.

Previous genetic studies have provided valuable insights regarding ALB's invasion history. However, to validate and complement these studies, as well as to obtain further information about ALB's adaptive capabilities, more extensive sampling and informative genome-wide markers, such as single nucleotide polymorphism (SNP), are needed. With the publish of the draft genome of ALB (710 Mb) by McKenna et al. (2016), it becomes more convenient to explore genome-wide SNP markers and enabling researchers to investigate focused genetic questions. Genetic analyses leveraging the power of genomics will provide far ranging information about the invasion of ALB and can better help regulate its management in the future. Given the continuous threat posed by ALB, it is crucial to gain a deeper understanding of its invasion dynamics and use the insights obtained from genomics for practical applications.

Research hypotheses and objectives of thesis

The aim of this PhD thesis is to inform the design of genomic tools to contribute to the biosurveillance of the invasive insect *A. glabripennis*. It consists of three chapters; each one serves a different objective and addresses different questions (Figure 4).

Chapter One:

Research question: How accurately can genome-wide SNP markers assign individuals to *A.*

glabripennis populations? What is the minimum number of SNPs required to achieve accurate population assignment?

Hypothesis: Genome-wide SNP markers can provide equal or better resolution than a limited number of microsatellites or mitochondrial markers when examining population structure of *A. glabripennis* populations. A small subset of these SNP markers can be identified to distinguish between genetic groups within these populations.

Research objectives: This chapter aims to (1) characterize the native population structure of *A. glabripennis* populations, and (2) identify the minimum number of SNP markers needed to assign individuals to populations with high probability.

Chapter Two:

Research question: Is there extensive non-neutral genetic variation among *A. glabripennis* populations? What are the genomic signatures underlying local adaptation/cold tolerance?

Hypothesis: It is hypothesized that there is geographical variation in cold tolerance of *A. glabripennis* populations and that it can be detected at the genome level.

Research objectives: This chapter aims to identify the genomic determinants of cold tolerance based on whole-genome sequence data.

Chapter Three:

Research question: Is there genetic variation among invasive *A. glabripennis* populations in North America? What are the native source populations? Do they originate from single introductions?

Hypothesis: It is hypothesized that there is genetic variation among the North American invasive populations that can help us to reconstruct the invasion history of *A. glabripennis*.

Research objectives: This chapter aims to identify the source populations of invasive *A. glabripennis* in North America.

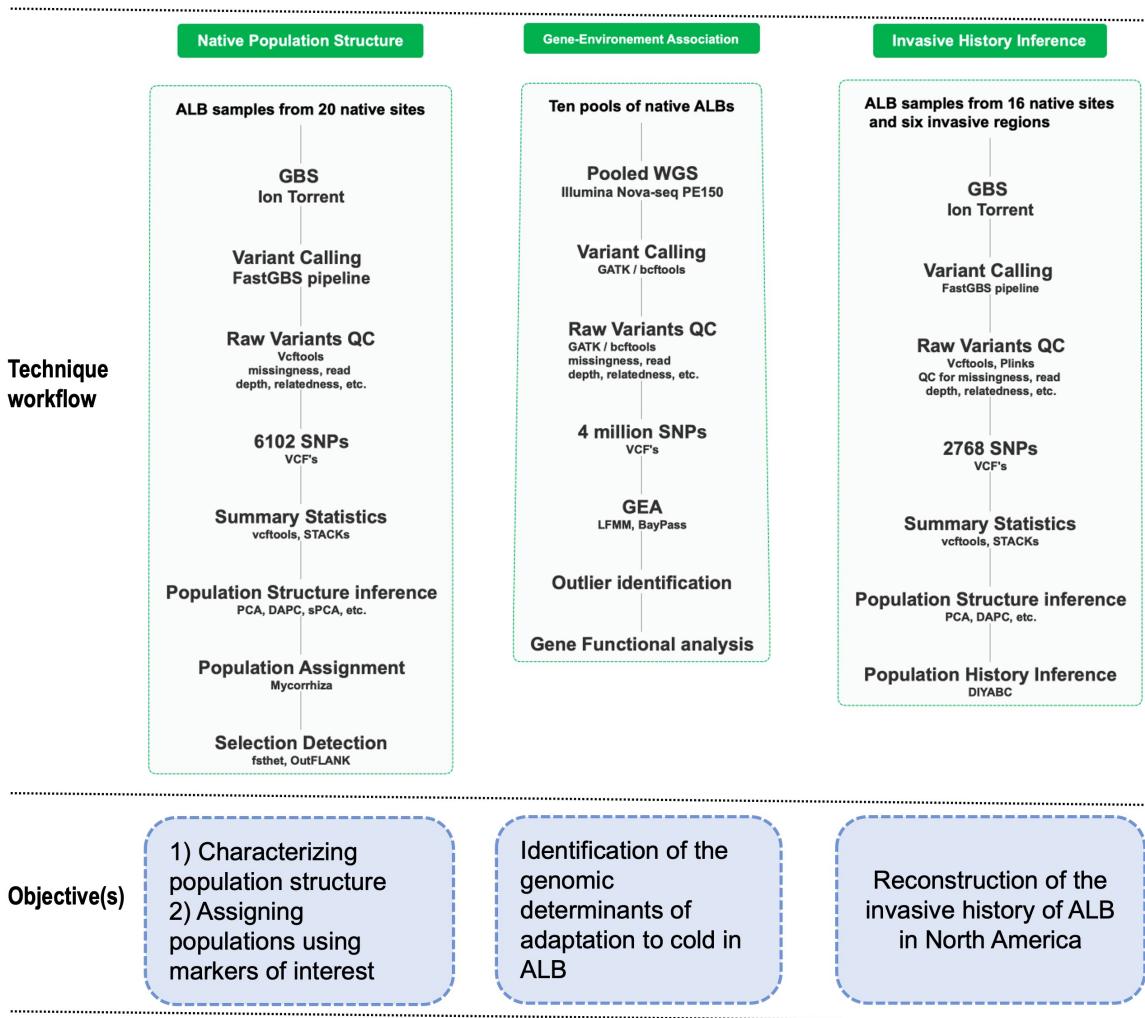


Figure 4: Flowchart of research

Chapter 1

Genome-scale phylogeography resolves the native population structure of the Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky)

1.1 Résumé

Les échanges mondiaux ont permis au longicorne asiatique (ALB, *Anoplophora glabripennis* (Motschulsky)) de se propager au-delà de sa zone d'origine et de devenir un nuisible envahissant au niveau mondial. Dans sa zone d'origine, qui comprend la Chine et la péninsule coréenne, la dispersion liée à l'homme a également causé la translocation cryptique d'insectes, résultant en une structure populationnelle complexe. Les études précédentes ont utilisé des méthodes génétiques pour démêler cette complexité mais n'ont pas réussi à délimiter clairement les populations natives, ce qui est nécessaire pour développer des outils de biosurveillance.

Nous avons utilisé des marqueurs à l'échelle du génome pour définir la structure historique des populations natives de l'ALB et le mouvement contemporain entre les régions. Nous avons utilisé le séquençage par génotypage pour générer 6,102 polymorphismes nucléotidiques (SNPs) et le séquençage d'amplicons pour génotyper 53 microsatellites. Au total, nous avons génotypé 712 individus de la distribution native de l'ALB. Nous avons observé six groupes distincts parmi les populations natives de l'ALB, avec une délimitation claire entre les groupes du nord et du sud. La plupart des individus de Corée du Sud étaient distincts des populations en Chine. Nos résultats indiquent également une divergence historique entre les populations et suggèrent un mélange génétique limité à grande échelle, même si le nombre de cas de mouvement contemporain identifié entre les régions était restreint.

Nous avons identifié des SNPs sous sélection et décrit un schéma de fréquence allélique clinale pour une variation génétique « faux-sens » associée à la glycéro kinase, une enzyme importante dans l'utilisation d'un cryoprotecteur chez les insectes. Nous démontrons en outre qu'un petit nombre de SNPs peut attribuer des individus à des régions géographiques avec une grande probabilité, ouvrant la voie à de nouveaux outils de biosurveillance de l'ALB.

1.2 Abstract

Human assisted movement has allowed the Asian longhorned beetle (ALB, *Anoplophora glabripennis* (Motschulsky)) to spread beyond its native range and become a globally regulated invasive pest. Within its native range of China and the Korean peninsula, human-mediated dispersal has also caused cryptic translocation of insects, resulting in population structure complexity. Previous studies used genetic methods to detangle this complexity but were unable to clearly delimit native populations which is needed to develop downstream biosurveillance tools. We used genome-wide markers to define historical population structure in native ALB populations and contemporary movement between regions. We used genotyping-by-sequencing to generate 6,102 single nucleotide polymorphisms (SNPs) and amplicon sequencing to genotype 53 microsatellites. In total, we genotyped 712 individuals from ALB's native distribution. We observed six distinct population clusters among native ALB populations, with a clear delineation between northern and southern groups. Most of the individuals from South Korea were distinct from populations in China. Our results also indicate historical divergence among populations and suggest limited large-scale admixture, but we did identify a restricted number of cases of contemporary movement between regions. We identified SNPs under selection and describe a clinal allele frequency pattern in a missense variant associated with glycerol kinase, an important enzyme in the utilization of an insect cryoprotectant. We further demonstrate that small numbers of SNPs can assign individuals to geographic regions with high probability, paving the way for novel ALB biosurveillance tools.

1.3 Introduction

Globalization has increased the spread and establishment of invasive species throughout the world, causing irreversible damage to forest ecosystems (Millar & Stephenson, 2015; Seebens et al., 2017). As it is difficult and expensive to manage established invasive species, the prevention and early detection of new invasive species and populations are the cornerstones of an effective management response (Reaser et al., 2020). In this respect, biosurveillance represents a knowledge framework that facilitates early detection and rapid response to new invasive threats, reducing the risk and impact of invasive species in novel habitats (Blackburn et al., 2020).

Tracking introduction pathways and identifying the geographic sources of alien pests is an

essential component of an invasive species biosurveillance pipeline (Cristescu, 2015; Bilodeau et al., 2019). Knowledge of invasion sources allows management responses to focus on high-risk points of entry and routes of spread, which facilitates invasive species monitoring, trade negotiations, and future risk assessments (Bilodeau et al., 2019). However, this approach requires clear delimitation of an invasive species' population structure in its native range so that intercepted individuals can be genetically assigned to a source population (Hamelin & Roe, 2020; Manel et al., 2005; Roe et al., 2019). Thus, accurately characterizing the population structure and genetic diversity within the native range of an invasive species is an essential step towards effective biosurveillance and management of high-risk pests.

The Asian longhorned beetle (ALB, Figure 1.1) (Coleoptera: Cerambycidae: *Anoplophora glabripennis* (Motschulsky)) is a wood-boring insect native to China and the Korean peninsula that has become a highly destructive invasive pest and poses a global threat to temperate broadleaved forests (Lingafelter & Hoebeke, 2002). This species is broadly distributed in East Asia, spanning much of the temperate forested regions in mainland China and the Korean Peninsula and extending south of 35 °N into subtropical regions in China and it broadly overlaps with the citrus longhorned beetle (CLB, *A. chinensis* (Forster)), another important invasive pest. Within its native range, ALB is considered an important forest pest, damaging and killing many tree species, such as poplar (*Populus* spp.), willow (*Salix* spp.), and maple (*Acer* spp.) (Lingafelter & Hoebeke, 2002; Sjöman et al., 2014). In less than two decades, the area that ALB populations occupied had increased by seven-fold, and by 1994 had impacted over 333,000 ha of broadleaf forests within China (Luo et al., 2000).

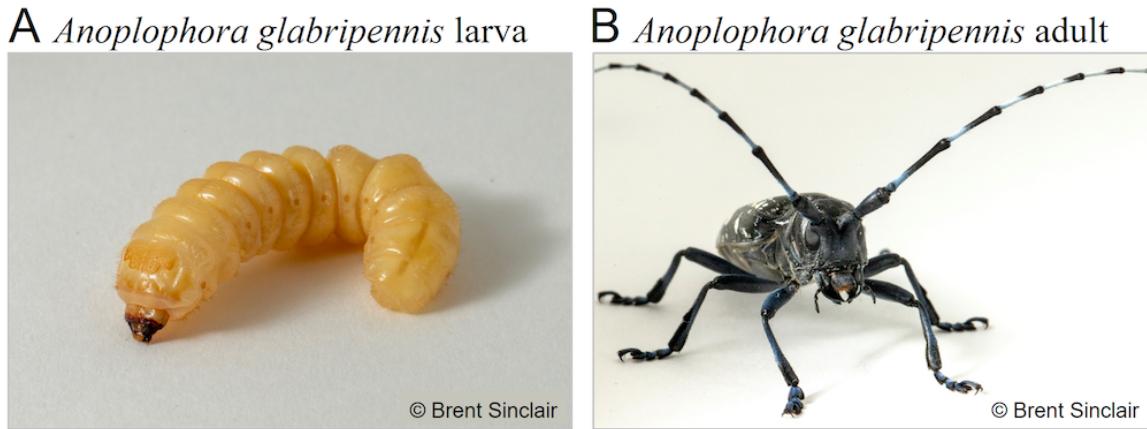


Figure 1.1: Larva and adult of *Anoplophora glabripennis*. Pictures were kindly provided by Dr. Brent Sinclair.

ALB was initially detected in North America in 1996 and then in 2001 in Europe where it has established and spread, causing severe ecological and economic losses in these newly invaded regions (Meng et al., 2015). ALB larvae tunnel and feed within the host xylem, making visual

detection challenging. Wood damaged by ALB is inferior in quality and is often used to construct solid wood packaging material (e.g., pallets, crates, dunnage) which is then used to transport goods. Indeed, infested wood packaging material is considered a crucial pathway for ALB's invasion, facilitated by increasing globalization and trade (Wu et al., 2017). In response to the ALB invasion, the Food and Agriculture Organization (FAO) established International Standards for Phytosanitary Measures (ISPM) 15, which set minimum heat and fumigation treatments for wood packaging material used in global trade to render potential invasives nonviable (FAO, 2002). However, prior to global acceptance of ISPM 15, ALB was transported throughout the world and introduced to many new broadleaf ecosystems. Despite these new phytosanitary protocols, ALB continues to pose a significant threat to temperate forests outside its native range with recent interceptions and introductions (Haack, 2020; Pedlar et al., 2020) throughout the invaded range. Therefore, biosurveillance tools are needed to manage the threat posed by ALB and identify the regional source(s) of invasion.

Complex biogeographic patterns are typical among native species within China (Lyu et al., 2020). While East Asia is known to maintain complex, deeply divergent population structure (Qiu et al., 2011), this historic structuring can be eroded with contemporary migration between distinct populations, often facilitated through anthropogenic movement of individuals beyond their ability to disperse naturally (Eloff et al., 2020). In ALB's native range, population differentiation is driven by both evolutionary and anthropogenic processes (Colunga-Garcia et al., 2010; Jixia Huang et al., 2020; Shatz et al., 2013). Indeed, forestry practices in northern China from the 1950s-1970s, including the Three-North Shelter reforestation project which is the largest artificial forest in the world, created forest conditions that caused ALB populations to irrupt and spread beyond their historic range. Previous studies reported moderate levels of population differentiation, but also found high levels of admixture (Carter et al., 2009; Javal, Lombaert, et al., 2019; Javal, Roques, et al., 2019). These authors hypothesized that this complex population structure was due to widespread contemporary movement of ALB associated with movement of infested wood out of China's Three-North Shelter reforestation plantation. This assisted migration caused gene flow between different populations, blurring the historic population structure. However, these previous methods used random amplified polymorphic DNA markers (An et al., 2004), mitochondrial DNA (Carter et al., 2009, 2010; Javal, Roques, et al., 2019), and small sets of microsatellite markers (Carter et al., 2010; Javal, Lombaert, et al., 2019), which lack the resolution needed to describe complex patterns of historic population structure and contemporary gene flow and therefore, more information-rich marker systems are needed.

Genomic data provide unparalleled insights into the invasion history of an invasive pest (Hamelin & Roe, 2020). These data can quantify variation among populations and allow us to identify individuals that share common ancestors and geographic origins. Genomic technologies, such as reduced representation libraries (Altshuler et al., 2000; Van Tassell et

al., 2008), provide access to variable genetic loci throughout the genome. Genotyping-by-Sequencing (GBS) (Elshire et al., 2011), double digest restriction-site associated DNA sequencing (ddRADseq) (Peterson et al., 2012) and DNA simple sequence repeats (microsatellites)-based genotyping (Jarne & Lagoda, 1996; Luikart et al., 2003) have been used to characterize population structure in a range of invasive insect pests, such as *Lymantria dispar* from Asia (Picq et al., 2018; Wu et al., 2020), the Asian granulated ambrosia beetle *Xylosandrus crassiusculus* (Storer et al., 2017), and the fall webworm *Hyphantria cunea* (Cao et al., 2016). With these approaches, high numbers of polymorphic markers can resolve fine-scale patterns of genomic differentiation and assign invasive samples to distinct populations with high levels of accuracy. Furthermore, genomic markers were able to identify signatures of population expansion within the invaded range, providing further insight into the patterns of invasion and spread within invasive pest populations (Cao et al., 2016; Picq et al., 2018; Storer et al., 2017; Wu et al., 2020).

In this study, we comprehensively defined the native ALB population structure and quantified the scale of contemporary gene flow in this species by leveraging the informative power of genome-wide molecular markers. To complement and cross-validate our inferred population structure, we used two sets of independent data: SNPs obtained from GBS, and microsatellites obtained from amplicon sequencing. First, we generated SNPs using a reduced representation library approach to characterize genomic diversity and population structure in ALB’s native range. We then validated our proposed population structure with an independent set of insect samples and 53 microsatellites developed with ddRADseq and genotyped through amplicon sequencing. Using our panel of informative SNPs, we delineate ALB population structure, reconstruct population history scenarios to describe the observed population structure, and identified loci that were under selection that may show adaptive significance within ALB populations. We identified loci that proved diagnostic for our delineated ALB populations and provided high assignment accuracy for downstream use in an amplicon-based ALB biosurveillance tool for this high-risk invasive forest pest.

1.4 Materials and Methods

1.4.1 ALB sampling and DNA preparations for SNP genotyping

We collected 480 samples from 20 sites across China and South Korea (Figure 1.2A; sampling details in Table S1.1), which covered c.40% of ALB’s distribution in the native range according to its most current distribution record (Yan, 1985; CABI, 2021). Our samples also covered several Three-North regions, including Heilongjiang, Jilin, Liaoning, Hebei, Beijing City, Inner Mongolia, and Ningxia. These represent the major biogeographic regions in northern China, with the Greater Khingan Range dividing the Northeast and the Helan Mountains dividing the Northwest (He et al., 2017; Jihong Huang et al., 2012). In the south, the Huai River basin

is a major transition zone between northern and southern climatic regions as it approximates the 0 °C January isotherm (Shi et al., 2014). The Yangtze River is also recognized as an important North-South boundary within the region. Based on the regions separated by these major geographic barriers, we defined the three regions located north of the Huai River as “Northeast (NE)”, “North (N)”, and “Northwest (NW)”, respectively, and the one southern region, where ALB sampling occurred, “South (S)”. Bengbu (BB) is located within the Huai River basin and was therefore included in the “South” region together with Cixi (CIX).

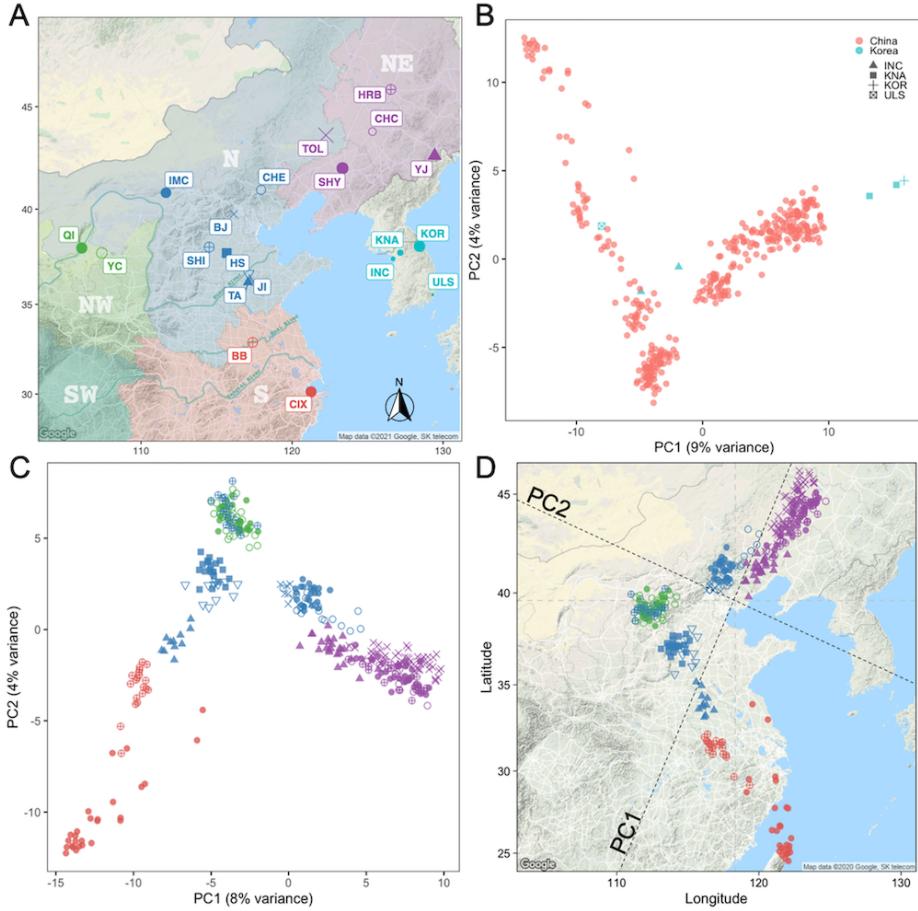


Figure 1.2: Sampling and Principal Component Analysis (PCA) of *Anoplophora glabripennis*. **A** Sampling map. The codes for the populations are shown in Table 1.1. The icon size of each population is relative to the sample size. We divided the ALB sampling locations in China into four areas, according to biogeography, distinguished by different colors: N, north China (blue); NW, northwest China (green); NE, northeast China (lilac); S, south China (red); in addition to China, Korea (turquoise), was also sampled for ALB. Each point with a different color or shape represents a distinct sampling location. TOL is treated as a NE population given its geographical location. The three rivers shown are Huang River, Huai River, and Yangtze River from north to south. **B** PCA of ALB populations in the native range based on 6102 SNPs. **C** PCA of Chinese ALB populations. **D** Procrustes-transformed PCA of Chinese ALB populations based on the geographical coordinates of each sample. The colors and symbols correspond to those shown in A.

Before DNA isolation, 95% ethanol was used to remove any contaminations from the beetle's surface. For each sample, a single leg or larval tissue was cut into smaller pieces, and the tissue was ground using a mixer mill (Retsch MM400, Germany) at 29 Hz for 1 min. DNA was extracted using the DNeasy 96 Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. The quality and quantity of DNA samples were assessed using the NanoDrop ND-1000 spectrophotometer and Qubit 2.0 Fluorometer, respectively. For each sample, a total

of 1 μ g or 2 μ g DNA was prepared for sequencing.

1.4.2 Genome-wide SNP markers: genotyping and analysis

Reduced representation libraries (RRLs) were prepared and then sequenced at the Institut de Biologie Intégrative et des Systèmes (IBIS; Université Laval, Quebec City, Canada) with the Ion Proton sequencer (Thermo Fisher Scientific, Waltham, MA, USA) following the procedure described in detail by Abed et al. (2019). We modified their protocol by adding NsiI to the PstI/MspI double digest to create a triple digest library and adopted double-size selection steps to increase the efficiency and accuracy of marker discovery (Bayona-Vásquez et al., 2019; Peterson et al., 2012).

The Fast-GBS pipeline v1.0 was adopted to process raw sequencing reads (Torkamaneh et al., 2017). Within this pipeline, SABRE 1.0 was used to demultiplex barcoded reads into separate files (Joshi, 2011), Cutadapt 2.1 to remove adapter sequences (M. Martin, 2011), Burrows-Wheeler Aligner (BWA) 0.7.17 (Li & Durbin, 2010) to align reads to the ALB reference genome which consists of 10,474 scaffolds (GCA_000390285.1) (McKenna et al., 2016), Samtools 1.8 to convert sam files to bam format and index (Li et al., 2009), and lastly, PLATYPUS 0.8.1.1 to call the SNP variants (Rimmer et al., 2014). The resulting SNP variants were filtered using VCFtools 0.1.16 (Danecek et al., 2011). Basic filters were applied to retain only biallelic SNPs, remove indels and variants with a FILTER flag other than PASS, and remove loci with more than 50% missing data. Samples with > 60% missing data were removed. Finally, only loci with read depth > 5, minor allele frequency (MAF) > 0.05, and missing data per site < 10% were retained. We estimated relatedness between samples through the PLINK method of moment (MoM) using the "snpgdsIBDMoM" function in SNPRelate (Zheng et al., 2012) implemented in R (Team, 2015). Samples with an identity-by-descent coefficient exceeding 0.25, which are considered full sibs, were removed. However, genetically related individuals from Korea, where we had limiting sampling, were initially retained for analyses to explore our Korean samples' overall genetic diversity and compare them to the Chinese populations. Following these comparisons, these analyses were repeated without the related Korean individuals. The SNP variant filtering pipeline is outlined in Table S1.2.

Genetic diversity

We calculated population genetic indices such as nucleotide diversity (π), observed heterozygosity (H_o), and expected heterozygosity (H_e) using the function "populations" in STACKS v2.3e (Catchen et al., 2013). We calculated the observed heterozygosity for each individual via VCFtools-0.1.16 and the pairwise F_{ST} between populations via GenoDive 3.04 (Meirmans, 2020). The relationships of H_e or H_o with latitude or longitude were assessed using stat_cor() in R (Team, 2015).

Population structure

We assessed population structure among native ALB populations using three complementary multivariate ordination methods: principal component analysis (PCA) (Price et al., 2006), discriminant analysis of principal components (DAPC) (Jombart, Devillard, & Balloux, 2010), and spatial principal component analysis (sPCA) (Jombart et al., 2008). We used PCA to demonstrate genetic similarities among individuals and then DAPC to delimit distinct genetic clusters. Finally, we used sPCA to explicitly incorporate spatial information to quantify geographic patterns of genetic variation. We performed all three analyses using adegenet v2.1.2 (Jombart, 2008). For PCA analysis, we used the "glPca" function with retaining 20 PCs. For DAPC analysis, we first ran the "find.clusters" function to determine the number of groups that best summarize variations in the data running from $K = 1$ to 16. We then used the "dapc" function while retaining 250 PCs and five discriminant analysis eigenvalues. In the sPCA analysis, we ran the "spca" function by choosing a Delaunay triangulation as a connection network and kept the first positive eigenvalue. To assess similarities between geographical and genetic distributions, we performed a Procrustes analysis and a Mantel test (Lisboa et al., 2014). We used the "procrustes" function in the R package MCMCpack 1.4-7 (A. Martin et al., 2011) to perform a Procrustes transformation analysis on the first two PCs identified in the PCA analysis, in which the PCA coordinates were scaled to geographical coordinates. We determined the Pearson correlation coefficient between the pairwise genetic distance (F_{ST}) and the geographical distance between populations using stat_cor() in R.

Based on the distinct groups identified by DAPC analysis, we performed an analysis of molecular variance (AMOVA) using an infinite allele model in GenoDive 3.04 (Meirmans, 2020) to separate total genetic variance into among-individuals within-populations, among-populations within-groups, and among-groups covariance components.

Contemporary movement of ALB was previously hypothesized by authors (Carter et al., 2009; Javal, Lombaert, et al., 2019), therefore, we wanted to quantify levels of admixture within each population. We used a sparse non-negative matrix factorization (sNMF) approach to estimate individual ancestry coefficients using LEA 2.0 (Frichot & François, 2015). This method allowed us to estimate homozygote and heterozygote frequencies and avoid the assumption of Hardy-Weinberg equilibrium (HWE) (Frichot et al., 2014). We ran ten replicates with K range 1 to 16 using the "snmf" function. It used a cross-validation approach to estimate the entropy of each K , where the minimum K value is the best estimate. We then ran "snmf" again with the ascertained K and an alpha-value (regularization parameter) of 100 to estimate the individual ancestry coefficients. We also computed a maximum likelihood (ML) phylogeny using 1,000 bootstrap replicates under the GTRGAMMA model in RAxML v8.2.9 (Stamatakis, 2014). We visualized the ML tree with FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Gene flow and population history models

We calculated recent gene flow among populations with a Bayesian Markov Chain Monte Carlo (MCMC) resampling method in BayesAss 3.04, which estimates recent migration rates between populations as the proportion of migrants per generation (Mussmann et al., 2019). First, we ran 1,000,000 iterations without burn-in and then adjusted the parameter for allele frequencies to 0.3 to align with the suggested acceptance rate of 20% 60% (Wilson & Rannala, 2003). Then, we performed three independent runs using 10 million iterations with a burn-in of one million and a sampling interval of 1,000. We used the default settings (0.1) for the mixing parameters associated with proposed moves (i.e., inbreeding coefficient and migration rates). We assessed convergence of the estimated parameters by plotting the combined trace plot and the probability distribution using Tracer 1.7.1 (Rambaut et al., 2018).

We used Bayesian inference with the structured coalescent estimator Migrate 4.4.4 (Beerli et al., 2019) to test population history models for the four regions N, S, NE, and NW based on our population structure results. Thus, the following different population scenarios were evaluated: seven scenarios (Figure S1.1, models 1-7) that assumed the most recent common ancestry occurred in the N or NW regions based on their high genetic diversity, and among which, four of the scenarios (Figure S1.1, models 2-5) assumed that the S and NE populations were from different sources; we included two admixture models (Figure S1.1, models 8-9) to assess whether population admixture has occurred within the North region, given our admixture result. We randomly selected 20 individuals from each region and ran the models with a strictly filtered SNP dataset that contained no missing data ($n = 422$ SNPs). We used one long chain with four replicates for each run with 1,000,000 iterations and a burn-in of 10,000. We compared the models using the log marginal likelihood (Beerli et al., 2019). To further estimate effective population sizes, divergence time, and migration rates of the optimal model identified with Migrate, we used an alternative coalescent-based simulation method, Fastsimcoal 2.6 (Excoffier & Foll, 2011). We applied easySFS.py (<https://github.com/isaacovercast/easySFS>) to generate the joint site frequency spectrum (SFS). We performed 100 independent parameter estimations to obtain the maximum composite likelihood of the joint SFS, each run with 100,000 iterations and 40 cycles of expectation conditional maximization (ECM) (Lanier et al., 2015) using a maximum likelihood method. The mutation rate μ was set to 2.9e-09 per base and per generation based on *Heliconius Melpomene* (Keightley et al., 2015), assuming a 1-year life cycle (generation time) for ALB (Hu et al. 2009). The best run was selected based on the highest maximum likelihood. We computed the 95% confidence intervals of the parameter estimates of the best run by simulating 50 SFS of the estimates and re-estimating parameters each time (Lanier et al., 2015; Y. Zhao et al., 2019). PGDSpider 2.1.1.5 was used to convert the vcf file to the required file formats for the above analyses (Lischer & Excoffier, 2012).

Selection detection

We used the F_{ST} -based genome scan method OutFLANK v0.2 (Whitlock & Lotterhos, 2015) to detect loci putatively under selection. It calculates the likelihood based on a trimmed distribution of F_{ST} values to infer the F_{ST} distribution for neutral markers. We used a q-value of 0.05 to identify outlier loci. We also used fsthet v1.0.1 (Flanagan & Jones, 2017) as an alternative method to detect loci under selection in R. It calculates F_{ST} and expected heterozygosity values for individual SNPs, from which it determines smoothed quantiles to identify loci with elevated or low-lying F_{ST} values relative to their heterozygosity. Here, we designated a confidence level of 95% to detect F_{ST} outliers. We annotated outlier SNPs using SnpEff v5.030 (Cingolani et al., 2012). The SnpEff database for ALB was built manually with the reference genome GCA_000390285.1 (McKenna et al., 2016). The flanking sequence for each locus identified was compared to the NCBI database using blastn to identify similar insect genes (Altschul et al., 1990), and details for the annotated regions were derived from insect-based records from the Gene Ontology database (Ashburner et al., 2000; Gene Ontology Consortium 2021), UniProt database (UniProt Consortium 2021), and FlyBase (Larkin et al., 2021) to determine molecular function and biological processes. We identified a potentially adaptive outlier SNP within glycerol kinase and conducted a phylogenetic analysis for related insect genes using the Neighbor-joining method with 1,000 bootstraps in the MEGA X software (Kumar et al., 2018). The allele frequencies of the candidate SNP across all populations were visualized in R (Team 2015).

Microsatellite discovery and analysis

To independently assess population structure with a different set of data, we genotyped 232 new ALB specimens collected independently from 16 locations in China (Table S1.3) within the biogeographic regions (i.e., NW, N, NE, S) using highly polymorphic microsatellite markers. Here, we sampled within the same regions, but with limited overlap with localities of the GBS study, with only five duplicated sampling locations, Beijing (BJ), Hengshui (HS), Cixi (CIX), Tongliao (TOL), and Yanchi (YC). DNA was extracted using the same protocol outlined above. Following DNA extraction, we mined polymorphic microsatellite motifs from clean reads obtained from ddRAD sequencing, based on a library of 25 freshly preserved specimens. For ddRAD sequencing, we used restriction enzymes *SbfI* and *MspI* while simultaneously ligating P1 (*SbfI*) and P2 (*MspI*) adaptors to the fragmented DNA and targeting large fragments (sizes selected for 350–550 bp) for 2×300 bp paired-end sequencing on an Illumina MiSeq platform (Cornell University's Biotechnology Resource Center). After an initial quality check using FastQC, raw Illumina reads were assembled into contigs (unique consensus sequences from multiple reads) using NGen (v.11. DNASTAR. Madison, WI) and default parameters. The resulting contigs were used as the reference file to re-align all raw Illumina reads to assess coverage and the presence of microsatellite loci for each contig.

We assembled over 26,500 contigs and discovered around 1,300 non-duplicated microsatellite loci, including dimers, trimers, and tetramers. Only tetramers were developed for the subsequent multiplex assay because they normally are less prone to stuttering caused by Taq polymerase slippage (DeWoody et al., 2006). Primers were designed for 69 tetramer microsatellite loci for genotyping using BatchPrimer3 (You et al., 2008) with default parameters. They were divided into four multiplexes. PCR conditions consisted of 1 min initial denaturation at 94 °C, followed by 40 cycles with 15 sec denaturing at 94 °C, 15 sec annealing at 55 °C, 30 sec extension at 72 °C, and a final extension time of 10 min at 72 °C. Raw genotypic data were trimmed to exclude microsatellite loci with > 20% missing data and specimens with > 8 (15%) missing loci. The final dataset included 53 microsatellite loci (primer sequences available in Table S1.4) for 202 individuals (1–32 individuals per location, average 12.6). We assessed linkage disequilibrium between all pairs of microsatellite loci in the R package *genepop* 1.1.7 (Rousset, 2008) to avoid genetically linked loci.

Mean number of alleles per locus (A_n), average gene diversity over loci (D), observed heterozygosity (H_o), and expected heterozygosity (H_e) under HWE were estimated in Arlequin 3.5 (Excoffier & Lischer, 2010). The probability test for deviation from HWE was calculated in the R package *genepop* 1.1.7 (Rousset, 2008). We used the R package *PopGenReport* 3.0.4 (Adamack & Gruber, 2014) to estimate mean allelic richness (A_s) for each location and pairwise F_{ST} after correcting for sample size differences. Samples with fewer than five specimens were omitted from the estimation.

We applied DAPC to characterize genetic clusters using Ward's hierarchical clustering method (Ward, 1963). We further assessed the distribution of genetic variance between and within genetic clusters as assessed through an AMOVA based on the clusters according to our DAPC results. Statistical significance of covariance associated with each hierarchical level was calculated with 1,000 permutations.

Congruence between genome-wide marker sets

To assess the congruence between the two marker sets, we performed a Mantel test using a Monte-Carlo method on the pairwise F_{ST} generated from SNPs and microsatellites in *ade4* (Chessel et al., 2004). We used F_{ST} values calculated from the five sampling locations (i.e., BJ, HS, CIX, TOL, and YC) present in both datasets. We performed the ‘mantel.rtest’ analysis on the two F_{ST} matrices with 9,999 permutations.

Population assignment

To identify markers for downstream diagnostic tool development, we wished to identify SNPs that could accurately assign individuals to their source populations. To do so, we used Mycorrhiza 0.0.28, a genotype assignment software that employs machine learning and phylogenetic networks to identify informative SNPs that can assign individuals to their source population

(Georges-Filteau et al., 2020). The SNPs were ranked by discriminatory power based on mutual information. Mycorrhiza works in two steps. First, it generates a pairwise genetic distance matrix from the genotype data and uses it to construct a phylogenetic split system using the Neighbor-Net method in the program SplitsTree 4.14.6 (Huson, 1998; Huson & Bryant, 2006). Second, the split system is used in a two-fold cross-validation procedure via a random split method in the scikit-learn library.

With this approach, we generated two SNP assignment sets: (1) North vs South, as defined by our sPCA results; (2) Six regional groups, as defined by our DAPC results (Figure S1.2).

1.5 Results

1.5.1 SNP genotyping and analysis

We obtained 524 million reads for 480 individuals ($\approx 1M$ per individual) from GBS and identified 664,178 variants with the fastGBS pipeline. After variant filtering (details in Tables S1.1, S1.2), we retained 6,102 SNPs and 365 individuals (359 from China and six from South Korea).

Population genetic diversity

H_o in the Chinese populations ranged from 0.222 ± 0.003 to 0.306 ± 0.002 , with the highest H_o in populations around $35\text{--}40^\circ\text{N}$ and the lowest in Northeast China. This contrasts with the low heterozygosity we observed in our Korean samples, Kangwon (KOR, $H_o = 0.019$), and Pocheon (KNA, $H_o = 0.021 \pm 0.001$) (Table 1.1). However, 31 Korean individuals in the populations Kangwon (KOR) and Pocheon (KNA) were identified as full siblings by SNPRelate. The H_o was lower among most Korean samples compared to Chinese ALB (Figure S1.3). Expected heterozygosity and nucleotide diversity in Chinese populations showed patterns similar to H_o with peak values $35\text{--}40^\circ\text{N}$ and the lowest values in the northeast. Similarly, H_e and nucleotide diversity values in Kangwon (KOR) and Pocheon (KNA) were lower than for all Chinese populations (Table 1.1). Among Chinese samples, we observed significant isolation-by-distance (IBD, $R=0.56$, $p=2.2\text{e-}11$; Figure S1.4A) with clear latitudinal and longitudinal patterns among sampling sites (H_o/H_e , Figure S1.4B-E). Both H_o and H_e were negatively correlated with longitude, with heterozygosity decreasing from west to east, although these relationships were insignificant. Northern populations showed low H_o and H_e , but values increased with decreasing latitude until 37°N , then dropped again for the southernmost site (Figure S1.4B, D). However, this relationship appeared to be driven by a single population, Cixi (CIX), and when we removed this population from the analysis, the latitudinal cline was no longer significant (Figure S1.5).

Table 1.1: Asian longhorned beetle sampling and basic population genetic analyses using 6,102 SNPs, 16 locations in China and four locations in Korea. Indices are calculated based on within-population sample sizes ≥ 5 ; N , the exact number of samples per population. Ho , observed heterozygosity; He , expected heterozygosity; π , nucleotide diversity. For further sampling details, see Table S1.1; for sampling and basic population genetic analyses involving microsatellites, see Table S1.3.

ID	Locality	Country/Region	Latitude	Longitude	N	Ho	He	π
HRB	Harbin, Heilongjiang	Northeast China	45.8	126.54	21	0.226	0.233	0.24
CHC	Changchun, Jilin	Northeast China	43.82	125.32	10	0.222	0.223	0.238
TOL	Tongliao, Inner Mongolia	Northeast China	43.65	122.24	58	0.228	0.242	0.244
YJ	Yanji, Jilin	Northeast China	42.66	129.44	25	0.263	0.282	0.288
SHY	Shenyang, Liaoning	Northeast China	42.05	123.36	33	0.244	0.251	0.256
QI	Qingtongxia, Ningxia	Northwest China	38.02	106.08	28	0.297	0.302	0.308
YC	Yanchi, Ningxia	Northwest China	37.75	107.4	24	0.3	0.301	0.307
CHE	Chengde, Hebei	North China	40.98	117.95	17	0.283	0.274	0.284
IMC	Huhhot, Inner Mongolia	North China	40.83	111.66	21	0.261	0.256	0.262
BJ	Beijing	North China	39.75	116.14	10	0.284	0.281	0.296
SHI	Shijiazhuang, Hebei	North China	38.05	114.51	22	0.306	0.302	0.309
HS	Hengshui, Hebei	North China	37.74	115.67	20	0.299	0.303	0.311
JI	Jinan, Shandong	North China	36.65	117.12	12	0.278	0.279	0.293
TA	Taian, Shandong	North China	36.2	117.09	11	0.284	0.295	0.309
BB	Bengbu, Anhui	South China	32.92	117.39	20	0.297	0.301	0.308
CIX	Cixi, Zhejiang	South China	30.17	121.27	27	0.23	0.246	0.251
KOR	Kangwon	Korea	38.11	128.46	29	0.019	0.015	0.015
KNA	Pocheon	Korea	37.75	127.17	5	0.021	0.016	0.018
INC	Incheon	Korea	37.45	126.7	2	-	-	-
ULS	Ulsan	Korea	35.54	129.31	1	-	-	-

Population genetic structure

In the PCA analysis, we observed Chinese and Korean populations were distinct (three individuals from KOR and KNA, upper right quadrant, Figure 1.2B). However, the three additional Korean samples from INC and ULS, collected in urban areas, were nested within the cluster

containing samples from China.

Chinese ALB populations in the NW, S and NE regions formed distinct clusters, while populations from the N region separated into three clusters, one that was unique, and the others were split among the NE and NW regions (Figure 1.2C-D; Figure S1.6). The sPCA analysis exposed a distinct north-south genetic break ($P\text{-value}=0.001$) (Figure 1.3A), which was consistent with the F_{ST} values, where populations in the north were all highly divergent from those in the south (Table S5). DAPC analysis (Figure 1.3B) further resolved distinct subclusters among ALB populations within the northern and southern regions, with two distinct populations in the Northeast (SHY/CHC/HRB/YJ and TOL), one in North China (IMC/BJ/CHE), one in Northwestern China (QI/YC/SHI), and two along the eastern coast (HS/JI/TA/BB and CIX). We used these DAPC groupings in our AMOVA (i.e., six DAPC groups; refer to Figure S1.7A for BIC) and identified significant among-group (14.8%) and among-population within-group (7.9%) genetic covariances (Table S1.6; $p < 0.05$).

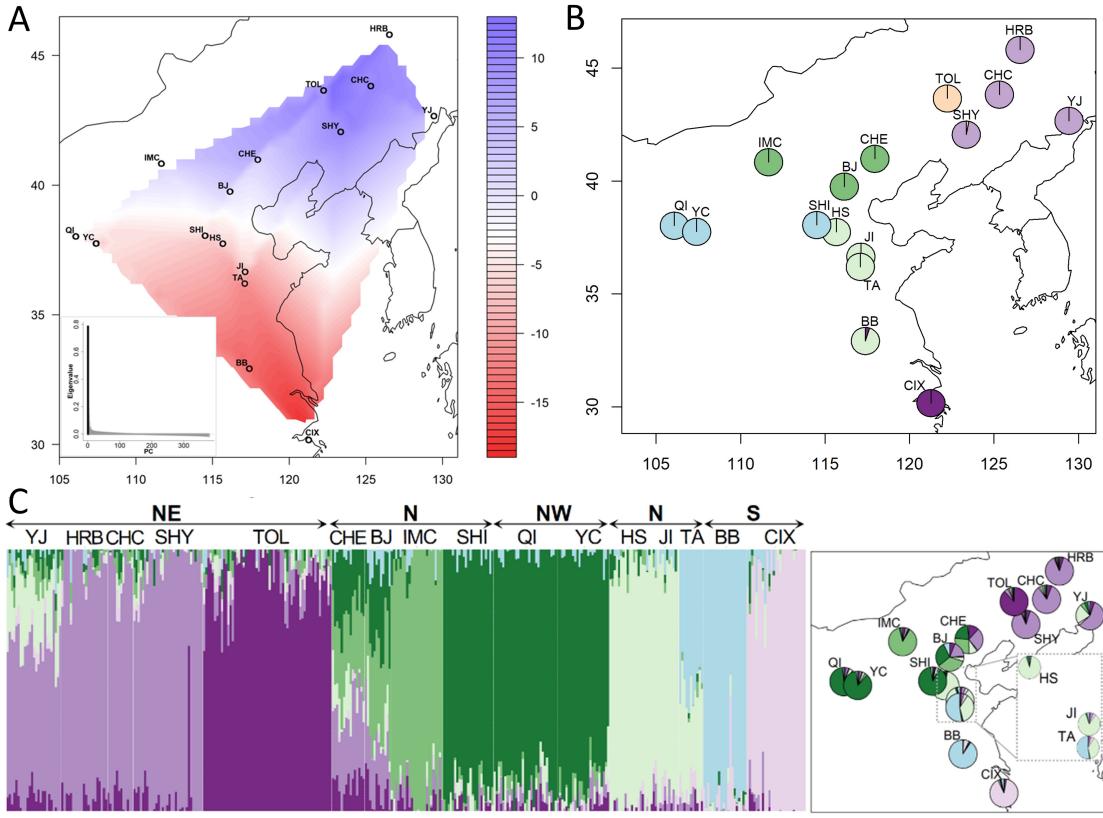


Figure 1.3: Spatial PCA, DAPC and Admixture. **A** Interpolated gradient map of the spatial genetic structure of the Asian longhorned beetle based on the first positive eigenvalue of spatial principal component analysis (sPCA). Eigenvalues of sPCA are displayed on the inset. Positive eigenvalues correspond to global genetic structure, and negative eigenvalues indicate local genetic structure. The color ramp shows values for lagged scores. **B** DAPC analysis using SNPs of 16 sampled ALB populations in China. Map with the 16 populations color-coded as per the DAPC population clustering results. **C** Bar plot and interpolated map of ancestry coefficients across 16 populations ($k = 7$). A total of 6102 SNPs were used in the analysis. Each color-coded pie represents one of the 16 sampled ALB populations in China. Population codes are explained in Table 1.1. A total of 6102 SNPs were used.

Individual admixture estimates identified eight distinct populations (Figure 1.3C) based on the cross-entropy criterion (Figure S1.7B). Admixture analyses showed that Chengde (CHE) and Beijing (BJ) (N region), as well as Yanji (YJ) (NE region), have the most complex admixture patterns, a result supported by relatively low F_{ST} values for these populations (Table S1.5). We also show that Yanchi (YC) and Qingtongxia (QI) (NW region), and Shijiazhuang (SHI) (N region) show similar genetic composition, further supporting our PCA, DAPC, and F_{ST} results. Finally, we observed similar genetic composition among Harbin (HRB), Changchun (CHC), and Shenyang (SHY) populations (NE region).

The ML phylogenetic tree (Figure 1.4) revealed that subregions NW, NE, and S formed distinct, well-supported lineages. However, the N region was polyphyletic relative to the other subregions. Individual populations formed distinct, well-supported lineages with a few exceptions. Within the NW, individuals from Qingtongxia (QI), Yanchi (YC), and individuals from Shijiazhuang (SHI, from the N region) formed a single cluster. Both Chengde (CHE) and Yanji (YJ) did not form single lineages; instead, these populations were divided into several distinct clades. In the S subregion, the Bengbu (BB) population formed a distinct lineage, except for a single individual which grouped with Cixi (CIX), a population located further to the south (Figure 1.4, see also Figure 1.2A).

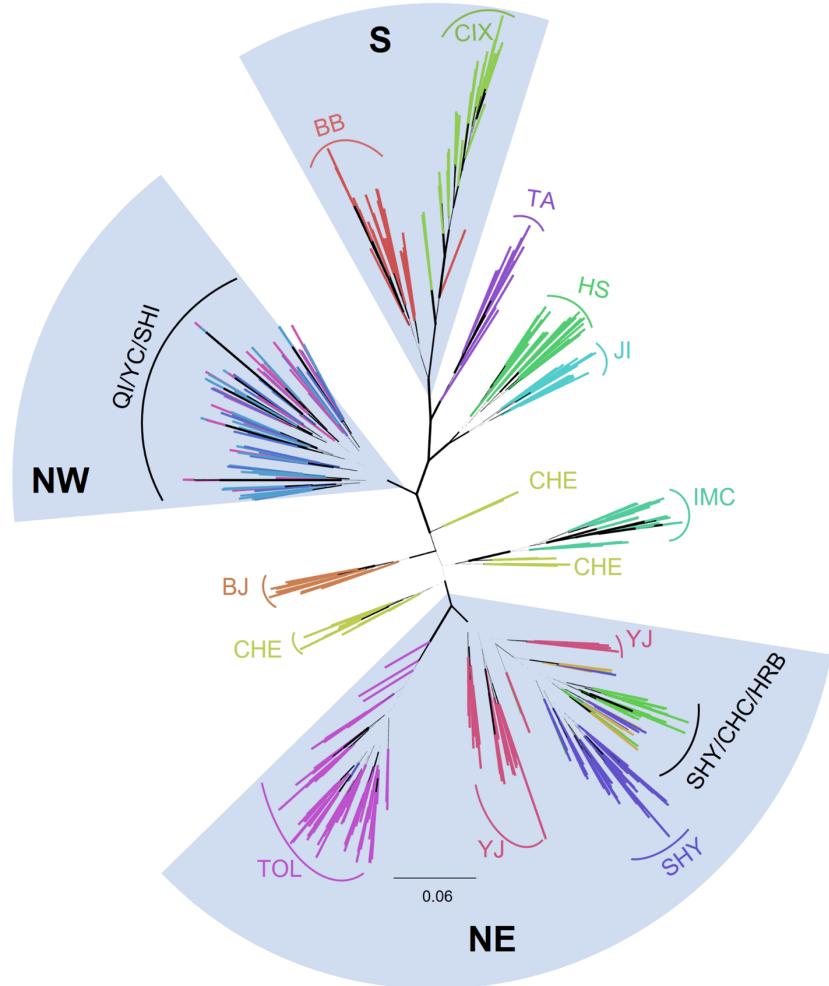


Figure 1.4: Maximum likelihood phylogenetic tree for Chinese Asian longhorned beetles analyzed in RAxML. A total of 6102 SNPs were used in the analysis. Each branch represents a sample. The branches are colored by their population codes (explained in Table 1.1). The width of each branch corresponds with their bootstrap value (widest branch with a bootstrap value of 100%). Fan-like sections indicate subregions NE, S, and NW (whereby SHI is actually in the N region). The rest of the populations are from the N region.

Gene flow and population history

We estimated contemporary gene flow within and among regions using our SNP data set. The migration rates ranged from 0.004 ± 0.004 to 0.124 ± 0.060 migrants per generation (Nm), with an average across all populations estimate of 0.01 ± 0.0006 Nm (Figure 1.5A). We see moderate gene flow among sites within the NW (0.104 ± 0.046 Nm) and NE (Shenyang (SHY)→Changchun (CHC)) and Harbin (HRB)→Changchun (0.058 ± 0.032 Nm)). The highest rates of gene flow occurred between sites in the NW (Qingtongxia (QI) and Yanchi (YC)) to Shijiazhuang (SHI) in the N (0.124 ± 0.060 and 0.031 ± 0.019 Nm, respectively). The trace plot and probability distribution are shown in Figure S1.8.

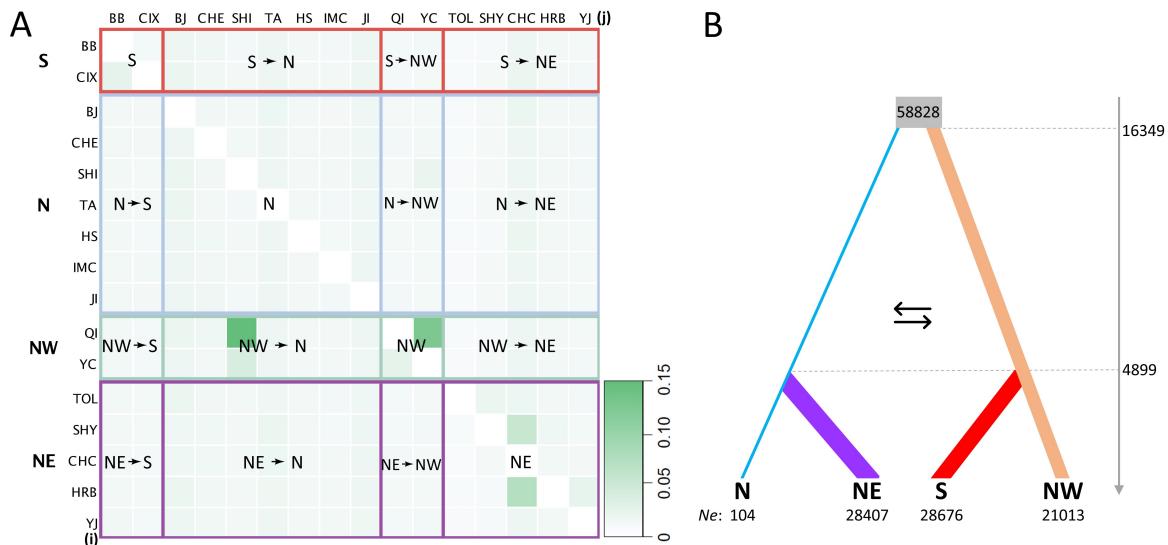


Figure 1.5: Migration rates and population history model. **A** Migration rates among the 16 Asian longhorned beetle populations sampled throughout China. The migration direction reads from populations in vertical (i) to populations in horizontal (j) order. Light to dark green color represents the migration rates from low to high. Contemporary gene flow within and between S, N, NW, and NE groups are indicated. Gene flow within each population is shown along the diagonal with empty values. A total of 6102 SNPs were used in the analysis. Population codes are explained in Table 1.1. **B** Population history model. The ancestral population is shown in gray. The estimated population divergence time (vertical axis) and effective population sizes (Ne values) are indicated in the figure. The width of each column reflects the relative effective population size.

Out of the nine scenarios that included present-day populations tested in Migrate (Figure S1.1), model 3 was the most probable based on marginal likelihood (Table S1.7). This model assumes that the North population first diverged from the Northwest population with ongoing gene flow and an estimated migration rate of 5e-03 between populations (Fastsimcoal output). Through modelling by Fastsimcoal, the divergence between the North and the Northwest populations was traced back to 16,349 generations ago (Figure 1.5B). Furthermore, the South population diverged from the Northwest population, and the Northeast population diverged from the North population (see Figure 1.2A for delineation of regions). The respective splits of the Northeast and the South populations were estimated to have occurred 4,899 generations ago (Figure 1.5B). The highest effective population size (N_e) was predicted within the hypothetical ancestral population ($N_e = 58,828 \pm 4,449$), with a moderate reduction in N_e following divergence of the NE ($28,407 \pm 4,711$), NW ($21,013 \pm 4,509$) and S ($28,676 \pm 4,749$) descendant populations. The population of ALB in the North exhibited surprisingly low ($N_e = 104 \pm 4,796$) (Table S1.8).

Selection detection

We did not detect outliers in OutFLANK using a q-value cutoff of 0.05, while we did detect SNP loci putatively under selection through *fsthet* (Figure S1.9; Table S1.9). Among these 360 outliers (48% under positive selection; 52% under balancing selection), there were 185 (51%) intergenic variants, and 172 variants (48%) were annotated as related to protein coding (of which 29% were intron variants; 37% in upstream or downstream regulatory regions; 34% directly in the coding DNA sequence CDS). Among the F_{ST} outliers directly associated with genes, we found SNPs within 134 genes. For 26 genes, multiple outlier loci were detected (i.e., two to four within a single gene). Of those 58 SNP outliers directly within the CDS, there were 31 synonymous variants and 26 missense variants, and one splice variant. We assigned functional annotations to each locus based on flanking sequence similarity with previously published results. Over half of the flanking sequences (n=213) could not be characterized. For loci that were functionally similar to previously annotated regions, we identified loci associated with a range of molecular functions and biological processes (Table S1.9). Of particular interest was the glycerol kinase (GLK) gene (AGLA000593, i5K database, https://i5k.nal.usda.gov/Anoplophora_glabripennis, Accessed Jan 2022). The SNP under selection was annotated as a missense mutation under positive selection ($F_{ST} = 0.335$) (Table S1.9). We compared the ALB GLK gene to previously published insect GLK genes. We observed that the ALB GLK gene is closely related to genes described in other beetle taxa (Figure 1.6A), including *Dendroctonus ponderosae* and *Tribolium castaneum*, in agreement with the species' phylogenetic relationships shown in McKenna et al. (2016). We plotted the allele frequency of the SNP (A/G) and showed that the allele frequency exhibits a distinct clinal trend within our ALB populations (Figure 1.6B; for allele count details, see Table S1.10).

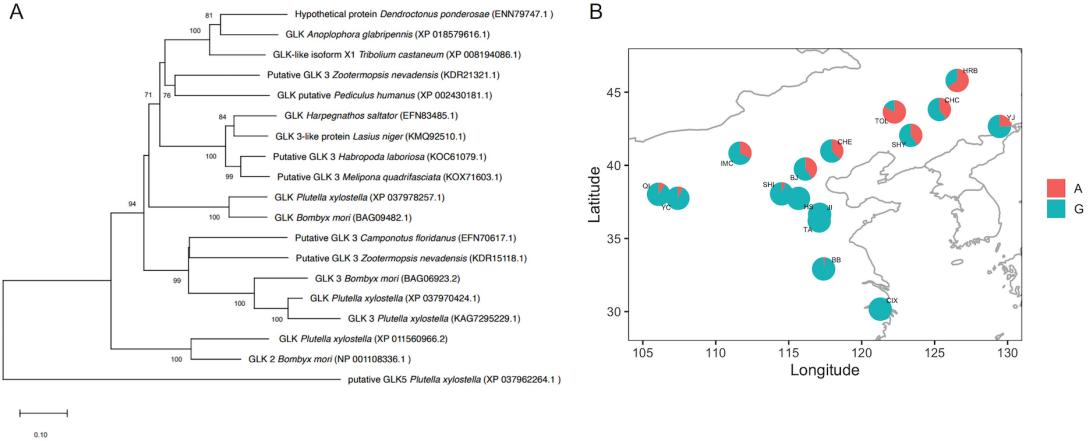


Figure 1.6: Exploring the potential functional relevance of a missense mutation under positive selection ($F_{ST} = 0.335$) within the *Anoplophora glabripennis* glycerol kinase gene *AGLA000593*. **A** Phylogenetic tree of several glycerol kinase proteins from different insect species, including *A. glabripennis*. All NCBI accessions are provided. The neighbor-joining tree was done in MEGAX (with 1000 bootstraps). Only the branch supports above 70% are indicated. The closest homology for the ALB protein was found with *D. ponderosae*. **B** Geographic map of the allele frequency distribution for the glycerol kinase gene *AGLA000593* across the 16 Chinese ALB populations studied by GBS technology. G refers to the reference allele and A to the alternative allele (the missense mutation).

1.5.2 Microsatellites analyses

After correcting for sample size variation, allelic richness is similar across locations (Table S1.3). A slight deficiency of heterozygotes was observed in all populations (all highly significant, $P < 0.01$). Linkage disequilibrium was minimal among those 53 microsatellite loci, as only six out of all 1,378 (0.44%) pairwise comparisons had significant false discovery rate-adjusted P -values. LD was also checked for within each population. There were no loci pairs found showing LD in every population, and the maximum number of populations that shared an LD loci pair was seven, which was less than half of the total number of populations.

The Ward algorithm did not determine a single optimum number of genetic clusters and found approximately equal support for models from a range of $K = 4$ to 6 (24 PCs retained), with $K = 5$ having the lowest Bayesian Information Criterion (BIC) score (Figure S1.7C). When the native populations were assigned into five genetic clusters, cluster 1, the largest, was formed mostly by specimens from North China. Three populations outside North China, Dalian (DAL), Pengzhou (PEZ), and Zunyi (ZUY), were genetically embedded in cluster 1 (Figure 1.7). The NE populations Tongliao (TOL) and HuiChun (HUC) (cluster 2) and the eastern coastal population Cixi (CIX, cluster 3) can clearly be distinguished from cluster 1 on the second discriminant function. Two NW populations Yanchi (YC) and Pengyang (PEY), merely 200 kilometers apart, were very distinct, forming the separate clusters 4 and 5,

respectively. For the model with $K = 4$, Yanchi (YC) merged with cluster 1. And if clustering was increased to $K = 6$, an additional cluster was formed by seven specimens from Zunyi (ZUY) that became separated from cluster 1. However, another five specimens collected in the same location remained with cluster 1. For the AMOVA with $K = 5$, most of the genetic variance (84.56%, $P < 0.001$) is found within locations, with 7.22 % ($P < 0.001$) of the variance among locations within clusters, and 8.22% ($P < 0.001$) of the total variance among clusters.

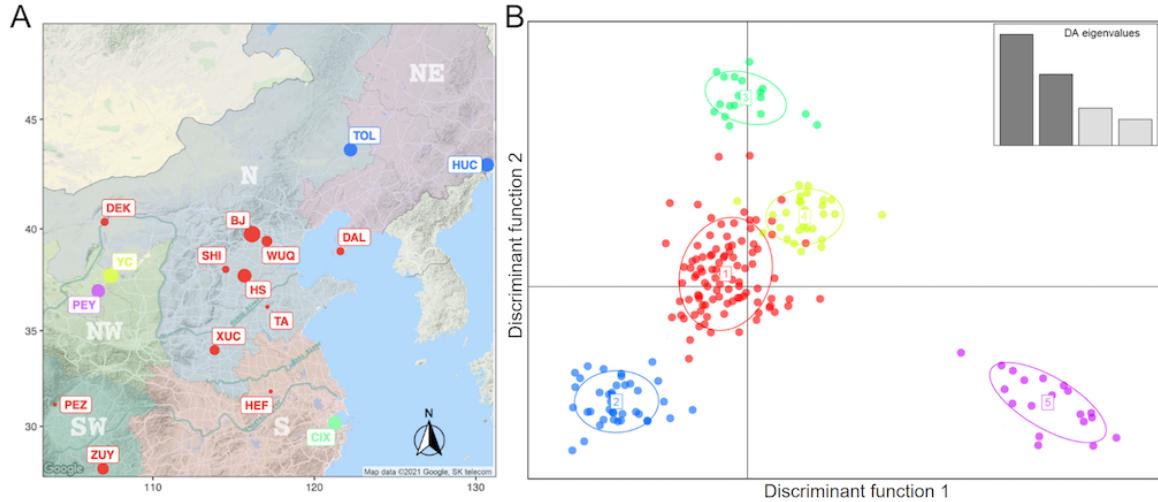


Figure 1.7: Discriminant analysis of principal components (DAPC) of the Asian longhorned beetle using microsatellites. ALB sampling for microsatellites analysis covers different regions in China. **A** Sampling locality. The color for each locality corresponds to its cluster in DAPC. The bubble size of each population is relative to the sample size. The codes for the populations are shown in Table S1.3. **B** Five clusters were identified in DAPC.

1.5.3 Congruence of genetic markers

We used a Mantel test between the two F_{ST} matrices generated from SNPs and microsatellites to assess congruence between our marker and specimen sets (Tables S1.5,S1.11). Here, we observed a significant congruence between our data sets ($r = 0.951$, P -value = 0.039) based on 9,999 replicates between the two datasets. We observed similar, although not identical, clustering in our two DAPC analyses on these distinct data sets (Figures 1.3B and 1.7). Furthermore, we also found that the positions of the N-S break between regions in the two sPCA analyses were the same (Figure 1.3A; Figure S1.10).

1.5.4 Population assignment using SNP markers

The accuracy of assignments to different groups was conducted using an increasing SNP set (up to 500 SNPs; Figure 1.8). We assigned individuals to the N and S groups with $> 90\%$ accuracy using only 20 SNPs that we ranked as the most discriminant and with near 100% with 100-500

SNPs (Figure 1.8A). Assigning individuals to one of the six DAPC groups reduced accuracy to 96.4% with 200 SNPs (Figure 1.8B). Assignment accuracy for each population assignment class using 500 SNPs was shown in Figure S1.11. When the whole set (6,102 SNPs) was used in the tests, accuracies were 100% (two groups) and 98.6% (six groups), respectively.

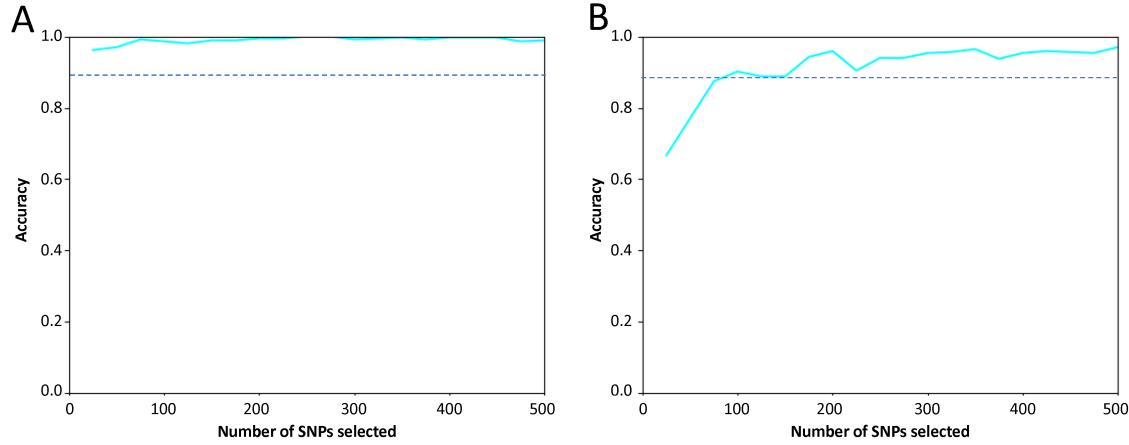


Figure 1.8: Prediction accuracy of Asian longhorned beetle individual assignment with an increasing number of SNPs selected (ranging from 20 to 500). **A** ALBs were assigned to two groups as identified in sPCA. **B** ALBs were assigned to six groups identified in DAPC. Dashed line indicates 90% accuracy of assignment

1.6 Discussion

The Asian longhorned beetle is a widespread pest found throughout the temperate forests in East Asia. Global spread of this species underlies the urgent need for biosurveillance tools that can trace its pathways of invasion and monitor its invasion dynamics. However, such tools require detailed knowledge of the population structure of ALB in its native range. Here, our results show pronounced regional differences, indicative of historical biogeographic structuring, as well as evidence of contemporary migration between China and South Korea and within regions in China. We also identified several genomic regions under selection and that may underlie adaptive differences between populations. Our data also provide a foundation for future genomic-based surveillance tools. To that end, we identified smaller sets of SNPs that accurately assigned individuals to regions and sub-regions within the native range. We further highlight how these markers can be used to assign individuals to regions within the native range.

East Asia has a complex biogeographic history. Temperate forests and their associated communities experienced repeated fragmentation and range shifts due to historic changes in climate and sea levels, combined with diverse topography in the region. These biogeographic processes created ideal conditions for the evolution of high inter- and intraspecific diversity (Bai et al., 2010; Chen & Lou, 2019), which can result, at least partially, from the biogeographic history of the area. We observed some distinct phylogeographic patterns within ALB consistent with earlier publications and patterns seen in other forest species (Carter et al., 2009; Du et al., 2019, 2021). First, we observed a distinct cluster of Korean ALB specimens from forested areas in Kangwon (KOR) and Pocheon (KNA) (except three individuals sampled from urban areas, which will be discussed further below). These populations were distinct from ALB populations in China, consistent with previously published results (Carter et al., 2010; Javal, Lombaert, et al., 2019; Lee et al., 2020) and were collected in the native forest from the northeast region of the country, within the hypothesized historic distribution of the species (Lee et al., 2020). Our limited sample of insects from this forested region showed low genetic diversity, contrary to the more extensive sampling in Lee et al. (2020). Given that our samples from Kangwon (KOR) were collected from a few trees at a rest stop in Korea and proved to be highly related based on kinship estimation performed in SNPRelate (Table S1.12), these can therefore be considered isolated populations with low genetic diversity due to inbreeding and sampling bias. Despite these differences, the unique Korean population structure and genetic diversity we observed in ALB, relative to mainland China, were consistent with phylogeographic patterns observed in other forest species (reviewed in Qiu et al., 2011), including assassin bugs (Du et al., 2019), raccoon dogs (Kim et al., 2013), orchids (Tian et al., 2018) and oaks (Zeng et al., 2015). Baekdudaegan, the main mountain range that runs the length of the Korean Peninsula, has been identified as an important biodiversity hot spot that served as a glacial refugium for boreal and temperate forest species during the last glacial maximum (LGM) (Chung et al., 2018). Sea level changes created and flooded land bridges between the Korean Peninsula and mainland China, isolating species with previously continuous distributions (Qiu et al., 2011). Consequently, ALB populations in this forested region of Korea are genetically differentiated from mainland China through evolutionary processes.

Complex phylogeographic structure is also typical among forest species in eastern China (Qiu et al., 2011). We found that ALB genetic variation was hierarchically structured, with sharp regional genetic breaks and internal population subdivisions with varying levels of admixture. Regionally, ALB populations were divided into northern and southern populations, similar to, but more clearly defined than, results in earlier publications (Carter et al., 2009; Javal, Lombaert, et al., 2019). This North-South population break is seen in a wide range of forest species (reviewed in Qiu et al., 2011), including ALB host plants (*Betula*, Chen & Lou, 2019; *Acer*, Guo et al., 2014; *Populus*, Hou et al., 2018). In our SNP dataset, this North-South division aligned with the hypothesized location of an “aridity belt” along the Yellow River (= Huang He) in East Asia (Milne & Abbott, 2002). Geographic barriers such as mountain chains,

arid regions, rivers, and geographic distance are key factors driving population structure in many widely distributed Asian insect species (e.g., Du et al., 2021).

Population structure can be driven by a combination of geographic, spatial, and environmental conditions. Often, these vary along spatial or temporal gradients; temperature conditions, light regime, and moisture availability can show latitudinal, altitudinal, or longitudinal gradients across a species range (De Frenne et al., 2013; Laiolo & Obeso, 2017). Geographic barriers such as rivers and mountains, particularly when aligned along these longitudinal or latitudinal axes, can profoundly influence population structuring and gene flow within populations (Bester-van der Merwe et al., 2011). We observed latitudinal gradients in ALB with genetic diversity peaking between 35°N to 40°N (Figure S1.4), similar to previous findings in ALB (Javal, Lombaert, et al., 2019), and spotted lanternfly (Du et al., 2021; Zhang et al., 2019). Thus, the higher genetic diversity found for the central populations compared to the edge populations would support the ‘center-periphery hypothesis’ for ALB’s native distribution.

Climate is critical in shaping biogeography within continents (Ficetola et al., 2017; Mckown et al., 2014). For ectothermic species such as insects, temperature is particularly influential in defining species ranges (Angilletta, 2009). Surviving sub-zero temperatures is physiologically challenging and can limit the northward expansion of insects (Brightwell et al., 2010). This hypothesis aligns well with our results, where we observed that the genetic diversity in ALB is negatively correlated with latitude. A notable exception represents the Cixi (CIX) ALB population (30°N) which shows relatively low genetic diversity. Contrary to the other population sampled in the South (Bengbu, 33°N), Cixi is situated south of the Huai River, thus already within the subtropical region with winter temperatures above 0°C. While ALB is found throughout East China, it is generally considered a temperate species (Feng et al., 2016a; Javal et al., 2018; Keena & Moore, 2010), and the greatest population density is north of the 0°C isotherm (Yan 1985). Therefore, although ALB has been found throughout China, their populations may be limited by biotic pressures (Fragnière et al., 2015), hence, limiting their effective population size and leading to the lower genetic diversity observed in the southernmost population CIX, and the northernmost populations HRB and CHC in terms of latitude.

1.6.1 Population history and contemporary movement

The relative contribution of contemporary dispersal and historic population structure represent two opposing forces when seeking to reconstruct the population structure of an organism. Contemporary movement is thought to blur the historic population boundaries that evolved over thousands of years. Contemporary dispersal, as influenced by anthropogenic movement of ALB throughout the native range, was hypothesized to have significantly disrupted the natural population boundaries (Carter et al., 2009; Javal, Lombaert, et al., 2019). We detected evidence of gene flow and admixture within the native range. Globally, gene flow was low be-

tween regions, except for specific sites in the NW and N contributing to gene flow between the two regions. Our population history modelling dated evidence to c.16,300 yrs ago in our most probable demographic scenario. Indeed, such dispersal for ALB might have been driven by changing environmental factors coinciding with the retreat of the glacial coverage that began c.18,000 yrs ago as the end of the LGM. It is probable that ALB experienced similar fragmentation and distribution restrictions as its host plants, allowing congruent phylogeographic structuring to evolve. Then, this regional population structure would have been maintained given ALB's low dispersal ability (Smith et al., 2001, 2004; Williams et al., 2004) and limited movement beyond the natal tree. The second major population split and scattering of ALB was dated to c.5,000 yrs ago, thus following the Holocene climatic optimum in China (6,000 yrs ago), when temperate deciduous forest vegetation reached almost 1,000 km further north (48°N) than at present and extended further southwestwards at higher elevations than today (Yu et al., 2000). As climatic restrictions eased, ALB expanded its range in concert with its host plants, leading to localized admixture between nearby regions while still maintaining distinct regional structure. Similar population structuring, including regional admixture, has been reported in temperate forest trees, including *Juglans species* (Bai et al., 2016) and *Betula platyphylla* (Chen & Lou, 2019). We also observed evidence of movement out of the North to outlying regions in Zunyi (ZUY) and Pengzhou (PEZ) in the microsatellite data set (Figure 1.7), suggesting a possible long-distance spread pathway. Similarly, we detected two distinct genomic signatures in our South Korean samples. We had specimens collected from sites within the historic forested range of ALB (i.e., Kangwon, Pocheon), and three additional individuals collected within the urban centers of Incheon (INC) and Ulsan (ULS). Lee et al. (2020) suggests that ALBs in urban centers do not represent local populations but are the result of long-distance dispersal from China. Our results support these conclusions: individuals in Incheon (INC) and Ulsan (ULS) nest within the Chinese samples, and group with Jinan (JI) and Taian (TA) in Shandong Province, respectively (Figure S1.12) (see also: Javal, Lombaert, et al. (2019); Lee et al. (2020)), with Shandong being an important shipping hub between China and South Korea (X. Li, 2012). As ALB are poor dispersers (Smith et al., 2001, 2004; Williams et al., 2004), the current long-distance movement is likely linked to human activities. Although our findings do not support the hypothesis that contemporary anthropogenic movement was widespread and blurred regional population differences (Carter et al., 2009; Javal, Lombaert, et al., 2019; Javal, Roques, et al., 2019), we were, nonetheless, able to detect cryptic movement of ALB within the native range.

1.6.2 Selection Detection

We identified several loci under selection within our ALB populations. We were particularly interested in glycerol kinase (GLK), an enzyme that catalyzes an important rate-limiting step in the utilization of glycerol at diapause termination (Kihara et al., 2009). Glycerol is composed of two polyols and is a well-known cryoprotectant molecule used by insects to prevent

intracellular ice formation (e.g., Storey and Storey, 2012; Park & Kim, 2013; Cheng et al., 2014). Prior evidence also showed that ALB exhibits both seasonal and population differences in the accumulation of this important cryoprotectant (Feng et al., 2014, 2016b), which coincide with gene expression that peaks during winter (Xu et al., 2021). Moreover, recent laboratory experiments performed on ALB colonies showed that glycerol content changes within the hemolymph occur during diapause and that such changes are related to concurrent shifts in the supercooling point in ALB (Torson et al., 2021). In our study, we found that the missense mutation within the *GLK* gene occurs at much higher frequencies at higher latitudes and follows a clear clinal trend. Based on the observed allele frequency pattern, it suggests that the northern allele variant may provide an adaptive advantage to the northern populations, and further study of the functional implications of this missense variant would be fruitful. We also suggest that this locus be prioritized in future genomic biosurveillance panels (Roe et al., 2019), given its potential link with cold tolerance in ALB. However, accurate SNP calling requires a reference genome, which is not often available for non-model organisms. Moreover, functional genomics research would benefit from a properly annotated genome of the organism under study, especially for the gene space, and we note that over half of the protein coding loci could not be identified. Much of ALB’s genome annotation is still under development because its genome sequence is highly fragmented (McKenna et al., 2016), and an annotated chromosomal assembly would greatly improve detection of other loci of interest and adaptive differences among populations.

1.6.3 Data Set Congruence

Here, we used two different genetic marker systems and slightly different sampling strategies within the main geographic regions to verify the overall genetic pattern of ALB in its native range. In addition to complementary findings, in both cases, the overall genetic structure differentiated NE, NW, and S regions. Given the highly congruent results from the two sets of independent markers, it is worthwhile to consider the pros and cons of the two approaches. The biggest advantage of GBS is undoubtedly its capability to produce thousands of SNPs. Genome-wide SNP markers provide in-depth insights into the level of population structure that may not be revealed by other types of genetic markers. For example, our SNP data show a genetic cline along a latitudinal gradient in Eastern China, a result not reported by Sanger sequencing or microsatellite data (Carter et al., 2010; Javal, Lombaert, et al., 2019). For organisms whose population structure is obscured by recent long-distance dispersal (such as human-aided movement as seen in many invasive species), SNP-based analysis proves to be particularly powerful in resolving the underlying genetic pattern (Picq et al., 2018). Additionally, GBS can generate SNPs that are candidate markers for identifying associations between genotypes and phenotypes that may point to processes of selection or adaptation to novel environmental conditions (Wickland et al., 2017), which can ultimately improve our assessment of invasive risk.

On the other hand, microsatellite markers do have their own merits. Using the designed PCR primers, there is no need to sequence the genomes and run the bioinformatic pipeline again for joint SNP calling when new samples are added to the existing database. Although it is not always possible to compare different data sets produced by different laboratories due to inconsistencies in allele size calling (Vignal et al., 2002), combined with its proven capability to resolve population structures and less requirement on bioinformatics, microsatellite markers will continue playing an important role in population genetic studies.

1.6.4 Developing genomic tools for biosurveillance

Preventing the incursion of invasive forest pests is paramount to an effective biosurveillance program. One of the most critical applications in an effective biosurveillance program is (a) accurate detection of regulated pests to enable an early, rapid response, (b) ability to trace interceptions back to the source and inform trade regulations and policies (Bilodeau et al., 2019; Hamelin & Roe, 2020; Roe et al., 2019). Given the consequences of inaccurate information, it is imperative that the identification and assignment of intercepted invasive species is robust and accurate. As such, genomic tools are becoming the ‘gold standard’ to support regulatory decision-making (Bilodeau et al., 2019; Roe et al., 2019). Our genomic results showed equally strong genetic differentiation among native ALB populations, with limited contemporary migration. Therefore, we were able to exceed 90% predictive accuracy with only 200 SNP markers for the six genetic groups delimited within our dataset. We needed even fewer SNP markers to confidently assign individuals to Northern or Southern regions with 95% accuracy. A study on *Lymantria dispar* spp. also showed that a high degree of genetic differentiation led to high population assignment success (Picq et al., 2018). They used laboratory colonies that consisted of eight groups that were genetically distinct, and the assignment success was generally high (86%–100%) using different sets of SNPs (12, 24, 48, and 96). However, for populations of white bass (*Morone chrysops*) with a moderate genetic differentiation between six populations ($F_{ST} = 0.083$), the assignment accuracy reached 99.78% with 57 SNPs (H. Zhao et al., 2019). Moreover, progress has been made in developing assignment tools using SNP markers. For example, an assay that contains 324 SNPs, allowing for predictions of phenotype, biogeographical ancestry, and male lineage has been demonstrated to be a robust and sensitive human forensics tool (Diepenbroek et al., 2020). Overall, such an approach is also very promising for translating our results into developing SNP panels for a biosurveillance purpose in forest protection. While our sampling design covered most of the regions in China, there are records of ALB throughout China (Yan, 1985), and we have not fully sampled the entire distribution. Given our microsatellite results, the NW population Pengyang (PEY) that is not included in the SNP data set formed a distinct cluster. There are likely other populations with a different genetic make-up that have not been collected. Therefore, it is necessary to consider the possibility of unsampled sources when moving forward with genomics-based biosurveillance.

1.7 Conclusion

Our study revealed clear regional differentiation among native ALB populations using two independent datasets that comprised complementary genetic data and sample sets. Biogeography, drift, and limited dispersal capacity are likely key factors that aided the formation and subsequent maintenance of the ALB population structure within its native range. Future research is poised to explore the environmental factors, including both abiotic (e.g., temperature, precipitation, etc.) and biotic (e.g., host plants), that shaped the complex genetic structure we observed in this pest. The patterns of genetic structure varied among regions and while large-scale human-assisted ALB population migration was limited, we were able to detect it in both datasets. Our ability to resolve this complex pattern of movement in the native range demonstrated the ability for our markers set to serve as a diagnostic tool to track invasive ALB populations outside their native range. Using our genome-wide markers in a diagnostic framework will lead to the development of biosurveillance tools to aid rapid screening and pathway analyses of new ALB interceptions in support of plant protection agencies and regulatory bodies.

1.8 Data availability statement

Raw sequence data are available at the Sequence Read Archive (SRA) with the BioProject ID PRJNA824548. Variant Call Format data have been deposited in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.866t1g1rb>. Scripts are available at <https://github.com/mimingcui/nativeALB>.

1.9 Supplementary materials

The supplemental tables for this chapter are available for access under data repository: <https://datadryad.org/stash/share/fCbWSHakbd2HBPpTFYfbVTBo3UNjnxGNofJlwkukm0>.

All supplementary information can also be found through <https://doi.org/10.1111/eva.13381>.

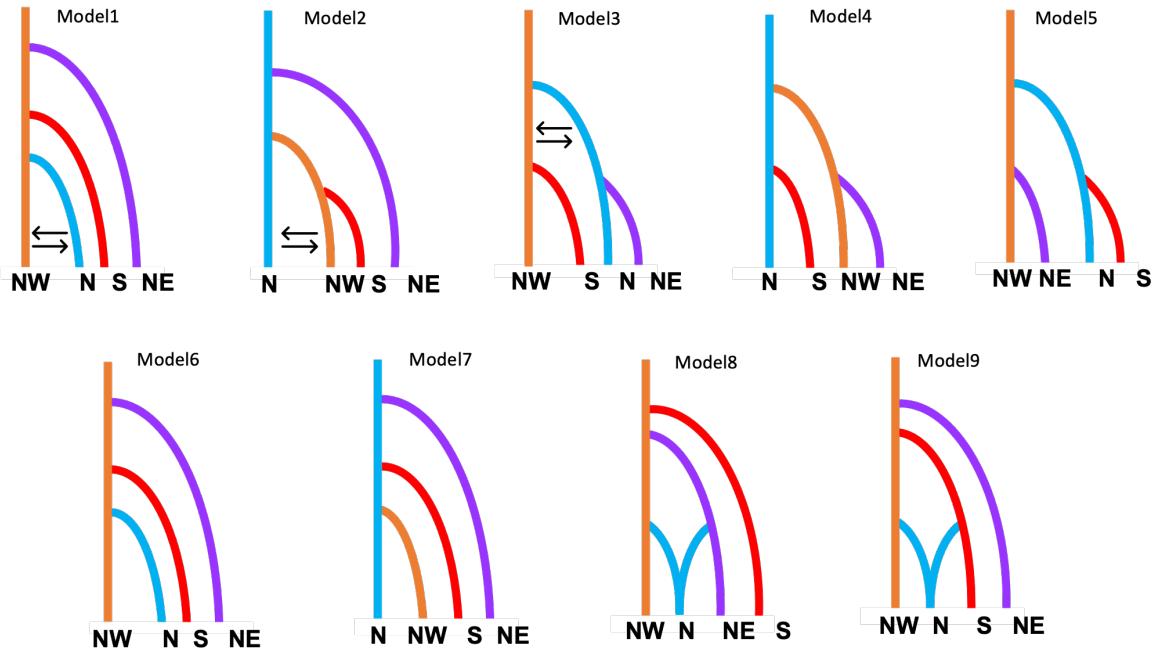


Figure S1.1: Nine demographic models tested in Migrate. 422 SNPs shared by 80 samples were used in the analysis of four groups identified by the PCA, i.e., Northwest (NW, orange), North (N, turquoise), South (S, red) and Northeast (NE, purple). Gene-flow is indicated by arrows. The divergence times between groups were not tested here.

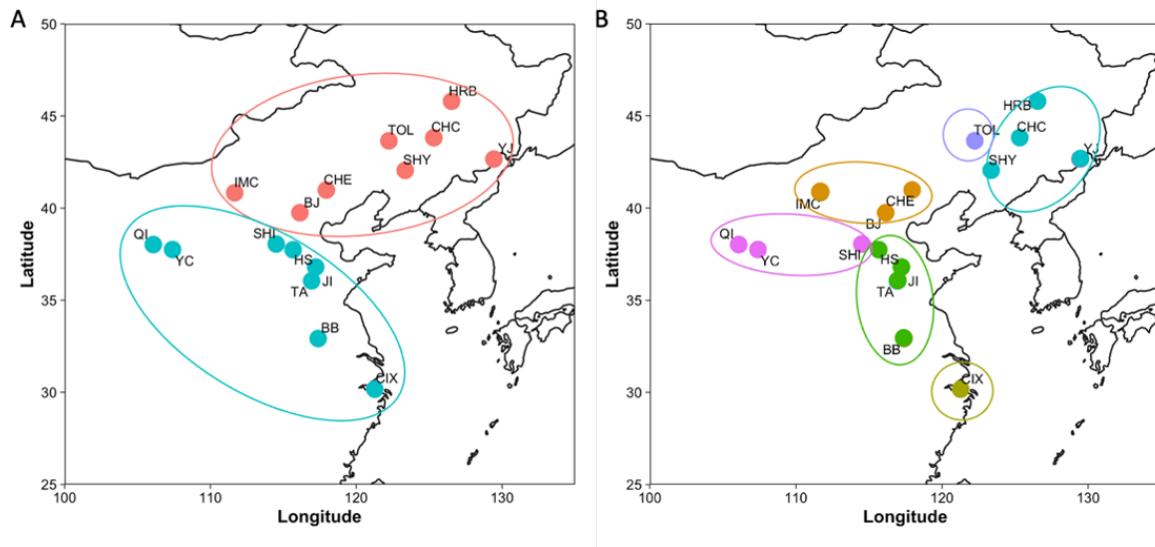


Figure S1.2: Grouping map. **A** Two groups based on sPCA results. **B** Six groups according to groups identified by DAPCs. For population codes, see Table 1.1.

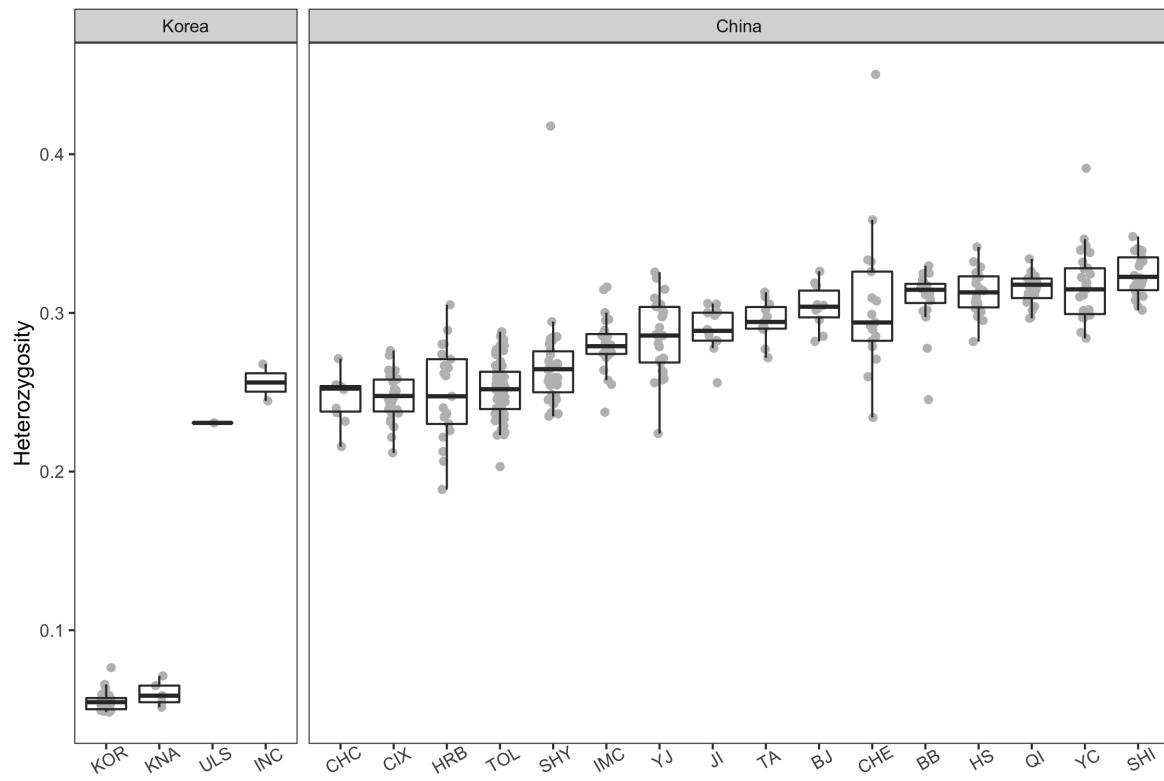


Figure S1.3: Observed heterozygosity plot for ALB individuals. Each dot represents an individual. The information for sampling site is shown in Table 1.1. Populations are ordered by mean values in ascending order. A total of 6,102 SNPs were used in the analysis.

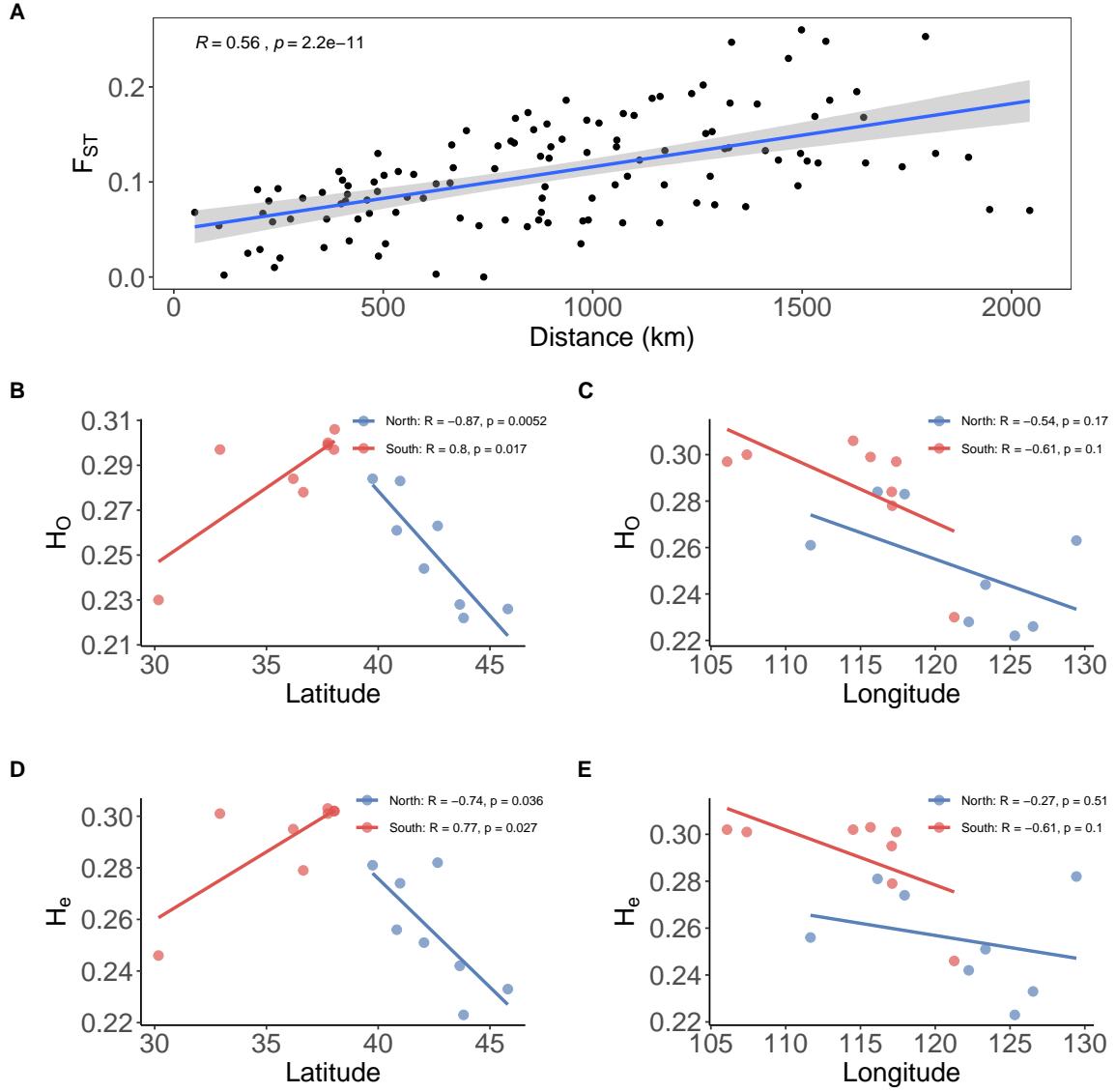


Figure S1.4: Isolation-by-distance (IBD) and Relationships between heterozygosity and latitude/longitude. **A** IBD shown as pairwise F_{ST} between populations against geographical distances. **B, C** Relationships between observed heterozygosity and latitude/longitude. **D, E** Relationships between expected heterozygosity and latitude/longitude. The eight populations in the north identified by sPCA are colour-coded in blue and the other eight populations in the south in red. Colour-coded lines are the trend lines. Pearson correlation coefficient and P-value are presented on the top right of each graph.

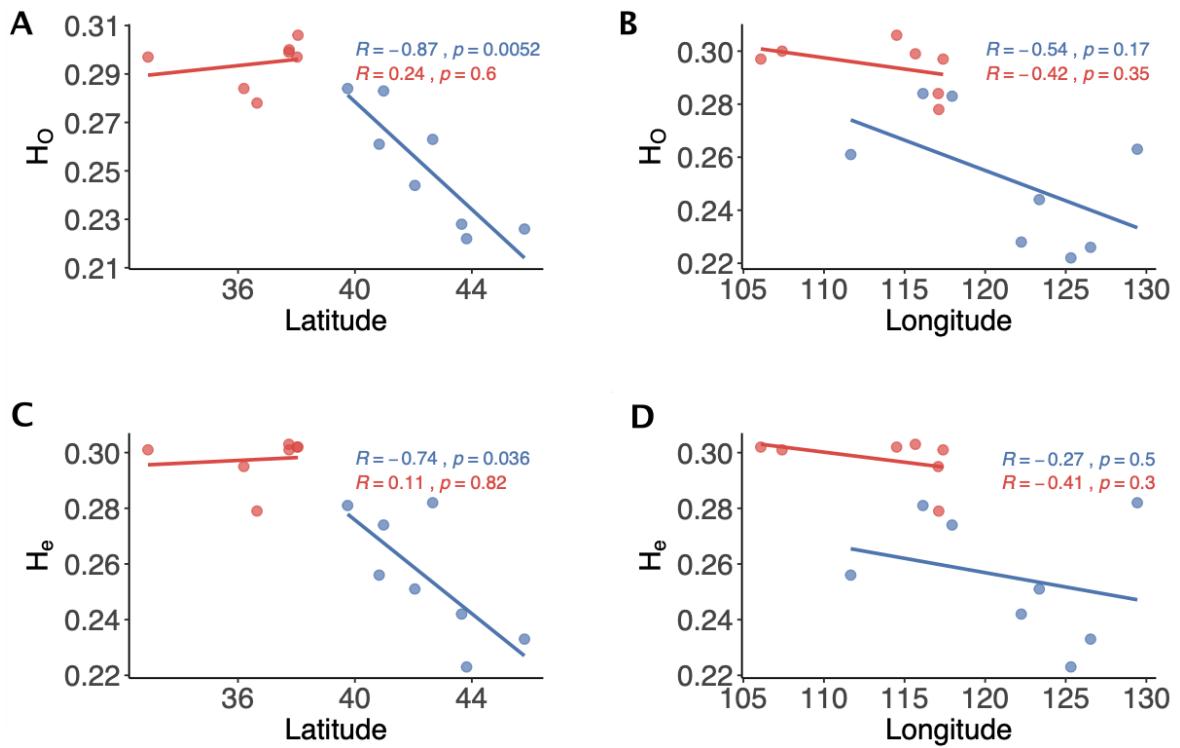


Figure S1.5: Relationships between heterozygosity and latitude/longitude without CIX. **A, B** Relationships between observed heterozygosity and latitude/longitude excluding population CIX. **C, D** Relationships between expected heterozygosity and latitude/longitude excluding CIX. The eight populations in the north identified by sPCA are colour-coded in blue and the other eight populations in the south in red. Colour-coded lines are the trend lines. Pearson correlation coefficient and P-value are presented on the top right of each graph.

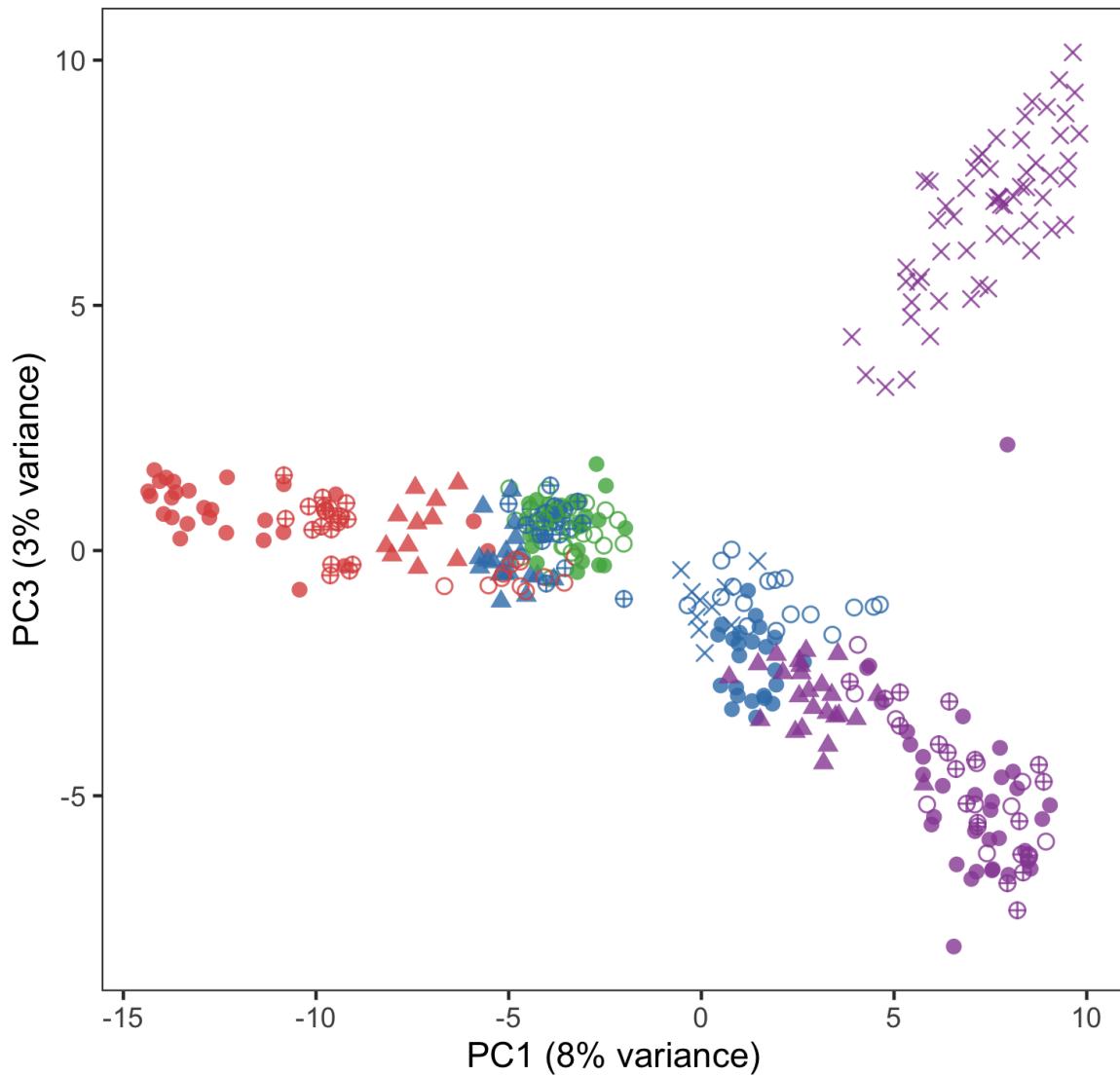


Figure S1.6: PCA analysis of the ALB populations in China. PC1 and PC3 explained 8% and 3% of the genetic variance, respectively. A total of 6,102 SNPs were involved in the analysis. For population colours and shapes, refer to Figure 1.2A.

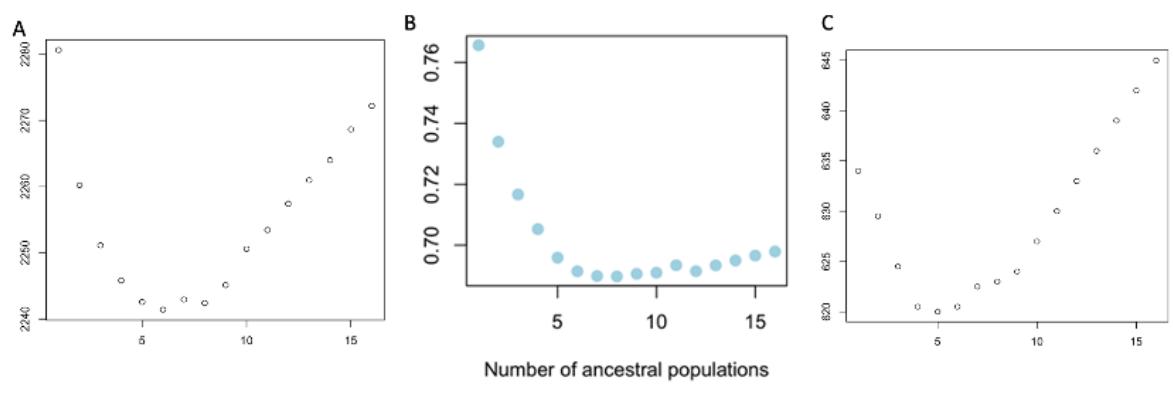


Figure S1.7: **A** Bayesian Information Criterion (BIC) for the SNP dataset used in the DAPC analysis. The x-axis indicates the number of clusters and the y-axis shows the BIC values. **B** The cross-entropy criterion (y-axis) for the respective number of ancestral populations (x-axis) as identified in the LEA analysis (SNP marker analysis only). **C** BIC for the microsatellite dataset used in the DAPC.

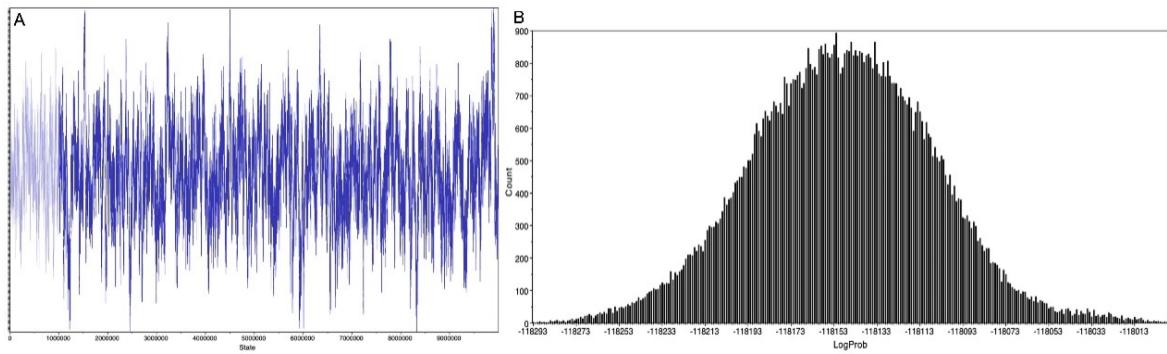


Figure S1.8: Trace file visualization of BayesAss results. **A** Trace plot of BayesAss analysis. **B** Posterior probability distribution.

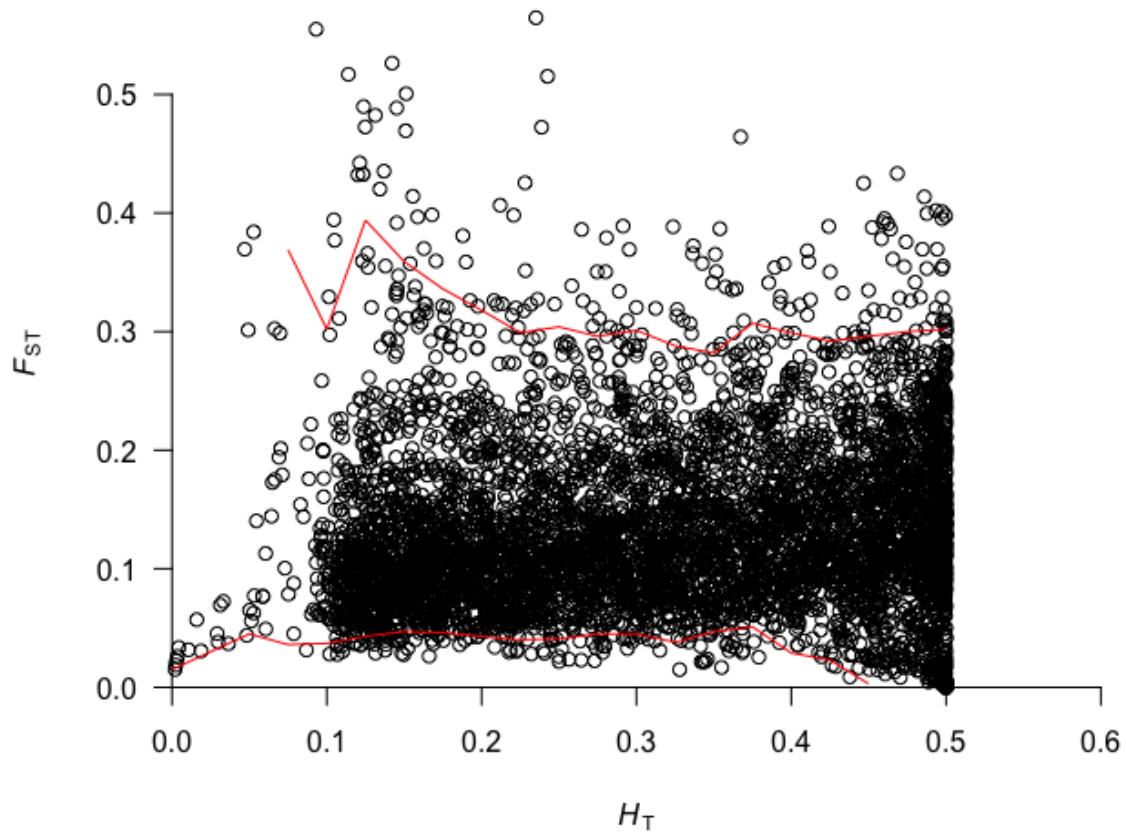


Figure S1.9: Heterozygosity (H_T) - F_{ST} relationship with smoothed quantiles in *fsthet*. Loci lying outside the quantiles were identified as outliers putatively under (positive or balancing) selection as determined under confidence intervals of 95% (red lines). The mean F_{ST} across all 6,102 loci is 0.134 ± 0.002 .

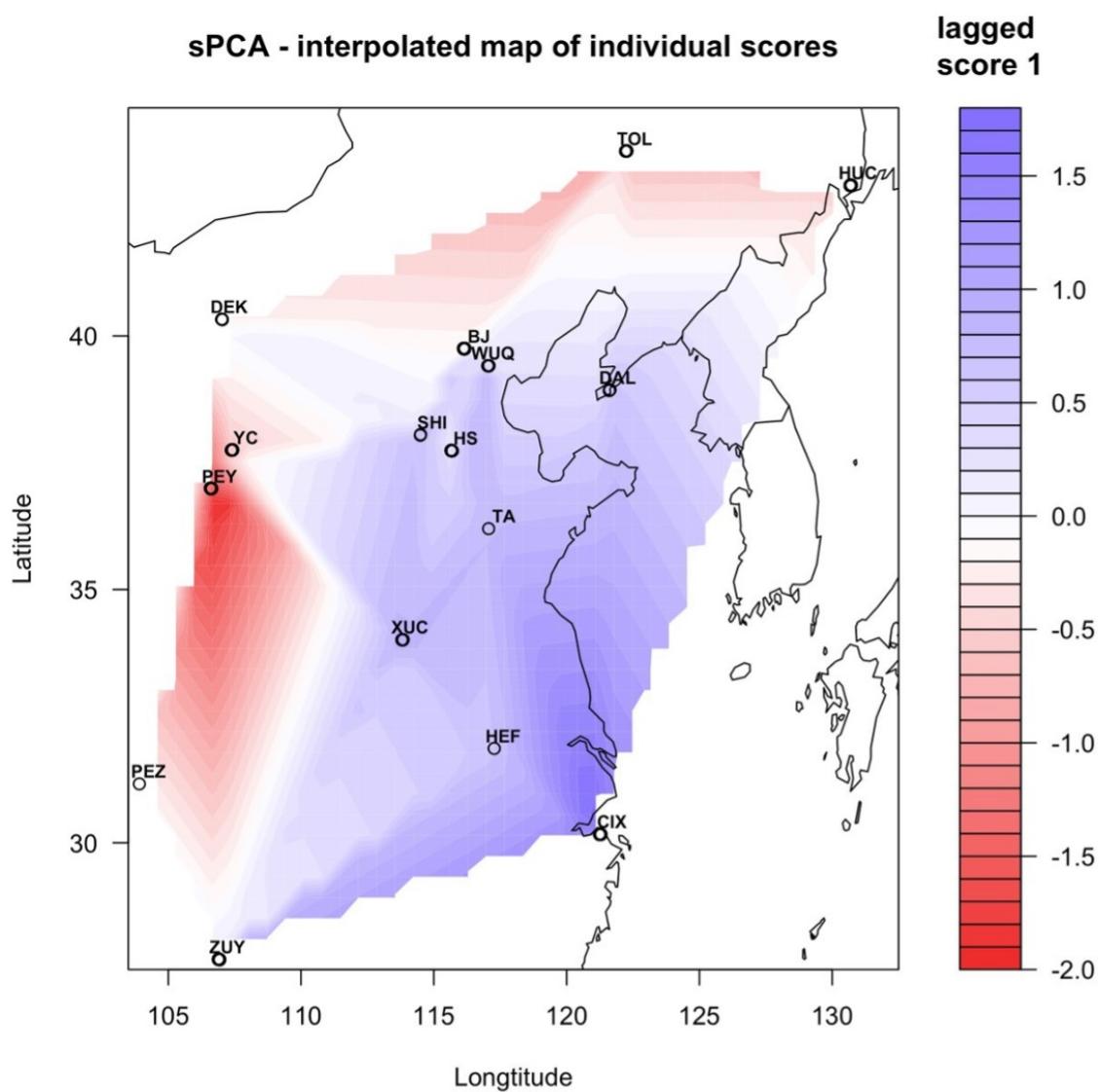


Figure S1.10: The sPCA analysis for the microsatellite dataset. The name codes for the populations are given in Table S1.3.

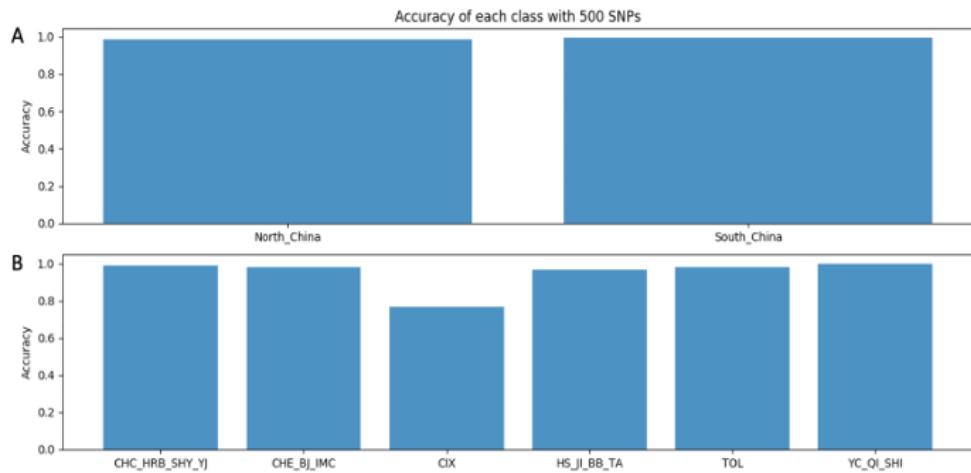


Figure S1.11: Assignment accuracy for each population assignment class in Mycorrhiza based on the most discriminant 500 SNPs (after testing the whole dataset of 6,102 SNPs). **A** ALB population divided into two groups (sPCA informed). **B** ALB populations divided into six groups (DAPC informed). Grouping information as in Figure S1.2.

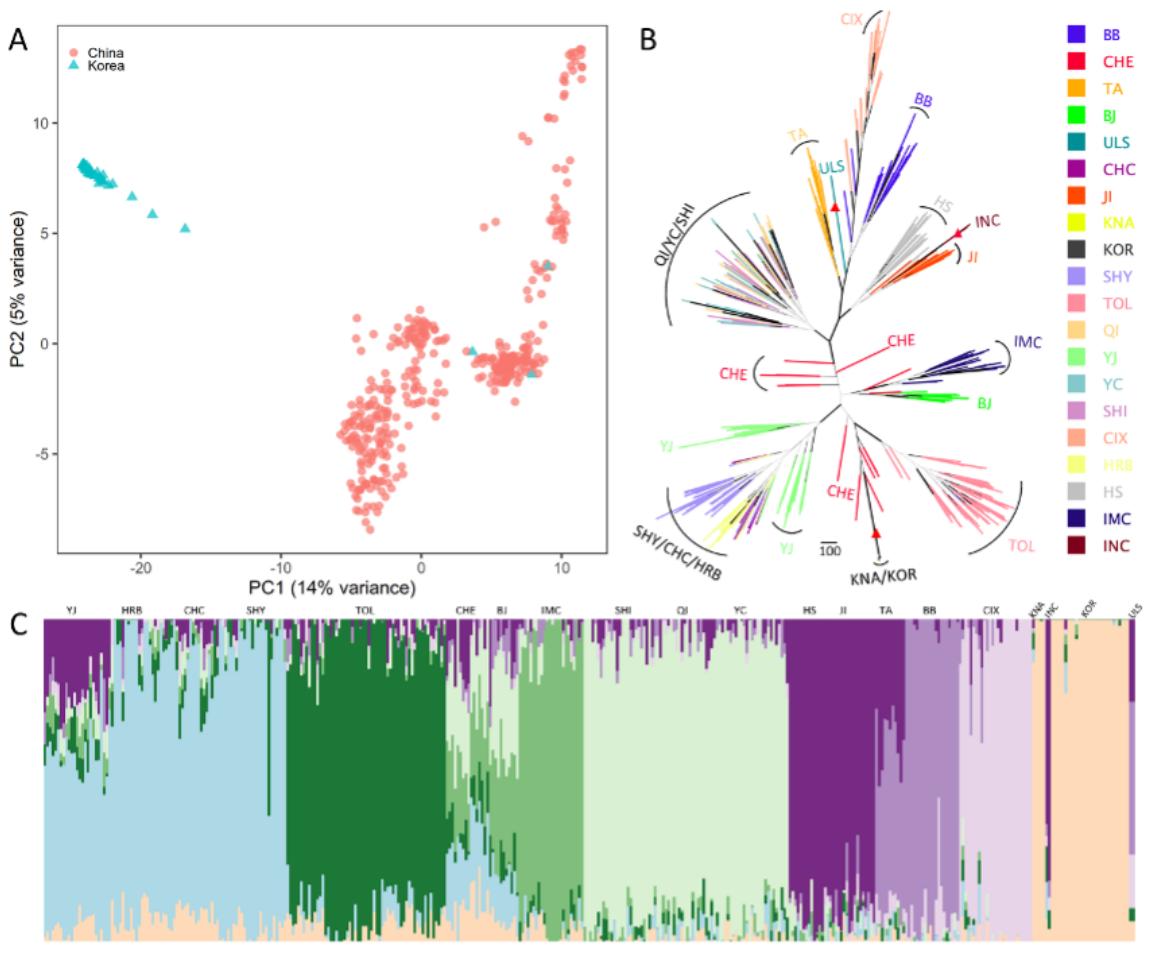


Figure S1.12: Population genetic analyses of native ALBs in China and Korea (including related Korean samples). **A** PCA analysis. **B** Maximum likelihood phylogenetic tree in RAxML. Each branch represents a sample. The width of each branch corresponds with the bootstrap value. The scale bar shows a bootstrap value of 100%. Branches with a red triangle point to Korean samples. **C** Bar plot of ancestry coefficients identified in LEA (optimal $K=8$). Rotated labels indicate the Korean samples. A total of 6,102 SNPs were used in the analyses. For population codes, see Table 1.1. For relatedness estimates of KNA/KOR, see Table S1.12.

1.10 Acknowledgements

We acknowledge funding of the study through Genome Canada, Genome British Columbia, Genome Québec, the Canadian Forest Service, the Canadian Food Inspection Agency, and FPInnovations through the Large Scale Applied Research Project LSARP 10106. This work was also supported by a four-year grant from the China Scholarship Council to MC. We are thankful to Gwylim Blackburn for feedback and discussions.

1.11 References

- Abed, A., Légaré, G., Pomerleau, S., St-Cyr, J., Boyle, B., & Belzile, F. J. (2019). Genotyping-by-sequencing on the ion torrent platform in barley. *Methods in Molecular Biology*, 1900, 233–252.
- Adamack, A. T., & Gruber, B. (2014). PopGenReport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, 5, 384–387.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215, 403-410.
- Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander, E. S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 407, 513–516.
- An, Y., Wang, B., Yang, X., Lin, X., Chen, J., Huang, X., & Mastro, V. (2004). Characterizing populations of *Anoplophora glabripennis* and related taxa with RAPD. *Acta Entomologica Sinica*, 47, 229–235.
- Angilletta, M. J. (2009). Thermal Adaptation: A Theoretical and Empirical Synthesis. In *Oxford University Press*.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25, 25-29.
- Bai, W., Liao, W., & Zhang, D. (2010). Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytologist*, 188, 892–901.
- Bai, W., Wang, W., & Zhang, D. (2016). Phylogeographic breaks within Asian butternuts indicate the existence of a phytogeographic divide in East Asia. *New Phytologist*, 209, 1757–1772.
- Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, 2019, e7724.
- Beerli, P., Mashayekhi, S., Sadeghi, M., Khodaei, M., & Shaw, K. (2019). Population genetic inference with MIGRATE. *Current protocols in bioinformatics*, 68.
- Bester-van der Merwe, A. E., Roodt-Wilding, R., Volckaert, F. A., & D'Amato, M. E. (2011). Historical isolation and hydrodynamically constrained gene flow in declining populations of the South-African abalone, *Haliotis midae*. *Conservation Genetics*, 12, 543–555.
- Bilodeau, P., Roe, A. D., Bilodeau, G., Blackburn, G. S., Cui, M., Cusson, M., ... Hamelin,

R. C. (2019). Biosurveillance of forest insects: part II—adoption of genomic tools by end user communities and barriers to integration. *Journal of Pest Science*, 92, 71–82.

Blackburn, G. S., Bilodeau, P., Cooke, T., Cui, M., Cusson, M., Hamelin, R. C., ... & Porth, I. (2020). An applied empirical framework for invasion science: confronting biological invasion through collaborative research aimed at tool production. *Annals of the Entomological Society of America*, 113, 230-245.

Brightwell, R. J., Labadie, P. E., & Silverman, J. (2010). Northward expansion of the invasive *Linepithema humile* (Hymenoptera: Formicidae) in the Eastern United States is constrained by winter soil temperatures. *Environmental Entomology*, 39, 1659–1665.

CABI, 2021. *Anoplophora glabripennis*. In: Invasive Species Compendium. Wallingford, UK: CAB International. <https://www.cabi.org/isc/datasheet/5557>.

Cao, L., Wei, S., Hoffmann, A., Wen, J., & Chen, M. (2016). Rapid genetic structuring of populations of the invasive fall webworm in relation to spatial expansion and control campaigns. *Diversity and Distributions*, 22, 1276–1287.

Carter, M., Smith, M., & Harrison, R. (2009). Patterns of genetic variation among populations of the Asian longhorned beetle (Coleoptera: Cerambycidae) in China and Korea. *Annals of the Entomological Society of America*, 102, 895–905.

Carter, M., Smith, M., & Harrison, R. (2010). Genetic analyses of the Asian longhorned beetle (Coleoptera, Cerambycidae, *Anoplophora glabripennis*), in North America, Europe and Asia. *Biological Invasions*, 12, 1165–1182.

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

Chen, T., & Lou, A. (2019). Phylogeography and paleodistribution models of a widespread birch (*Betula platyphylla* Suk.) across East Asia: Multiple refugia, multidirectional expansion, and heterogeneous genetic pattern. *Ecology and Evolution*, 9, 7792–7807.

Cheng, C., Yao, F., Chu, B., Li, X., Liu, Y., Wu, Y., Mei, Y., Wang, P., Hou, L. & Zou., X. (2014). Identification of the glycerol kinase gene and its role in diapause embryo restart and early embryo development of *Artemia sinica*. *Gene*, 537, 51-62

Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 Package. *R News*, 4, 5–10.

Chung, M., Son, S., Suh, G., Herrando-Moraira, S., Lee, C., López-Pujol, J., & Chung, M. (2018). The Korean Baekdudaegan Mountains: A glacial refugium and a biodiversity hotspot that needs to be conserved. *Frontiers in Genetics*, 9, 489.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M.

(2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6, 80–92.

Colunga-Garcia, M., Haack, R. A., Magarey, R. A., & Margosian, M. L. (2010). Modeling spatial establishment patterns of exotic forest insects in Urban areas in relation to tree cover and propagule pressure. *Journal of Economic Entomology*, 103, 108–118.

Cristescu, M. E. (2015). Genetic reconstructions of invasion history. *Molecular Ecology*, 24, 2212–2225.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

De Frenne, P., Graae, B. J., Rodríguez-Sánchez, F., Kolb, A., Chabrerie, O., Decocq, G., ... Verheyen, K. (2013). Latitudinal gradients as natural laboratories to infer species' responses to temperature. *Journal of Ecology*, 101, 784–795.

DeWoody, J., Nason, J. D., & Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, 6, 951–957.

Diepenbroek, M., Bayer, B., Schwender, K., Schiller, R., Lim, J., Lagacé, R., & Anslinger, K. (2020). Evaluation of the ion Ampliseq™ phenotrivium panel: MPS-based assay for ancestry and phenotype predictions challenged by casework samples. *Genes*, 11, 1–24.

Du, Z., Ishikawa, T., Liu, H., Kamitani, S., Tadauchi, O., Cai, W., & Li, H. (2019). Phylogeography of the assassin bug *Sphedanoolestes impressicollis* in east asia inferred from mitochondrial and nuclear gene sequences. *International Journal of Molecular Sciences*, 20, 1234.

Du, Z., Wu, Y., Chen, Z., Cao, L., Ishikawa, T., Kamitani, S., ... Li, H. (2021). Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evolutionary Applications*, 14, 915–930.

Eloff, J., Veldtman, R., Bulgarella, M., & Lester, P. J. (2020). Population genetics of the invasive wasp *Vespula germanica* in South Africa. *Insectes Sociaux*, 67, 229–238.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.

Excoffier, L., & Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of ge-

nomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27, 1332–1334.

FAO. (2002). International Standards for Phytosanitary Measures Integrated Measures.

Feng, Y., Xu, L., Tian, B., Tao, J., Wang, J., & Zong, S. (2014). Cold hardiness of Asian longhorned beetle (Coleoptera: Cerambycidae) larvae in different populations. *Environmental Entomology*, 43, 1419–1426.

Feng, Y., Tursun, R., Xu, Z., Ouyang, F., & Zong, S. (2016a). Effect of three species of host tree on the cold hardiness of overwintering larvae of *Anoplophora glabripennis* (Coleoptera: Cerambycidae). *European Journal of Entomology*, 113, 212–216.

Feng, Y., Xu, L., Li, W., Xu, Z., Cao, M., Wang, J., Tao, J., & Zong, S. (2016b). Seasonal changes in supercooling capacity and major cryoprotectants of overwintering Asian longhorned beetle (*Anoplophora glabripennis*) larvae. *Agricultural and Forest Entomology*, 18, 302–312.

Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology and Evolution*, 1, 1–7.

Flanagan, S. P., & Jones, A. G. (2017). Constraints on the FST–heterozygosity outlier approach. *Journal of Heredity*.

Fragnière, Y., Bétrisey, S., Cardinaux, L., Stoffel, M., & Kozlowski, G. (2015). Fighting their last stand? A global analysis of the distribution and conservation status of gymnosperms. *Journal of Biogeography*, 42, 809–820.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925–929.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196, 973–983.

Georges-Filteau, J., Hamelin, R. C., & Blanchette, M. (2020). Mycorrhiza: Genotype assignment using phylogenetic networks. *Bioinformatics*, 36, 212–220.

Guo, X., Wang, H., Bao, L., Wang, T., Bai, W., Ye, J., & Ge, J. (2014). Evolutionary history of a widespread tree species Acer mono in East Asia. *Ecology and Evolution*, 4, 4332–4345.

Haack, R. A. (2020). Asian longhorned beetle: A 2020 update.

Hamelin, R. C., & Roe, A. D. (2020). Genomic biosurveillance of forest invasive alien enemies: A story written in code. *Evolutionary Applications*, 13, 95–115.

He, J., Kreft, H., Gao, E., Wang, Z., & Jiang, H. (2017). Patterns and drivers of zoogeographical regions of terrestrial vertebrates in China. *Journal of Biogeography*, 44, 1172–1184.

- Hou, Z., Wang, Z., Ye, Z., Du, S., Liu, S., & Zhang, J. (2018). Phylogeographic analyses of a widely distributed *Populus davidiana*: Further evidence for the existence of glacial refugia of cool-temperate deciduous trees in northern East Asia. *Ecology and Evolution*, 8, 13014–13026.
- Hu, J., Angeli, S., Schuetz, S., Luo, Y., & Hajek, A. E. (2009). Ecology and management of exotic and endemic Asian longhorned beetle *Anoplophora glabripennis*. *Agricultural and Forest Entomology*, 11, 359–375.
- Huang, Jihong, Chen, B., Liu, C., Lai, J., Zhang, J., & Ma, K. (2012). Identifying hotspots of endemic woody seed plant diversity in China. *Diversity and Distributions*, 18, 673–688.
- Huang, Jixia, Qu, B., Fang, G., Li, X., & Zong, S. (2020). The drivers of the Asian longhorned beetle disaster show significant spatial heterogeneity. *Ecological Indicators*, 117, 106680.
- Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*, 14, 68–73.
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267.
- Jarne, P., & Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11, 424–429.
- Javal, M., Lombaert, E., Tsykun, T., Courtin, C., Kerdelhué, C., Prospero, S., ... & Roux, G. (2019). Deciphering the worldwide invasion of the Asian long-horned beetle: A recurrent invasion process from the native area together with a bridgehead effect. *Molecular Ecology*, 28, 951–967.
- Javal, M., Roques, A., Haran, J., Hérard, F., Keena, M., & Roux, G. (2019). Complex invasion history of the Asian long-horned beetle: fifteen years after first detection in Europe. *Journal of Pest Science*, 92, 173–187.
- Javal, M., Roques, A., Roux, G., & Laparie, M. (2018). Respiration-based monitoring of metabolic rate following cold-exposure in two invasive *Anoplophora* species depending on acclimation regime. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*, 216, 20–27.
- Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.
- Jombart, T., Devillard, S., Dufour, A. B., & Pontier, D. (2008). Revealing cryptic spatial

- patterns in genetic variability by a new multivariate method. *Heredity*, *101*, 92–103.
- Joshi, N. (2011). Sabre: a barcode demultiplexing and trimming tool for FastQ files. *GitHub: San Francisco, CA, USA*.
- Keena, M. A., & Moore, P. M. (2010). Effects of temperature on *Anoplophora glabripennis* (Coleoptera: Cerambycidae) larvae and pupae. *Environmental Entomology*, *39*, 1323–1335.
- Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., ... & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*, 239–243.
- Kihara, F., Itoh, K., Iwasaka, M., Niimi, T., Yamashita, O., & Yaginuma, T. (2009) Glycerol kinase activity and glycerol kinase-3 gene are up-regulated by acclimation to 5 degrees C in diapause eggs of the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, *39*, 763–9.
- Kim, S. I., Park, S. K., Lee, H., Oshida, T., Kimura, J., Kim, Y. J., ... Min, M. S. (2013). Phylogeography of Korean raccoon dogs: Implications of peripheral isolation of a forest mammal in East Asia. *Journal of Zoology*, *290*, 225–235.
- Kumar S, Stecher G, Li M, Knyaz C, & Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549.
- Laiolo, P., & Obeso, J. R. (2017). Life-history responses to the altitudinal gradient. In *High mountain conservation in a changing world* (pp. 253–283). Springer, Cham.
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, *24*, 3688–3705.
- Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., Dos Santos, G., Garapati, P. V., ... & Thurmond, J. (2021). FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic acids research*, *49*, D899–D907.
- Lee, S., Lee, Y., & Lee, S. (2020). Population genetic structure of *Anoplophora glabripennis* in South Korea: Invasive populations in the native range? *Journal of Pest Science*, *93*, 1181–1196.
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*, 589–595.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

- Li, X. (2012). Study on Logistics Industry Cooperation between Shandong and South Korea. *The Journal of Industrial Distribution & Business*, 3, 23–27.
- Lingafelter, S. W., & Hoebeke, R. E. (2002). Revision of the Genus *Anoplophora* (Coleoptera: Cerambycidae). In *The Entomological Society of Washington, Washington D. C. Entomological Society of Washington Washington, DC*.
- Lisboa, F. J., Peres-Neto, P. R., Chaer, G. M., Da Conceicao Jesus, E., Mitchell, R. J., Chapman, S. J., & Berbara, R. L. (2014). Much beyond Mantel: Bringing procrustes association metric to the plant and soil ecologist's toolbox. *PLoS ONE*, 9, e101238.
- Lischer, H., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299.
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, 4, 981–994.
- Luo, Y., Huang, J., & Li, J. (2000). The main achievements, problems and prospects in research of long-horned beetles in China. *Entomological Knowledge*, 37, 116–121.
- Lyu, Y., Wang, X., & Luo, J. (2020). Geographic patterns of insect diversity across China's nature reserves: The roles of niche conservatism and range overlapping. *Ecology and evolution*, 10, 3305–3317.
- Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, Vol. 20, pp. 136–142.
- Martin, A., Quinn, K., & Park, J. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, 42, 1–21.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17, 10.
- McKenna, D. D., Scully, E. D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S. M., ... Richards, S. (2016). Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology*, 17, 1–18.
- Mckown, A. D., Guy, R. D., Klápník, J., Gerald, A., Friedmann, M., Cronk, Q. C., ... Douglas, C. J. (2014). Geographical and environmental gradients shape phenotypic trait variation and genetic structure in *Populus trichocarpa*. *New Phytologist*, 201, 1263–1276.
- Meirmans, P. G. (2020). genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, 20, 1126–1131.

- Meng, P. S., Hoover, K., & Keena, M. A. (2015). Asian longhorned beetle (Coleoptera: Cerambycidae), an introduced pest of maple and other hardwood trees in North America and Europe. *Journal of Integrated Pest Management*, 6, 1–13.
- Millar, C. I., & Stephenson, N. L. (2015). Temperate forest health in an era of emerging megadisturbance. *Science*, 349, 823–826.
- Milne, R. I., & Abbott, R. J. (2002). The origin and evolution of tertiary relict floras. *Advances in Botanical Research*, 38, 281–314.
- Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods in Ecology and Evolution*, 10, 1808–1813.
- Park, Y., & Kim, Y. (2013). RNA interference of glycerol biosynthesis suppresses rapid cold hardening of the beet armyworm, *Spodoptera exigua*. *Journal of Experimental Biology*, 216, 4196–4203.
- Pedlar, J. H., McKenney, D. W., Yemshanov, D., Hope, E. S., & Wallin, K. (2020). Potential Economic Impacts of the Asian Longhorned Beetle (Coleoptera: Cerambycidae) in Eastern Canada. *Journal of Economic Entomology*, 113, 839–850.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135.
- Picq, S., Keena, M., Havill, N., Stewart, D., Pouliot, E., Boyle, B., ... Cusson, M. (2018). Assessing the potential of genotyping-by-sequencing-derived single nucleotide polymorphisms to identify the geographic origins of intercepted gypsy moth (*Lymantria dispar*) specimens: A proof-of-concept study. *Evolutionary Applications*, 11, 325–339.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38, 904–909.
- Qiu, Y., Fu, C., & Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution*, 59, 225–244.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*, 67(5), 901–904.
- Reaser, J. K., Burgiel, S. W., Kirkey, J., Brantley, K. A., Veatch, S. D., & Burgos-Rodríguez, J. (2020). The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. *Biological Invasions*, 22, 1–19.

- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., Wilkie, A. O., ... Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, *46*, 912–918.
- Roe, A. D., Torson, A. S., Bilodeau, G., Bilodeau, P., Blackburn, G. S., Cui, M., ... Hamelin, R. C. (2019). Biosurveillance of forest insects: part I—integration and application of genomic tools to the surveillance of non-native forest insects. *Journal of Pest Science*, *92*, 51–70.
- Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106.
- Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., ... Essl, F. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, *8*, 1–9.
- Shatz, A. J., Rogan, J., Sangermano, F., Ogneva-Himmelberger, Y., & Chen, H. (2013). Characterizing the potential distribution of the invasive Asian longhorned beetle (*Anoplophora glabripennis*) in Worcester County, Massachusetts. *Applied Geography*, *45*, 259–268.
- Shi, P., Qiao, X., Chen, X., Zhou, M., Qu, S., Ma, X., & Zhang, Z. (2014). Spatial distribution and temporal trends in daily and monthly precipitation concentration indices in the upper reaches of the Huai River, China. *Stochastic Environmental Research and Risk Assessment*, *28*, 201–212.
- Sjöman, H., Östberg, J., & Nilsson, J. (2014). Review of host trees for the wood-boring pests *Anoplophora glabripennis* and *Anoplophora chinensis*: An urban forest perspective. *Arboriculture and Urban Forestry*, *40*, 143–164.
- Smith, M. T., Bancroft, J., Li, G., Gao, R., & Teale, S. (2001). Dispersal of *Anoplophora glabripennis* (Cerambycidae). *Environmental Entomology*, *30*, 1036–1040.
- Smith, M. T., Tobin, P. C., Bancroft, J., Li, G., & Gao, R. (2004). Dispersal and spatiotemporal dynamics of Asian longhorned beetle (Coleoptera: Cerambycidae) in China. *Environmental Entomology*, *33*, 435–442.
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.
- Storer, C., Payton, A., McDaniel, S., Jordal, B., & Hulcr, J. (2017). Cryptic genetic variation in an inbreeding and cosmopolitan pest, *Xylosandrus crassiusculus*, revealed using ddRADseq. *Ecology and Evolution*, *7*, 10974–10986.
- Storey, K. B., & Storey, J. M. (2012). Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology*, *90*, 456–475.

- Team, R. C. (2015). R: A language and environment for statistical computing.
- The Gene Ontology Consortium. 2021. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49: D325-D334.
- The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*. 49: D480-D489.
- Tian, H., Han, L., Zhang, J., Li, X., Kawahara, T., Yukawa, T., ... Chung, M. (2018). Genetic diversity in the endangered terrestrial orchid *Cypripedium japonicum* in East Asia: Insights into population history and implications for conservation. *Scientific Reports*, 8, 1–13.
- Torkamaneh, D., Laroche, J., Bastien, M., Abed, A., & Belzile, F. (2017). Fast-GBS: A new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinformatics*, 18, 1–7.
- Torson, A.S., Zhang, M.L., Ong, K., Mohammad, L., Smith, A.J., Doucet, D., Roe, A.D., & Sinclair, B.J. (2021). Cold tolerance of laboratory-reared Asian longhorned beetles. *Comparative Biochemistry and Physiology, Part A*, 257,110957.
- Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., ... Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5, 247–252.
- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics selection evolution*, 34, 275-305.
- Ward, J.H. Jr (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, 236–244.
- Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST} . *The American Naturalist*, 186, S24-S36.
- Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., & Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics*, 18, 1–12.
- Williams, D. W., Lee, H., & Kim, I. (2004). Distribution and abundance of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in natural Acer stands in South Korea. *Environmental Entomology*, 33, 540–545.
- Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177–1191.

- Wu, Y., Bogdanowicz, S. M., Andres, J. A., Vieira, K. A., Wang, B., Cossé, A., & Pfister, S. E. (2020). Tracking invasions of a destructive defoliator, the gypsy moth (Erebidae: *Lymantria dispar*): Population structure, origin of intercepted specimens, and Asian introgression into North America. *Evolutionary Applications*, 13, 2056–2070.
- Wu, Y., Trepanowski, N. F., Molongoski, J. J., Reagel, P. F., Lingafelter, S. W., Nadel, H., ... Ray, A. M. (2017). Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in tradeassociated solid wood packaging material using DNA barcoding and morphology. *Scientific Reports*, 7, 1–12.
- Xu, Y., Li, Y., Wang, Q., Zheng, C., Zhao, D., Shi, F., ... & Zong, S. (2021). Identification of key genes associated with overwintering in *Anoplophora glabripennis* larva using gene co-expression network analysis. *Pest Management Science*, 77, 805-816.
- Yan, J. (1985). Study on the distribution area of *Anoplophora glabripennis* in East China. *Journal of North-Eastern Forestry College*, 13.
- You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., ... & Anderson, O. D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC bioinformatics*, 9, 1-13.
- Yu, G., Chen, X., Ni, J., Cheddadi, R., Guiot, J., Han, H., Harrison, S.P., ... & Zheng, Z. (2000). Palaeovegetation of China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. *Journal of Biogeography*, 27, 635-664.Zeng, Y., Wang, W., Liao, W., Wang, H., & Zhang, D. (2015). Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: The Mongolian oak as a case study. *Molecular Ecology*, 24, 5676–5691.
- Zhang, L., Zhao, W., Wang, F., & Qin, D. (2019). Genetic diversity and population structure of natural *Lycorma delicatula* (white) (hemiptera: Fulgoridea) populations in China as revealed by microsatellite and mitochondrial markers. *Insects*, 10, 312.
- Zhao, H., Fuller, A., Thongda, W., Mohammed, H., Abernathy, J., Beck, B., & Peatman, E. (2019). SNP panel development for genetic management of wild and domesticated white bass (*Morone chrysops*). *Animal Genetics*, 50, 92–96.
- Zhao, Y. P., Fan, G., Yin, P. P., Sun, S., Li, N., Hong, X., ... & Ge, S. (2019). Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. *Nature Communications*, 10, 1-10.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326–3328.

Chapter 2

Pool-Seq reveals genes selective for cold tolerance across Asian longhorned beetle's native range

2.1 Résumé

La sensibilité des insectes à la température et leur capacité à s'adapter à des conditions de froid extrême ont suscité un intérêt considérable pour la recherche sur la tolérance au froid chez ces organismes. Le nuisible invasif qu'est le longicorne asiatique (ALB) a une large distribution à travers les régions tempérées et froides dans ses régions natales en Chine. Dans cette étude, nous avons réalisé un séquençage complet du génome en « pool » (Pool-Seq) des populations d'ALB provenant de son aire de répartition native, en utilisant dix pools de 20 individus chacun, pour obtenir une collection extensive de marqueurs de polymorphisme nucléotidiques (SNP) à l'échelle du génome (4 millions de SNPs filtrés). Nous avons utilisé des analyses génomiques de sélection et d'association gène-environnement pour identifier les signatures génomiques de sélection liées à l'adaptation au froid. Nous avons identifié 5,359 gènes potentiellement sous sélection, ainsi que 87 gènes trouvés grâce à l'analyse d'association. Des gènes communs de tolérance au froid, tels que la glycérol kinase, les cytochromes P450, la protéine antigel et les protéines de choc thermique, ont été identifiés à travers les scans de sélection, alors qu'ils n'étaient pas significatifs dans l'analyse d'association gène-environnement. Cela suggère que les facteurs environnementaux influencent un large éventail de gènes impliqués dans l'évolution de la tolérance au froid. Étant donné la nature complexe de l'adaptation, il est difficile de cibler un petit nombre de gènes candidats responsables de ce trait grâce aux analyses d'association. Cette étude réalise le premier profilage à l'échelle du génome complet de la variation génétique dans les populations de cette espèce invasive d'insecte. Nos résultats fournissent des informations précieuses sur les mécanismes génétiques qui dirigent l'adaptation locale chez ce nuisible invasif.

2.2 Abstract

Insects' sensitivity to temperature and their ability to adapt to extreme cold conditions has generated considerable interest in cold tolerance research in these organisms. The invasive insect pest Asian longhorned beetle (ALB) has a broad distribution across temperate and cold temperate regions in its native China. In this study, we conducted pooled whole-genome sequencing (Pool-Seq) of ALB populations from its native range, using ten pools of 20 individuals each, to obtain an extensive collection of genome-wide single nucleotide polymorphism markers (4 million filtered SNPs). We employed genome scans for selection and gene-environment analysis to identify genomic signatures of selection related to cold adaptation. We identified 5,359 genes putatively under selection, along with 87 genes found through the association analysis. Common cold tolerance genes, including glycerol kinase, cytochrome P450s, antifreeze protein, and heat shock proteins, were identified through selection scans, while they were not significant in the gene-environment association analysis. This suggests that environmental factors influence a broad array of genes involved in the evolution of cold tolerance. Given the complex nature of adaptation, it is challenging to pinpoint a handful of candidate genes responsible for this trait through association analysis. This study conducts the first whole genome-wide profiling of genetic variation in populations of this invasive insect species. Our findings provide valuable insights into the genetic mechanisms driving local adaptation in this invasive pest.

2.3 Introduction

Genetic markers can be used to study genetic variation between and within populations. Common usages include estimating genetic diversity of populations, examining population structure, and making inference for demographic history etc., which pertain to the field of population genetics and are referred to as *broad-sense* genomics by Garner et al. (2016).

The emergence of next-generation sequencing (NGS) technologies has significantly advanced genotyping technologies, enabling the discovery of massive genome-wide markers. It allows for detecting any loci or genomic regions associated with phenotypic traits along the genome more efficiently and on a finer scale than traditional methods. This type of research can be aptly termed as *narrow-sense* genomics, distinguishing it from *broad-sense* genomics in the level of complexity and the type of questions being addressed (Garner et al., 2016; Hohenlohe et al., 2018).

Reduced representation genomic libraries (RRLs) (Altshuler et al., 2000; van Tassell et al., 2008) in combination with common NGS technologies have become popular in generating large numbers of genome-wide SNP markers at a moderate cost, such as genotyping-by-sequencing (GBS) or restriction-site associated DNA sequencing (RADseq) (Peterson et al., 2012; Elshire et al., 2011). These technologies have proven to be sufficient for most *broad-sense* population

genomics, addressing genetics questions that typically rely on thousands of markers while also costing less than traditional genetic methods that use markers such as microsatellites (Tokarska et al., 2009). They have been widely applied to many non-model insect species. For instance, Chen *et al.* used RADseq to analyse the population structure and invasion pathway of the invasive coconut leaf beetle *Brontispa longissimi* (Chen et al., 2020). SNPs generated through GBS have also been applied to investigate genetic structure and gene flow of the oriental fruit moth *Grapholita molesta* (Silva-Brandão et al., 2015). The population structure for native Asian longhorned beetle (ALB) populations was also revealed using the RRLs approach (Cui et al., 2022).

Studying the genetic basis underlying selection is a fundamental aspect of understanding evolutionary processes and has been a central focus of molecular ecology and evolutionary studies (Luikart et al., 2003). In the genomics era, genomics in the *narrow-sense* involves genome scans for local adaption to identify signals of selection in a so-called “bottom-up” approach without prior knowledge of specific loci. Furthermore, genetic variation can be assessed against environmental or phenotypic variation to identify which loci are affected by selection and correlated with traits or environmental factors of interest (Sork et al., 2013). This can be used to identify associated loci among the genome-wide markers.

Generating high-density markers through whole-genome sequencing (WGS) remains costly for extensive population-level studies. An alternative approach has been proposed that sequences pools of individual DNA samples (Pool-seq); it thus consists of one or multiple sequencing libraries of pooled DNA samples and thereby eliminates the need for preparing and sequencing individual DNA libraries (Schlötterer et al., 2014). This pooled sample whole-genome sequencing (WGS) enables the discovery of high-density SNP markers in a cost-efficient way. It has previously been demonstrated to be as accurate as individual-based genotyping in estimating allele frequencies (Gautier et al., 2013). Furthermore, Schlötterer et al. (2014) estimated that, at the same cost, Pool-seq can produce a higher accuracy-to-cost ratio compared to individual-based genotyping.

However, Pool-seq comes with its own limitations, such as differential representation of individuals in pools when sample sizes are small, challenges in distinguishing sequence errors from low frequency alleles, and the inability to accurately calculate linkage disequilibrium (LD) due to the lack of individual genotype information. These factors can bias allele frequency estimations and thus limit certain downstream population genetic analyses (Fuentes-Pardo & Ruzzante, 2017; Schlötterer et al., 2014). To reduce such biases, it is recommended to use pool sizes ideally with more than 50 samples (Gautier et al., 2013; Schlötterer et al., 2014). In practice, Pool-seq has been applied to various non-model species to investigate their genetic adaptation processes to environmental changes. For example, Fisher et al. sequenced pooled samples of *Arabidopsis halleri* from five natural populations and identified highly differentiated genomic regions and SNPs. They tested the most strongly differentiated SNPs associated

with environmental factors using partial Mantel tests and found 175 genes highly correlated with climatic factors (Fischer et al., 2013). In another study, Pool-seq was used to identify genomic regions and candidate genes involved in ecotype formation and adaptation in the yellow monkeyflower for two different ecotypes, providing insights into the genetic mechanisms underlying plant adaptations to different habitats (Gould, Chen & Lowry, 2017). Ren et al. (2021) used Pool-seq to identify 26 genes potentially regulating feather color in domestic and feral geese. Koot et al. (2022) uncovered five geographically distinct gene pools of *Leptospermum scoparium* across 76 locations in New Zealand using pooled DNA. They also explored the evolutionary history of these populations.

Here, our study species is the Asian longhorned beetle (ALB), *Anoplophora glabripennis* (Motschulsky), an invasive forest insect pest that is native to China and the Korean peninsula, and which has been repeatedly introduced to North America and Europe since the 1990s. ALB attacks broadleaved trees, with larvae tunneling deep into the wood, causing the host trees to die and posing significant threats to forest ecosystems. ALB is a polyphagous pest known to feed primarily on poplars, willows, and elm trees in China, while favoring maple trees in North America. It has caused considerable damage to global forest ecosystems (Nowak et al., 2001; Pedlar et al., 2020; Straw et al., 2015; Turgeon et al., 2022). The ALB is generally native to temperate regions and exhibits excellent cold tolerance (Keena, 2006; Keena & Moore, 2010; Roden et al., 2009; Torson et al., 2021), which contributes to its invasive success. Although the structure of native ALB populations has been extensively studied using individual-based GBS data (Cui et al., 2022), understanding the cold adaptation of the pest is critical for pest risk assessment. Therefore, in this study, we used Pool-seq technology to obtain a large number of SNPs and identify genetic variants potentially relevant to cold adaptation and host use in native populations of *A. glabripennis*.

2.4 Materials and Methods

2.4.1 DNA pooling and sequencing

We used ALB samples from 16 sites in China (Figure 2.1) and designed 10 pools based on the population structure of native ALB populations from individual genotyping data published by Cui et al. (2022) and the similarity in regional temperature for the sampling locations. Six of these pools consisted of samples from the same site, while the remaining four pools (PNW, PNE, PN, and PE) contained samples from multiple locations. Pooled DNA individuals were selected with an Isolation-by-descent coefficient <0.13 (published data by Cui et al. (2022)) to avoid close relatedness among individuals. For each pool, we mixed equimolar amounts of DNA from 20 samples.

Pooled genomic DNA (500ng) was mechanically fragmented using a Covaris M220 Focused-ultrasonicator (Covaris, Woburn, MA, USA) and shotgun library preparation was performed

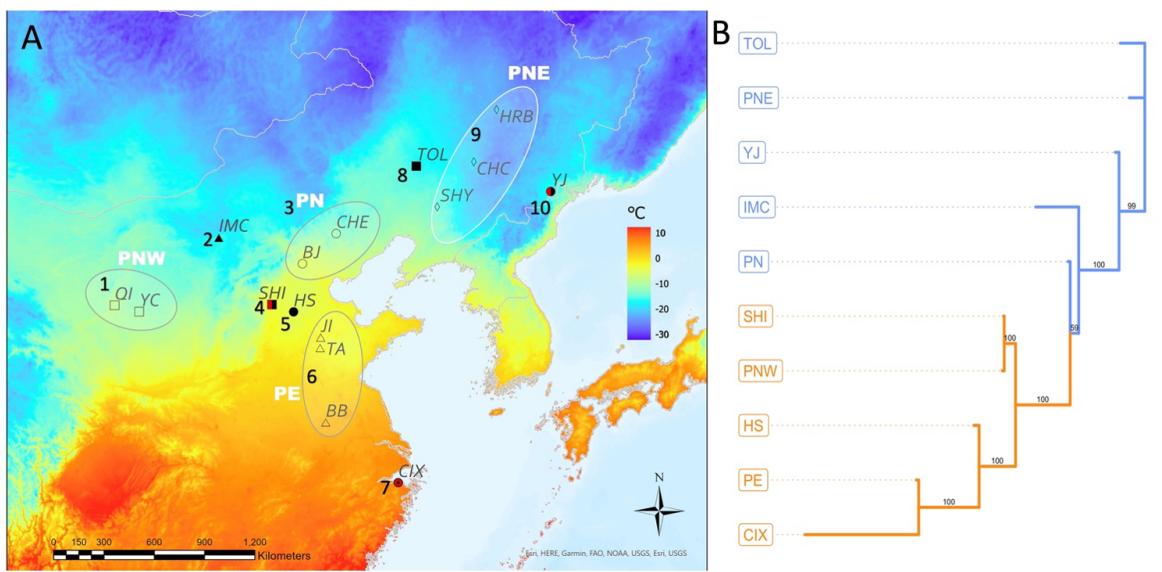


Figure 2.1: **A** 16 sampling sites and 10 DNA pools. Each pool consists of 20 individuals. The four pools PNW, PNE, PN, and PE containing samples from multiple locations are circled together. The map is produced using ArcGIS Pro v2.8 with bioclimatic variable (BIO 6, Min Temperature of Coldest Month) downloaded from WorldClim v2.1 at 2.5-min resolution. **B** Phylogenetic tree in RAXML using 109,062 SNPs. Bootstrap values (%) are indicated in each node. Pools with populations located in the north are shown in blue, southern population sample pools are in orange.

using the NEBNext UltraII kit (New England Biolabs, Ipswich, MA, USA) at the Plateforme d’Analyses Génomiques, Institut de Biologie Intégrative et des Systèmes (Université Laval, Quebec City, QC, Canada). Sequencing was performed on an Illumina NovaSeq 6000 S4 flowcell in PE150 mode at the Centre d’Expertises et de Services Génome Québec (Montreal, QC, Canada).

2.4.2 Variant calling and filtering

We checked the quality of raw sequence reads using FastQC v0.11.8 (Andrews, 2010) and trimmed adapters with TrimGalore v0.4.5 (Krueger, 2015). We mapped reads onto ALB’s reference genome GCA_000390285.1 (McKenna et al., 2016), which contains 10,474 scaffolds, using BWA-MEM (Li & Durbin, 2010), and we used Samtools v1.9 (Li et al., 2009) to sort and index aligned reads.

We performed variant calling using GATK v4.1.9.0 *HaplotypeCaller* (Auwera et al., 2013) with 40 as the pooled ploidy (20 individuals/pool, diploid organism). Before variant calling, we used picard v2.23.4 (Broad Institute, 2019) to mark and remove PCR duplicates. We filtered raw variants following GATK’s hard-filtering instructions (DePristo et al., 2011) based on the annotation values where the quality-by-depth value QualByDepth (QD) is suggested to better

filter data than either QUAL (variant quality) or DP (variant depth) individually, with the recommendation cut-off for any loci with a QD <2. However, for our data, the QD peak value was between 0 ~ 3, while this feature might be related to Pool-seq itself. Therefore, we applied QUAL and DP separately, using the *VariantFiltration* tool with the following cut-offs: QUAL <50.0, RMSMappingQuality (MQ) <40.0, FisherStrand (FS) <60.0, StrandOddsRatio (SOR) > 3.0, MappingQualityRankSumTest (MQRankSum) < -12.5, ReadPosRankSumTest (ReadPosRankSum) < -8.0 and missing-values-evaluate-as-failing as true. In addition, we measured the average read depth, based on which we filtered out variants with a read depth >850 across all pools to circumvent repetitive sequences as well as a read depth <40 per pool. We also removed low-frequency variants with an allele count (AC) <3 to minimize sequence errors. We kept only biallelic SNPs using the *SelectVariants* tool. For comparison, we employed a second method, bcftools v1.9 (Li, 2011) to call variants. Indels realignment was performed using the Samtools *calmd* tool. We then performed variant calling using bcftools *mpileup* and *call* tools with a -C50, minimum base quality of 4 (-q).

To evaluate the two variant data sets obtained from the two SNP callers on the Pool-seq data, we compared SNPs shared between Pool-seq and the individual genotyping SNP data published by Cui et al. (2022), by assessing the allele frequency correlations of the mutual SNPs for both caller SNP sets individually with the GBS obtained variants. Finally, we used SnpEff v4.2 (Cingolani et al., 2012) to annotate all SNPs against a database built *de novo* with ALB's reference genome.

2.4.3 Genetic diversity and phylogenetic analysis

We used NPStat v1 (Ferretti, Ramos-Onsins, & Pérez-Enciso, 2013) to estimate genetic diversity indices based on scaffold 1 which is the longest scaffold across ALB's genome. We calculated summary statistics of segregating sites S , Watterson estimator of θ , nucleotide diversity (P_i) and *Tajima's D* in non-overlapping 5 kb windows. We also generated an unrooted phylogenetic tree to assess the evolutionary relationships of the ten pools using RAxML v8.2.9 (Stamatakis, 2014) under the GTRGAMMA model, with a random selected SNP subset containing 109,062 SNPs. Data format conversion was completed using PGDSpider (Lischer & Excoffier, 2012). The tree was constructed with 1000 bootstrap replicates and visualized using the R package ggtree (Yu et al., 2017). Based on the findings from Cui et al. (2022), which indicated a genetic division between North and South in the spatial structure among populations, we accordingly categorized the ten pools in our tree to aid with interpreting the phylogenetic relationships.

2.4.4 Mantel tests for genetic with environmental and geographical distances

We obtained 19 bioclimatic variables from WorldClim v1.4 at a resolution of 10-min (Hijmans et al., 2017). We selected two variables that were not highly correlated with other variables ($|r| < 0.7$) that is mean temperature of the coldest month (MTCM) and isothermality (ISO). To obtain the values for the four pools that contain samples from different sites, we calculated the average across locations within the pool. We conducted Mantel tests between each of the matrix pairs using the *vegan* package in R (Oksanen et al., 2007) and the Pearson method with 9999 permutations. We examined the correlations between the environmental distance and the genetic distance which we calculated based on Wright's F_{ST} (Wright 1949):

$$F_{ST} = (H_T - H_S)/H_T$$

We performed the test using the genetic distance and the geographical distance calculated with the *geosphere* package in R (Hijmans et al., 2017). Additionally, we examined the correlation of the F_{ST} matrix and the one calculated from individual GBS data. In the Mantel tests we included only the pools containing single location.

2.4.5 Genome-wide scans for ALB loci under selection

We conducted the genome scan using BayPass v2.2 (Gautier, 2015). We ran the core model, which conducts genome scans for differentiation of allele frequencies using the XtX statistic (Günther & Coop, 2013). This statistic allows identifying SNPs that are over-differentiated across all populations while considering population structure resulting from the population demographic history. We performed 25 pilot runs on the SNP dataset, each with a burn-in of 5,000 (default) followed by 1,000 iterations (default). We used two separate XtX significance thresholds of 99% and 99.9% quantiles to determine SNPs over-differentiated across all populations, i.e., under selection.

2.4.6 Environmental association analyses

We also performed gene-environment association (GEA) for identifying genomic signatures related to environmental factors using BayPass v2.2 (Gautier, 2015). We performed this analysis with adding the population-specific environmental variables under the standard covariance model (STD). This model allows to evaluate to which extent a population covariate is associated to each SNP through a regression coefficient. We performed five independent runs with the same parameters as the core model runs. The support for association between the genetic variables and the environmental variables was determined by a Bayes Factor (BF, measured in decibans (dB) units) which was estimated through the importance sampling estimator (IS) approach (Gautier, 2015). We considered SNPs associated with the environmental variable with a BF threshold of 5, 10 and 15 dB according to Jeffrey (1961).

In addition, we conducted a second GEA analysis using latent factor mixed models (LFMM) (Frichot & François, 2015, Caye et al., 2019). We first simulated individual values using the beta distribution of allele frequencies per pool, as LFMM requires individual values for input. For our analysis, we used the ridge regression method, which conducts a PCA to account for potential confounding effects. We chose 10 latent factors based on the number of main components identified in the PCA (Figure S2.1). We further conducted association testing using the fitted LFMM models and calibrated *p*-values using the genomic inflation factor (GIF). To identify candidate loci, we estimated the adjusted *p*-values using the Benjamini-Hochberg FDR multiple testing correction (Benjamini & Hochberg, 1995), and selected loci with adjusted *p*-values below 0.05 ($q < 0.05$).

2.4.7 SNP and gene functional annotation

To annotate all outlier SNPs that we identified through the genomic scan and GEA analyses, we used SnpEff v4.2 (Cingolani et al., 2012) against the database we had built *de novo* with ALB’s reference genome (McKenna et al., 2016). To obtain gene function annotation, we first extracted the DNA sequences from the reference genome that enclosed a total of 1,000 bp flanking an outlier SNP, that is 500 bp upstream and 500 bp downstream, respectively. We then blasted the sequences against ALB’s nr database and SwissProt database (Camacho et al., 2009). Additionally, we used EggnoG-mapper v2.1.7 (Cantalapiedra et al., 2021) to annotate protein sequences functions under the Diamond mode.

We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the R package *clusterProfiler*. GO annotations were obtained directly from the EggnoG annotation. To retrieve gene annotations and associated pathways for ALB, we used the *KEGGREST* package.

2.5 Results

2.5.1 Sequencing and genotyping

In total, 2.9 billion reads corresponding to 285 gigabases of compressed sequence data were generated from the ten pools. The average base sequence quality was 35.4 (Table S2.1), and 98.3% of reads were mapped to the reference genome. Using GATK, we discovered about 28.4 million SNPs with an average coverage of 68 per pool. After filtering based on the distribution of the annotation values (Figure S2.2) and read depth, we retained 4,313,126 SNPs. Bcftools identified about 14 million SNPs. Before filtering, the two SNP callers discovered 13 million SNPs in common, with GATK identifying significantly more low-frequency variants (Figure S2.3). After filtering based on the same filter criteria, we kept 3.83 M SNPs (Table 2.1).

We compared the Pool-seq derived allele frequencies with those from the individual-derived

GBS dataset for 5,136 SNPs and found that most of the allele frequencies were similar, with both methods showing a peak value of $r = 0.9$ (Figure S2.4).

Among the c.4 million filtered SNPs, we discovered on average one SNP every 159 bases. Of the filtered SNPs, 54% were found in intergenic regions, 16% in introns, 13% in downstream regulatory regions, and 12% in upstream regulatory regions of genes. Directly within exons, we found 5% of the SNPs, including 163,171 synonymous and 87,447 missense variants (Table S2.2).

Table 2.1: Variant detection in ALB using two different SNP callers. (M: million)

SNP caller	Original SNPs	Shared SNPs	Filtered SNPs
GATK	28.4 M		
bcftools	14.0 M	13 M	3.8M

2.5.2 Genetic diversity and phylogenetic analysis

The summary statistics calculated per 5 kb windows is presented in Table 2.2. Across the ten pools, the average number of segregating sites per window ranged from 89 (IMC) to 149 (PE), Watterson's θ from 0.0069 (TOL) to 0.0099 (PE), nucleotide diversity (Pi) from 0.0070 (TOL) to 0.0095 (PE), and *Tajima's D* from -0.1830 (PE) to 0.0878 (TOL).

In the phylogenetic analysis, we used a dataset of 109,062 SNPs with RAxML (Figure 2.1B) to examine the relationships among ten pools, five from each of the northern and southern regions. The phylogenetic tree showed that the five southern pools clustered together with strong bootstrap support, indicating their close genetic similarity. Among these southern pools, SHI and PNW appeared closely related, while PE and CIX were nested within this group, with CIX exhibiting a longer branch suggesting a greater genetic divergence. However, the phylogenetic pattern among the northern pools was less distinct compared to that of the southern pools. The bootstrap support for the node connecting the northern pools to their southern counterparts was relatively low, at 59%, indicating a less confident grouping.

Table 2.2: Population genetics parameters from NPStats for the 10 ALB populations. Note: MTCM = Min Temperature of Coldest Month. ISO = Isothermality. Pool information: PNW (pool 1): Pool of samples from Qingtongxia (QI) and Yanchi (YC); IMC (pool 2): Inner Mongolia; PN (pool 3): Pool of samples from Beijing (BJ) and Chengde (CHE); SHI (pool 4): Shijiazhuang; HS (pool 5): Hengshui; PE (pool 6): Pool of samples from Jinan (JI), Taian (TA) and Bengbu (BB); CIX (pool 7): Cixi. TOL (pool 8): Tongliao; PNE (pool 9): Pool of samples from Harbin (HRB), Changchun (CHC) and Shenyang (SHY); YJ (pool 10): Yanji. Pools can be divided into two major regions, north and south, according to the special structure findings by Cui et al (2022). S : average segregating sites per 5 kb; θ : Watterson estimator; Pi : average pairwise nucleotide diversity; $Tajima's D$: test of selective neutrality. Table is sorted based on the values of MTCM from smallest to largest.

Pool (#)	Region	MTCM (°C)	ISO (%)	S	θ	Pi	<i>Tajima's D</i>
PNE (9)	North	-21.767	22.67	106	0.0072	0.0072	0.0123
YJ (10)	North	-21.1	27	115	0.0085	0.0084	-0.0527
TOL (8)	North	-20.5	25	98	0.0069	0.007	0.0878
IMC (2)	North	-18.8	27	89	0.0074	0.0074	-0.033
PNW (1)	South	-14.45	30	111	0.0084	0.0083	-0.0369
PN (3)	North	-12.55	27.5	130	0.0087	0.0086	-0.0274
HS (5)	South	-8.6	26	129	0.0087	0.0087	0.0387
SHI (4)	South	-8.3	28	120	0.008	0.008	-0.037
PE (6)	South	-5.567	26.33	149	0.0099	0.0095	-0.183
CIX (7)	South	1.3	24	99	0.0072	0.0073	0.007

2.5.3 Mantel tests for geographical, genetic, and environmental distance

The test result showed a strong positive correlation ($r = 0.997$) between the genetic distance matrices obtained from Pool-seq (Figure 2.2) and individual-based GBS sets (Cui et al 2022), with high significance ($p = 0.0014$).

We found a moderate positive correlation ($r = 0.51$) between the genetic distance (F_{ST}) and the geographical distance matrix, but it was not significant ($p > 0.05$); further, there was a strong positive ($r = 0.69$) and significant correlation ($p < 0.05$) between the genetic distance and the environmental distance matrix of mean temperature of the coldest month (MTCM); For Isothermality (ISO), we observed only a weak and non-significant positive correlation ($r = 0.34$) with genetic distance.

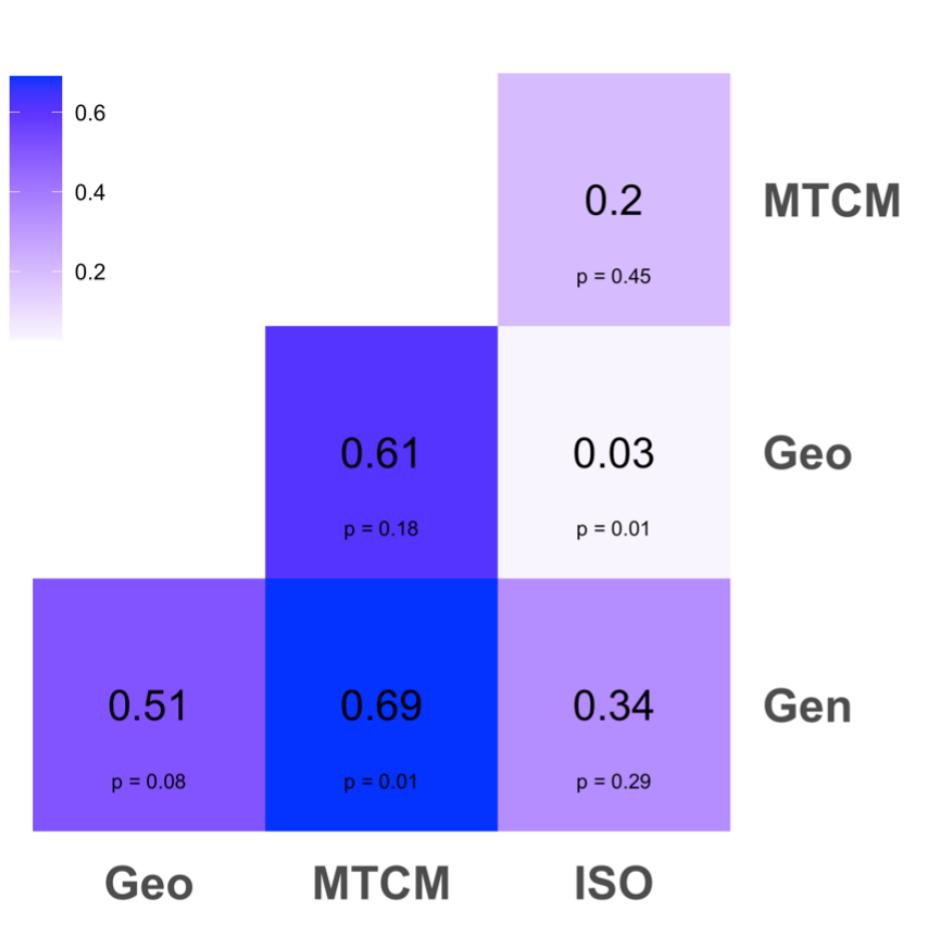


Figure 2.2: Correlations between genetic (Gen), geographical (Geo), and environmental (MTCM: mean temperature of the coldest month; ISO: isothermality) matrices. Each square of the heatmap represents the strength of the correlation, with the actual correlation coefficient and the corresponding *p*-value. *P*-value was derived by running 9999 permutations. Darker colors indicate a higher correlation coefficient according to the color code provided to the left.

2.5.4 Genome scans and outlier annotation

The BayPass null model was used to identify SNPs under balancing and diversifying selection, respectively. Under a 99-percentile threshold, we found, respectively, 34,424 SNPs under balancing and 35,684 SNPs under diversifying selection (Figure 2.3A). Applying a more stringent 99.9-percentile threshold, 3,754 SNPs were under balancing and 2,755 SNPs under diversifying selection, respectively. Of the total of 70,108 outliers (99-percentile), 56.64% of SNPs were in intergenic regions, 16.34% in introns, 8.68% in downstream regulatory regions, and 11.21% in upstream regulatory regions of genes. We found exonic SNPs (6.79%), of which 2,910 were synonymous and 1,569 missense variants (Table S2.3).

We annotated 5,359 genes out of all the 73,395 outlier SNPs (Table S2.4). Based on Gene On-

tology (GO) enrichment results, several key biological processes and molecular functions were identified: ‘transmembrane transport and transporter activity’, ‘peptidase and endopeptidase activities’, ‘guanyl-nucleotide exchange factor (GEF) activity’, ‘guanylate cyclase activity and cGMP metabolism’, ‘cell adhesion and development’, ‘signaling and receptor activity’, and ‘regulation of cellular processes’ (Figure S2.5, Table S2.5). There were five KEGG terms that were enriched including ‘fructose and mannose metabolism’, ‘glycolysis’, ‘citrate cycle’, ‘fatty acid degradation’ and ‘fatty acid biosynthesis’ (Figure S2.6).

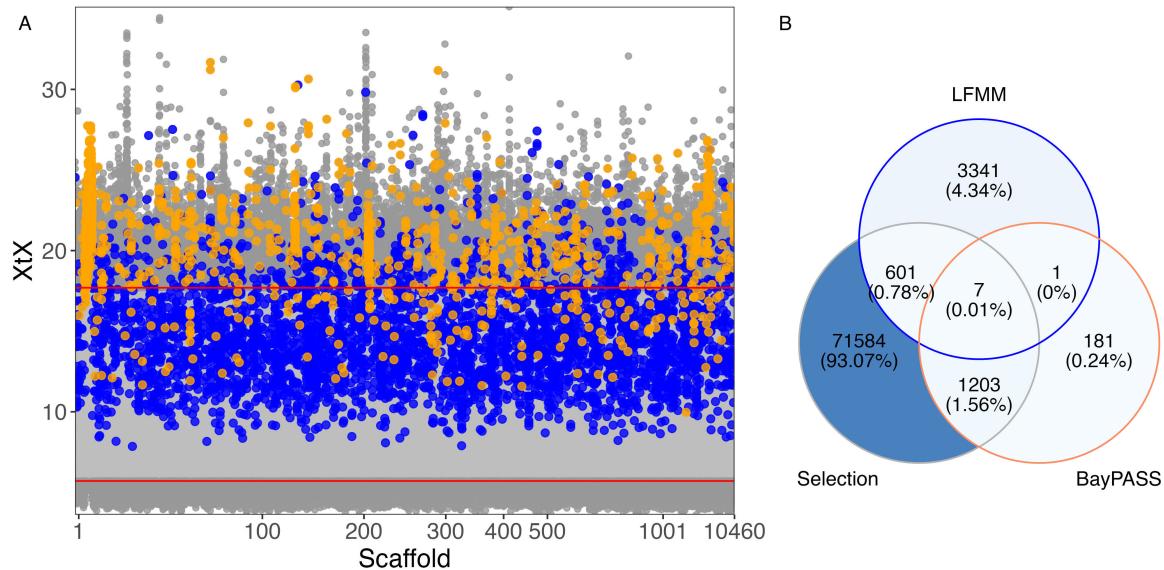


Figure 2.3: SNP outliers identified via genome-wide selection scan analysis (GWSS) and gene-environment analysis (GEA). **A** The x-axis represents scaffold positions, and the y-axis represents XtX values. Each dot represents a SNP. The two horizontal lines represent the 99th percentile of the XtX values. SNPs located below the lower line are outliers under balancing selection, while SNPs above the upper line are outliers under diversifying selection. Colored dots represent outliers (blue: LFMM; orange: BayPASS) associated with two environmental variables (MTCM; ISO). **B** Venn diagram illustrates the number and percentage of outliers shared among GWSS, LFMM GEA and BayPASS GEA.

2.5.5 Environmental association and outlier SNP/gene annotation

In the GEA analysis using BayPASS, we identified 1,392 SNPs with a Bayes Factor (BF) >5 dB. Among these, 305 SNPs were identified as outliers associated with MTCM and 1,271 SNPs as outliers associated with ISO, among which, 184 were shared outliers between both associations (Table S2.6).

In the LFMM analysis, we identified 3,950 SNPs with a q -value <0.05 . Among these, 2,710 SNPs were associated with MTCM and 2,439 SNPs with ISO, and among which, 1,199 were shared SNPs between the two associations (Table S2.7).

Of note, the GEA outliers identified in LFMM (Figure 2.3A; blue dots) overlapped with 608 outliers identified in the GWSS. Additionally, there were 1,210 SNPs in common between the BayPASS GEA analysis (Figure 2.3A; orange dots) and the GWSS. Notably, the two GEA methods only identified eight SNPs in common (Figure 2.3B).

When directly annotating the loci to the respective genes (Figure 2.4; Table S2.8), we found 87 genes in common between the two GEA methods, indicating that in these cases different SNPs, but within the same genes, were found by the two methods, demonstrating the need to perform gene annotation. Among these 87 genes, annotation information was available for 39 genes (Table 2.3). Each of these annotated genes contributes to different biological processes and pathways, such as transcription initiation factor, tubulin alpha-1 chain, ninjurin-2 isoform X1, etc. (Table 2.3).

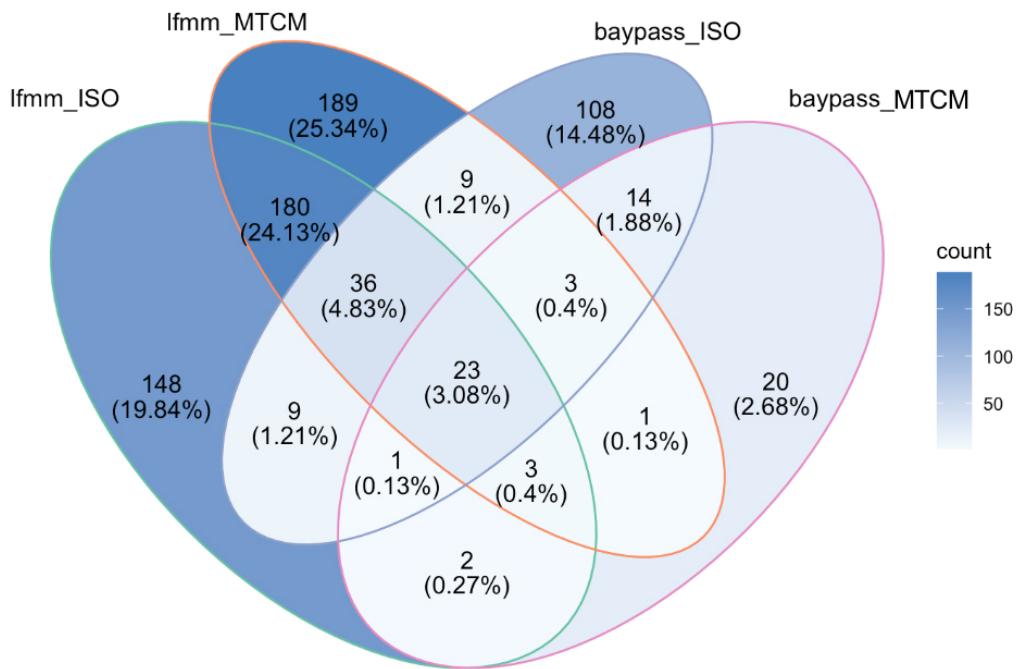


Figure 2.4: Shared numbers and percentages of genes found associated with MTCM (mean temperature of coldest month) and ISO (isothermality) using LFMM and BayPASS. Gene annotation is based on the blast results against the nr database of ALB 2.0.

Table 2.3: 39 shared genes identified with the two GEA analyses BayPASS and LFMM.

Accession_NR	Description_NR
XP_018561650.1	transcription initiation factor TFIID subunit 1 isoform X1
XP_018561662.1	tubulin alpha-1 chain
XP_018563037.1	ninjurin-2 isoform X1
XP_018564888.1	laminin subunit alpha-1-like
XP_018565941.1	Mariner Mos1 transposase OS=Drosophila mauritiana OX=7226 GN=mariner\T PE=1 SV=1
XP_018566333.1	zinc finger protein 184-like isoform X2
XP_018568328.1	pickpocket protein 28-like
XP_018568330.1	esterase FE4 isoform X2
XP_018569664.1	zinc finger protein basonuclin-2-like
XP_018573044.2	sonic hedgehog protein A
XP_018573715.1	putative mediator of RNA polymerase II transcription sub-unit 12
XP_018576492.1	PWWP domain-containing protein 2A
XP_018579404.1	probable serine/threonine-protein kinase kinX
XP_018579406.1	adenylate cyclase type 10-like
XP_018579563.1	echinoderm microtubule-associated protein-like 2
XP_018579901.1	transmembrane protein 131
XP_018580240.1	tubulin glycylase 3A-like
XP_023310320.1/	Probable RNA-directed DNA polymerase from transpo-
XP_023310562.1/	son BS OS=Drosophila melanogaster OX=7227 GN=RTase
XP_023311922.1/	PE=2 SV=1
XP_018569138.1	
XP_023310522.1	Transposable element Tcb1 transposase OS=Caenorhabditis briggsae OX=6238 PE=3 SV=1
XP_023310750.1/	histone-lysine N-methyltransferase SETMAR-like
XP_018568948.1/	
XP_023310010.1/	
XP_023310047.1/	
XP_023310469.1/	
XP_018564185.1/	
XP_018572538.2	
XP_023310853.1	neurexin-1-like
XP_023310855.1	myosin-3
XP_023311400.1	transcription factor IIIB 90 kDa subunit
XP_023311434.1	homeobox protein engrailed-2-A
XP_023311966.1/	Protein GVQW3 OS=Homo sapiens OX=9606
XP_018566191.1	GN=GVQW3 PE=2 SV=1
XP_023312445.1	pancreatic triacylglycerol lipase-like, partial
XP_023313014.1	Retrovirus-related Pol polyprotein from transposon 17.6 OS=Drosophila melanogaster OX=7227 GN=pol PE=4 SV=1
XP_023313090.1	vicilin-like seed storage protein At2g18540
XP_023313184.1	LOW QUALITY PROTEIN: KH domain-containing, RNA-binding, signal transduction-associated protein 2-like, partial

2.6 Discussion

2.6.1 Geographical influences on ALB population genetic differentiation

Natural populations often exhibit population genetic differentiation due to factors like geographic distance and barriers preventing random mating. While drift can lead to this differentiation, natural selection also plays a role. The challenge lies in determining the extent to which these factors influence the genetic variation.

Previous, individual GBS-based genotyping data indicated the existence of population structure among the ten selected ALB pools due to the insect's historical biogeographical structuring and the overall low contemporary gene flow among native ALBs, except for within certain regions (Cui et al., 2022). In the present study, we explored the genetic diversity and phylogenetic relationship of these pools, the results of which generally align with the findings from the individual GBS study (Cui et al., 2022). As evidenced in the phylogenetic tree, the southern pool populations displayed strong statistical support, while the northern ones exhibited less. This could be attributed to the presence of admixture in the BJ, IMC, and YJ populations, as identified by Cui et al. (2022). This added complexity to the interpretation of the evolutionary relationships within these populations.

Isolation-by-distance (IBD) was moderate ($r = 0.5$), which is comparable to findings with GBS data (Cui et al., 2022). The moderate IBD correlation suggests that genetic differentiation is partially influenced by geographic distance, where populations that are closer together geographically tend to have more similar genetic compositions than those located further apart. Despite this observation, the lack of statistical significance ($p = 0.075$) implies that this relationship between genetic and geographic distance may not be straightforward and could be influenced by other factors.

With limited gene flow, historical biogeography and random genetic drift have impacted the genomes differently, resulting in population differentiation, differences in allele frequencies and genetic diversity among ALB populations. As for how local adaptative processes impacted population differentiation in ALB, we lack prior knowledge to understand the underlying factors and their interaction. Here, we were interested in understanding how much natural selection contributed to population differentiation.

2.6.2 The role of natural selection in population differentiation: Isolation-by-environment

Besides IBD, the climate variables showed a latitudinal gradient, as temperature drops continually from southern to northern regions (Figure 2.1A). It can be difficult to distinguish IBD from isolation-by-environment (IBE) when the environmental variables covary with geographic distance (Tiffin & Ross-Ibarra, 2014; Porth et al., 2015). It is known that climatic

factors such as temperature and precipitation are important abiotic factors affecting survival, growth, and reproduction in insects. Insects are known to be highly sensitive to temperature due to their heterothermic nature and high surface area-to-volume ratio (Schowalter, 2022). In that case, given the low gene flow levels between ALB populations, it is likely the different selection pressures that resulted in the observed population differentiation patterns. Sexton et al. (2014) surveyed 70 studies that investigated IBE and IBD and found that IBE is the predominant pattern of gene flow restriction among invertebrates, suggesting environment factors play an important part in shaping population differentiation for invertebrates. ALB as a temperate insect is not only capable of overwintering under the broad temperature range found within its native geographic distribution but also within its invaded regions with extremely cold temperatures during winter, owing to its cold tolerance.

Climate as a key environmental factor in population differentiation

Geography is a factor influencing genetic variation; the environment also plays an important role. As mentioned above, climate factors highly impact population differentiation for invertebrates (Sexton et al., 2014). Climate is notably one critical environmental factor in ALB populations as the temperature of habitats of ALB populations can range from above zero (CIX) to -25°C (HRB) (Figure 2.1).

Huang et al. (2020) investigated the impact of various environmental factors on the occurrence of ALB in China and discovered that temperature and precipitation had the most extensive impacts on ALB. Association analysis allows us to link temperature variables to each genetic variable and identify which loci are associated with the traits/factors of interest. Here, we examined two environmental factors, mean temperature of the coldest month (MTCM) and isothermality (ISO) and identified candidate genes that responded to them in terms of differences in their allelic patterns.

The Isolation-By-Environment (IBE) test showed a strong correlation between the genetic distance and the Mean Temperature of Coldest Month (MTCM), with a Pearson correlation coefficient of 0.69 ($p = 0.01$). This environmental factor also showed a positive correlation with geographical distance, although the correlation was less strong ($r = 0.61$) and not statistically significant ($p = 0.18$). This suggests a complex interplay between the MTCM environmental factor and geography influencing the genetic differentiation in ALB populations. Spatial autocorrelation among environmental variables is commonly observed (Wang 2013; Shafer & Wolf, 2013). It is recommended to account for any correlation between environmental and geographic distances, as isolation by environment (IBE) should be independent of geographic distance (Wang et al., 2014).

The other environmental factor ISO, which measures how similar the temperature patterns are throughout the year, does not correlate with any other factors. In this case, it may be

easier to detangle the allelic effects for selected genes associated with this environmental factor from any confounding effect due to population structure.

The effect of climate factors on genes

The two factors used in our GEA analyses, MTCM, which is used to describe the coldness of a climate, and ISO, which is a measure for the amplitude of temperature fluctuations in a given region. We identified 87 shared genes or gene families (including 39 annotated genes as shown in Table 2.3) through GEA methods. These genes can all be found in differentially expressed gene sets discovered by Torson et al., (2023) in various tissues of ALB larvae responding to temperature changes, specifically from chilling to warm temperatures, during diapausing. Although they identified a large proportion of genes in their set, the high congruency between GEA and gene expression methods demonstrates the effectiveness of combining these two approaches for a more comprehensive understanding of gene-environment interactions. We also found many GO terms in common involved in their ALB diapause gene expression profiling, such as ‘catalytic activity’, ‘protein alkylation’, and ‘protein methyltransferase activity’.

Certain genes such as ninjurin genes are identified among the first being up-regulated in response to the peripheral nerve injury and have been found in a variety of species, e.g. humans, mice and *Drosophila*, and known to regulate multiple biological processes including injury, infection, immunity etc (Araki & Milbrandt, 1996; Thomas et al., 2014). We found multiple histone-lysine methylation genes which regulate gene methylation in a reversible manner. In some invertebrates, such as insects, DNA methylation has been shown to play a role in gene regulation and development, but the mechanisms and functions of DNA methylation in invertebrates are still being studied (Rose & Klose, 2014; Ren et al., 2021). These genes have also been identified to be associated with diapause in *Megachile rotundata* (Yocum et al., 2015) and between diapause and non-diapausing *Helicoverpa armigera* brains (Lu et al., 2013), as well as in diapause of ALB (Torson et al., 2023).

There is a consistent pattern of these genes being associated with cold response in insects, while it is not yet clear whether they play a direct role in conferring cold tolerance or if they are simply correlated with other factors that contribute to cold tolerance. Further research is needed to determine the exact role these genes may play in cold tolerance and their potential mechanisms of action in insects.

The impact of diverse vegetation and complex geography on ALB population adaptation

Through GWSS, we found several genes related to cryoprotectant regulatory processes. As we know, insects rely on cryoprotectants, such as, glycerol, trehalose, proline, glucose, mannose, lipid, and antifreeze proteins to overwinter (Storey & Storey, 2012). We found one glycerol kinase (GK) gene *XP_018561128.1* (Table S2.4). GK and glycerol-3-phosphate dehydroge-

nase (GPDH) which are involved in utilization of glycerol have been addressed in many cold tolerant studies on insects and were also shown to be up-regulated as a response to rapid cold hardening in red fire ants (Vatanparast, Sajjadian & Park, 2022). Similar results have been observed in a study of the oriental tobacco budworm by Cha & Lee, and they also showed that suppression of the two genes by RNA interference reduced the survival rate of the budworm under cold temperature (Cha & Lee, 2022). We also identified several lipid genes, such as diacylglycerol kinase encoding *DGKD* (diacylglycerol kinase delta) and *DGK3* (diacylglycerol kinase gamma), whose gene products are enzymes that convert diacylglycerol to phosphatidic acid. They play important roles in modulating lipid signaling pathways and are involved in various cellular processes, including stress responses (Sakane et al., 2007). We found genes that participate in lipid transporter activity which have previously been found enriched among differentially expressed genes under cold exposure in a transcriptomic study of a spider species (Lv et al., 2020). We also found fatty-acyl-CoA binding protein ACBD3 which belongs to a protein family coordinated with acyl-CoA esters that are key intermediates in numerous lipid metabolic pathways (Neess et al., 2015). We also found gustatory receptor genes which have been found to not only help insects detect volatile chemicals but also serve as thermosensors in *Drosophila* (Li & Gong, 2017; Montell, 2013).

Besides small molecular cryoprotectants encoding genes, heat shock proteins (HSPs) are also well known as protein families that respond to heat/cold temperature stress. We found *HSP70* and *HSP20* in our gene set. They play important roles in cellular stress responses, protein folding, and cellular protection (Feder & Hofmann, 1999). HSPs are molecular chaperones that are upregulated in response to various stress conditions, such as heat, cold, and other environmental or cellular stresses and have been frequently found in studies on insects (King & MacRae, 2015; Cui et al., 2017; Torson et al., 2023). We found genes that are also often identified in cold tolerance studies on insects, such as cytochrome P450 proteins which are involved in metabolism but also detoxification; cuticle proteins that form the exoskeleton of insects and other arthropods; as well as ribosomal proteins which are essential for protein synthesis and cellular functions (Enriquez & Colinet, 2019; Holmstrup et al., 2022; Qin et al., 2019; Vatanparast & Park, 2021, 2022; Xu et al., 2021; Zhao et al., 2021).

Beyond the abiotic climatic factors discussed earlier, biotic factors, such as host-insect interactions (the utilization of hosts by the insect), can act as important drivers of adaptive genetic variation among ALB populations. Other than the diverse climate features and complex geographical patterns in China, the country's vegetation is characterized by immense variation, offering a complex mosaic of hosts for the insects to utilize. It varies from deciduous to evergreen broadleaved forests, to grassland type (Zhao, 2010). This diversity in host availability could influence the genetic adaptations of the beetles, further complicating the factors that drive their evolutionary trajectory. Through our GO enrichment analysis, we found several enriched digestive enzymes associated with specific GO terms, such as alpha-1,4-glucosidase

activity, alpha-glucosidase activity and alpha-mannosidase activity, and terms that involve degradation of proteins, such as serine hydrolase activity, serine-type endopeptidase activity, and serine-type peptidase activity. Related genes such as glycoside hydrolase 28 protein (GH28) were previously found critical for ALB to help degrade rigid plant cell walls (McKenna et al., 2016).

2.7 Conclusion

Owing to fast-evolving sequencing technologies that are also increasingly cost-efficient allowed us to conduct WGS on pooled ALB populations and scan the entire genome to identify genes underlying adaptation. In comparison to individual-based genotyping methods, we examined how robust Pool-seq is regarding allele frequency estimation. We obtained high accuracy of allele frequencies despite having less samples pooled than the recommended pool size (>50 individuals per pool) (Fuentes-Pardo & Ruzzante, 2017). Through genome scan and GEA analyses, we identified a set of candidate genes putatively under selection. Further validation is needed to assess the candidate genes for their biological significance to an increased adaptive potential of ALB. However, the genes we identified in our GEA results are not commonly known to be insect cold tolerance-related genes (Table 2.3).

Temperature plays an important role in shaping the genetic differences among the populations. The sheer number of genes that might respond to changes in temperature, as evidenced by gene expression studies (cf. Torson et al. 2023), complicates the process of identifying the key gene sets related to temperature adaptation. The exact function and relevance of candidate genes associated with cold tolerance require further validation.

Considering the relationship between geography and specific environmental factors affecting ALB populations, it remains challenging to account for confounding effects resulting from population structure. Addressing this issue will aid in distinguishing authentic adaptive variants from false positives more effectively.

While genomics serves as a crucial resource for understanding evolution and adaptation, epigenomic factors may also significantly influence gene expression and function across different seasons. Therefore, it would be valuable to investigate the impact of epigenomic effects on overwintering in future studies. Our current findings could be applied to support assessment related to the adaptation of invasive populations and, more generally, to their biosurveillance using targeted genomics tools based on genetic markers of individuals' adaptive potential.

2.8 Supplementary materials

The supplemental tables associated with this work are available for access under data repository: <https://datadryad.org/stash/share/fCbWSHakbd2HBPpTFYfbVTBo3UNjnxKGNoJ1wkukm0>.

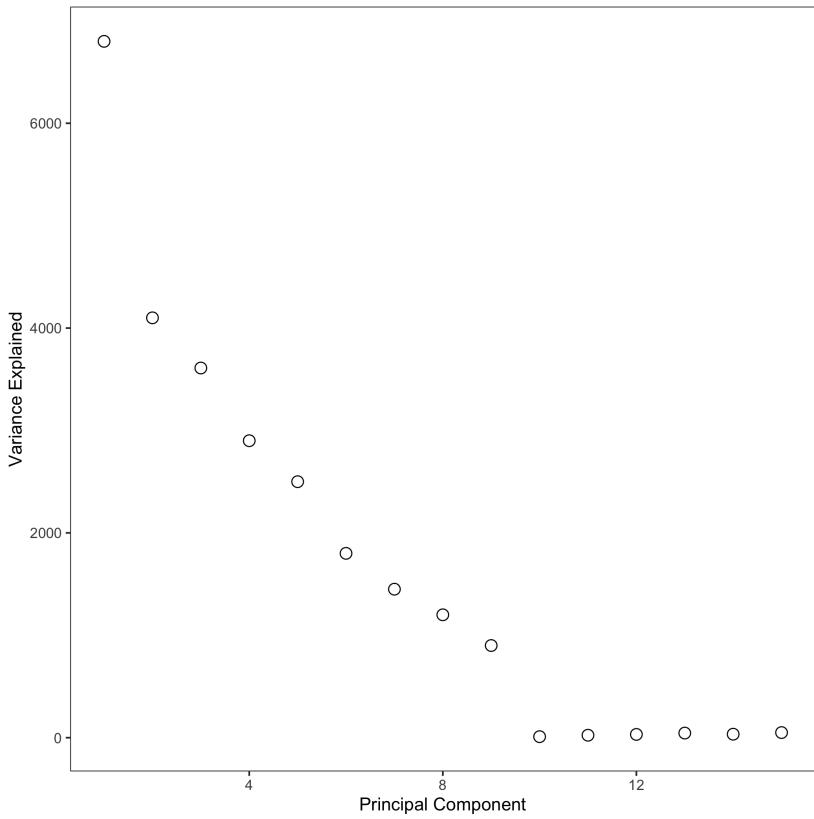


Figure S2.1: Screeplot of LFMM analysis showing the percentage of variance explained by each component in a principal component analysis. The elbow point at $K = 10$ indicates the presence of ten major genetic clusters in the data.

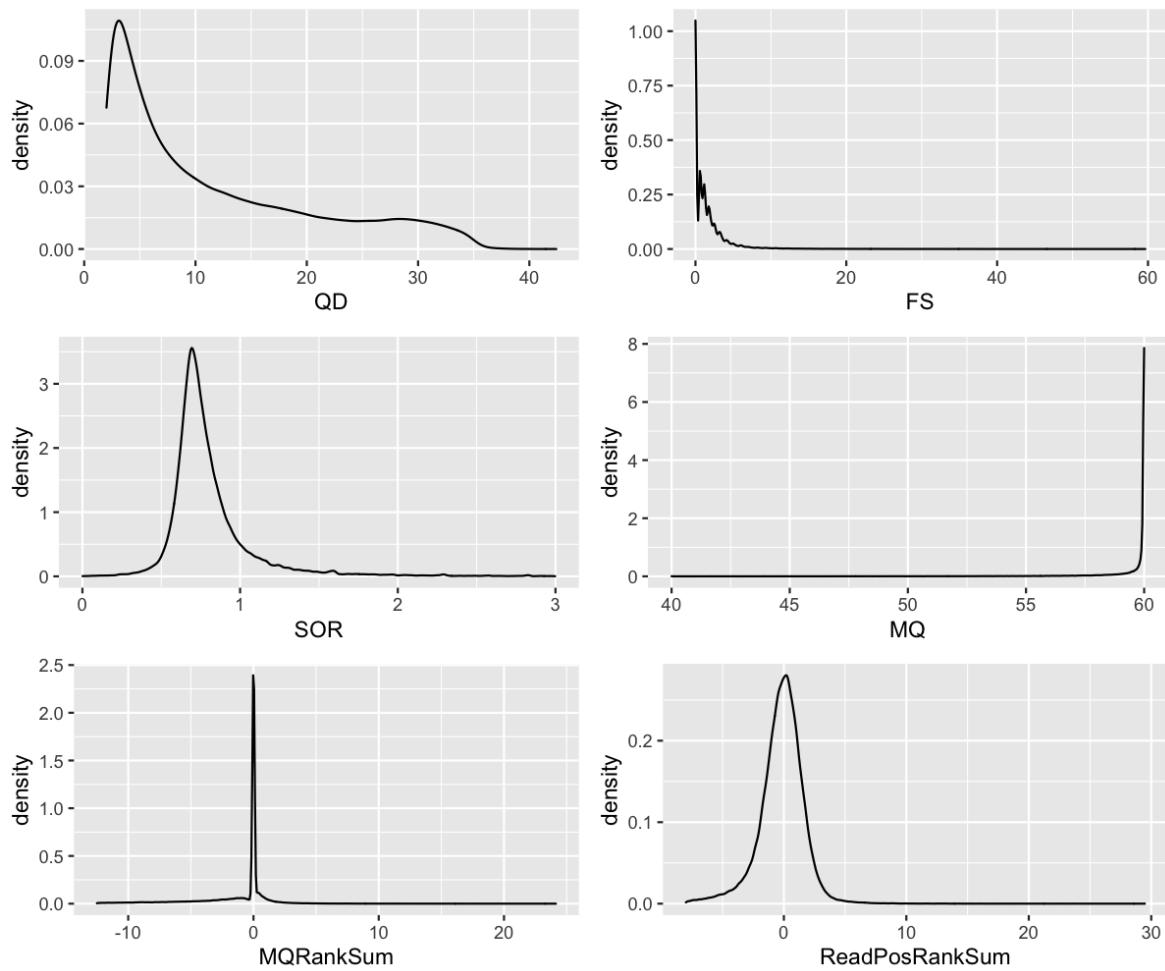


Figure S2.2: Annotation values distribution of all SNPs called in GATK. QD: quality of variant by depth of variant. FS: FisherStrand. SOR: StrandOddsRatio. MQ: RMSMappingQuality. MQRankSum: MappingQualityRankSumTest. ReadPosRankSum: ReadPosRankSumTest.

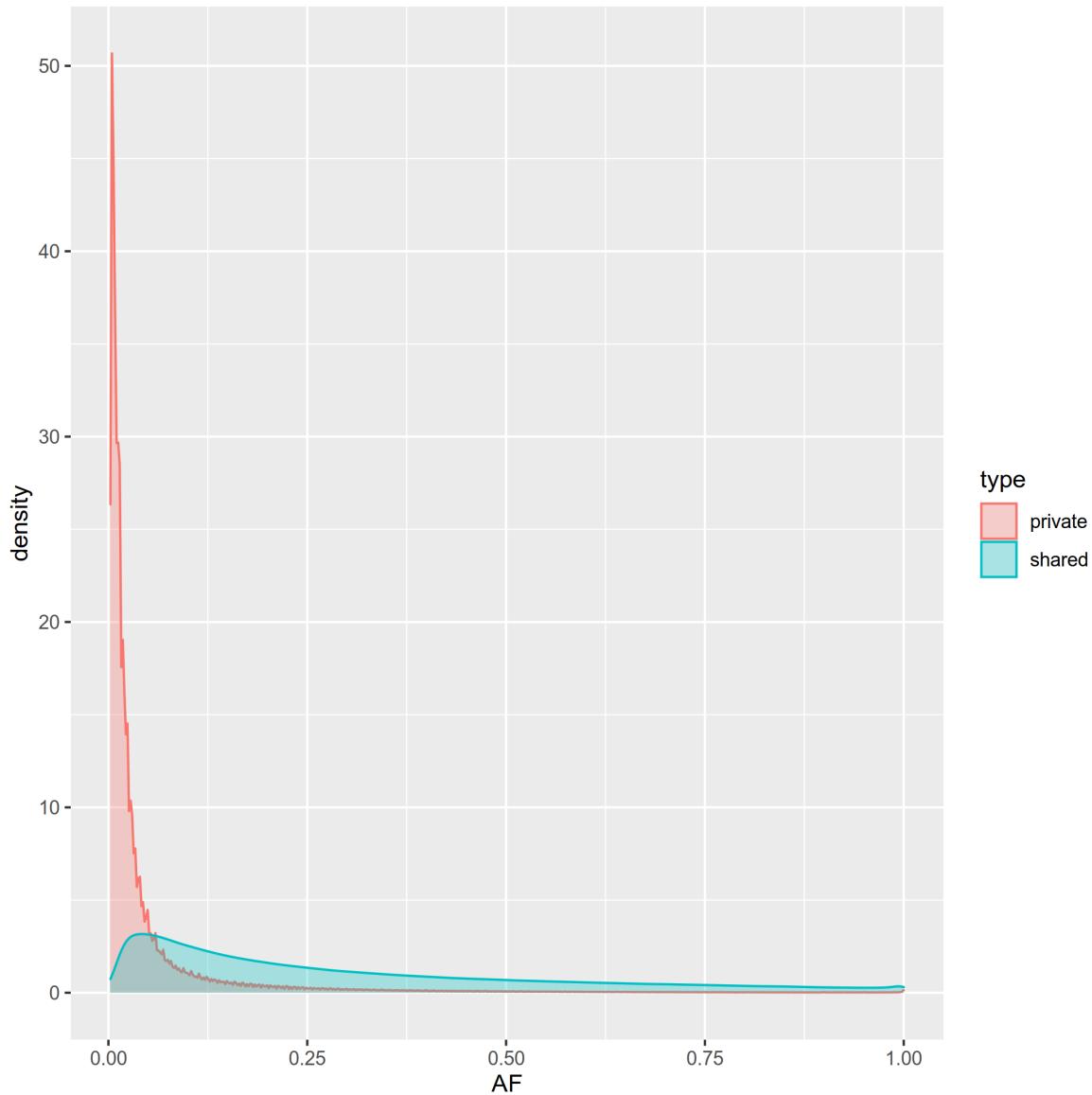


Figure S2.3: Allele frequency distribution of shared SNPs between the two SNP callers. Blue shading indicates shared SNPs between the two SNP callers, GATK and bcftools. Red shading shows SNPs only found in GATK.

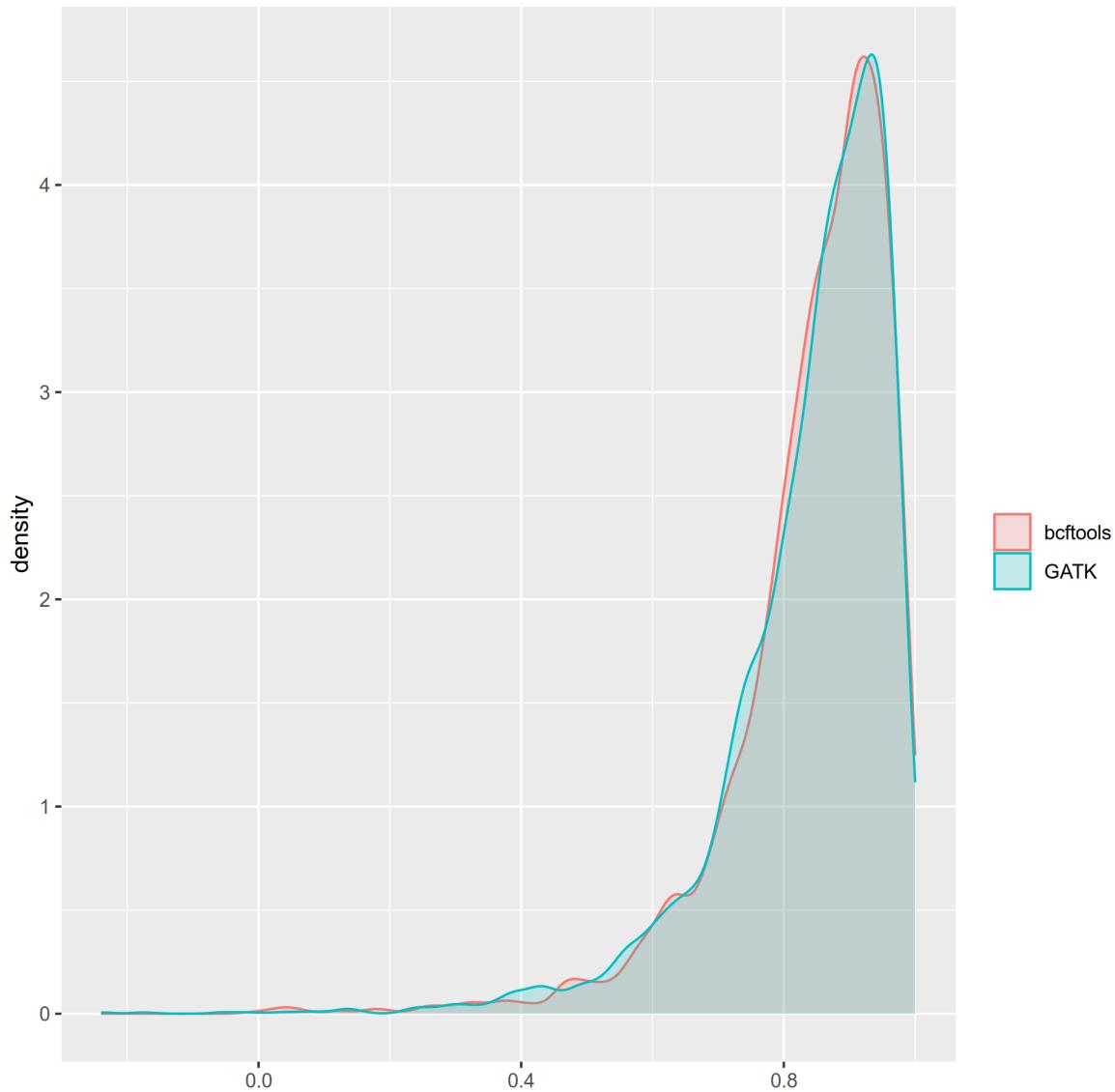


Figure S2.4: Poolseq and individual-genotyping comparison. 3000 SNPs are used in the comparison. Red shading represents the comparison between SNPs generated using bcftools and those obtained from individual genotyping by Cui et al. (2022), while blue shading indicates the comparison between SNPs called by GATK and the individual genotyping.



Figure S2.5: GO enrichment analysis of all the genes putatively under selection. The top 30 terms are displayed, and the complete list can be found in the Table S2.5.

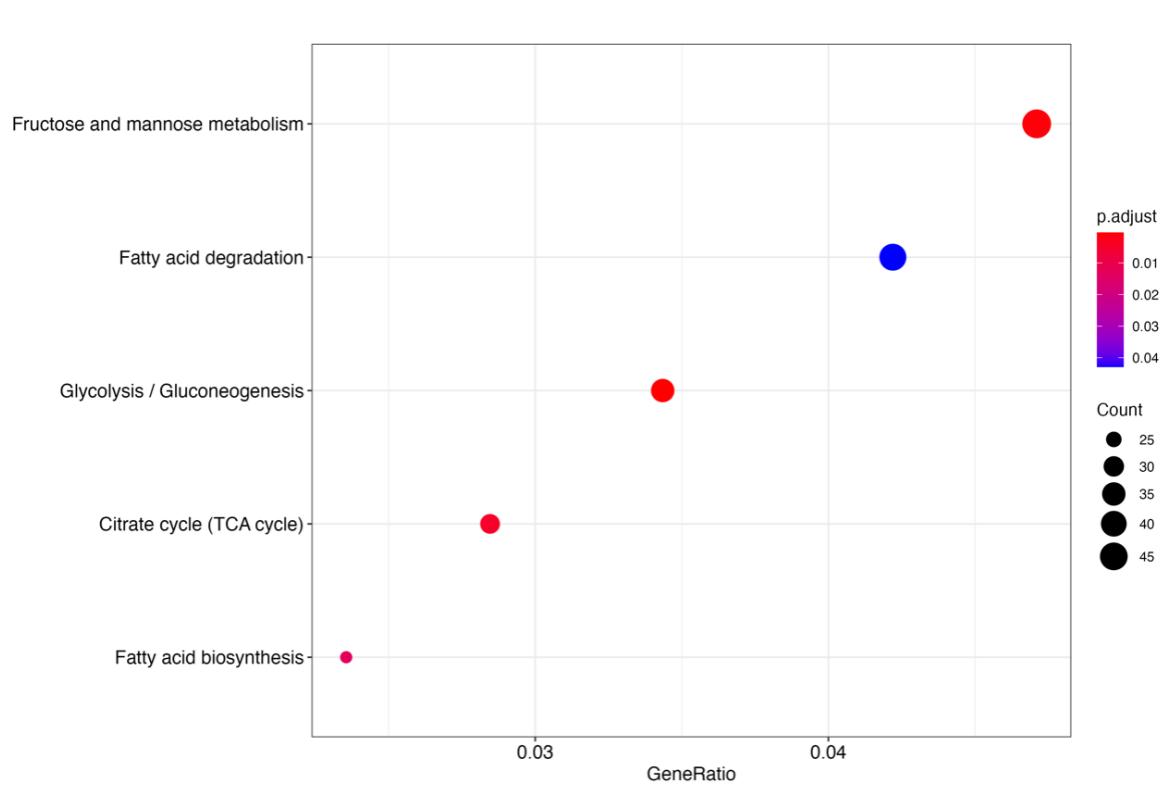


Figure S2.6: KEGG enrichment analysis of all genes identified as under selection.

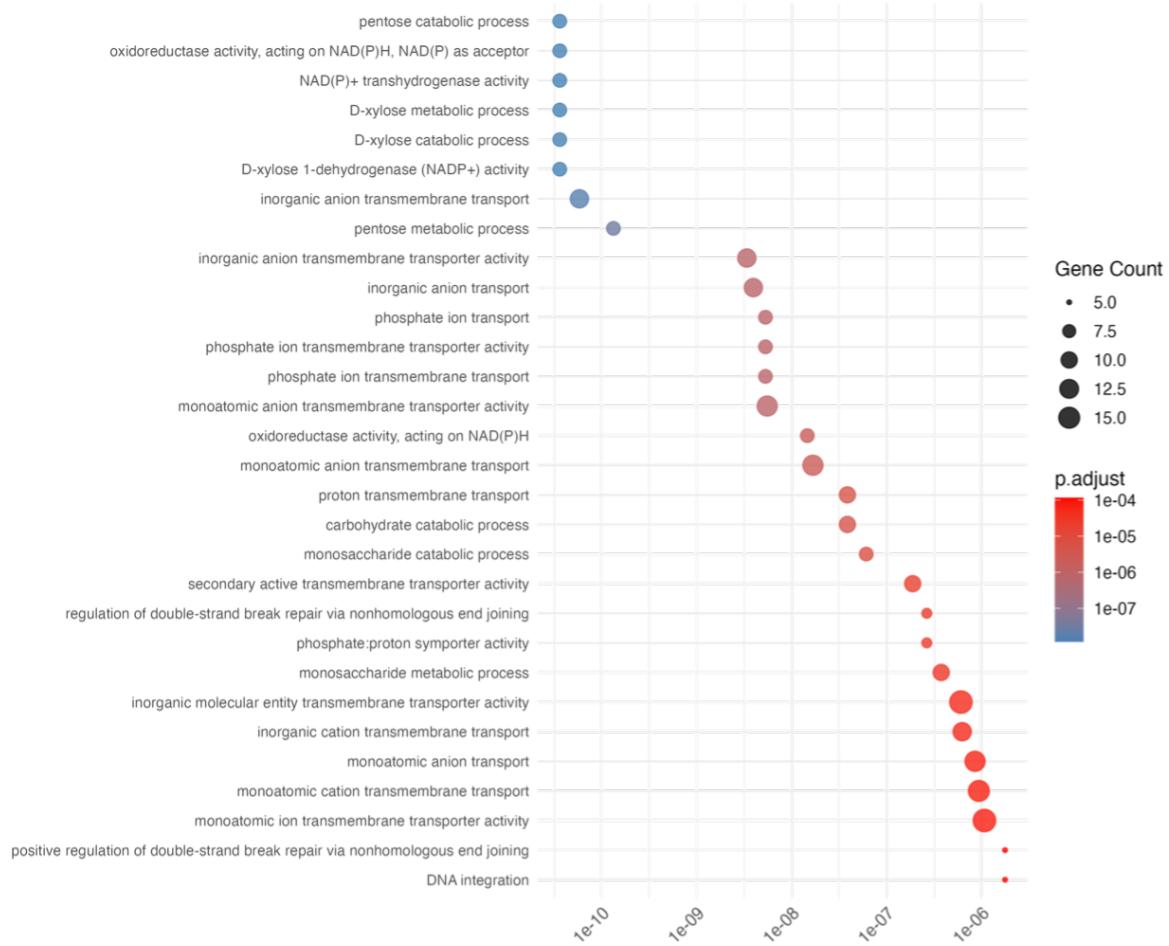


Figure S2.7: GO enrichment analysis of the 87 shared genes. The top 30 terms are displayed, and the complete list can be found in the Table S2.9.

2.9 References

- Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander, E. S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, *407*, 513–516.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Araki, T., & Milbrandt, J. (1996). Ninjurin, a novel adhesion molecule, is induced by nerve injury and promotes axonal growth. *Neuron*, *17*(2), 353-361.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*, *43*. <https://doi.org/10.1002/0471250833.bi43011s43>

//doi.org/10.1002/0471250953.bi1110s43

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Broad Institute. (2019). Picard Toolkit.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421. doi: 10.1186/1471-2105-10-421

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38, 5825–5829.

Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. *Link to MBE*, 36(4), 852–860.

Cha, W. H., & Lee, D. (2022). Suppression of glycerol biosynthesis-related genes decreases the effect of rapid cold hardening in *Helicoverpa assulta*. *Entomological Research*, 52(1), 24–32. doi: 10.1111/1748-5967.12561

Chen, Z., Wang, G., Li, M., Peng, Z., Ali, H., Xu, L., ... Hou, Y. (2020). Development of single nucleotide polymorphism (SNP) markers for analysis of population structure and invasion pathway in the coconut leaf beetle *Brontispa longissima* (gestro) using restriction site-associated DNA (RAD) genotyping in Southern China. *Insects*, 11, 230.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6, 80–92.

Cui, M., Hu, P., Wang, T., Tao, J., & Zong, S. (2017). Differential transcriptome analysis reveals genes related to cold tolerance in seabuckthorn carpenter moth, *Eogystia hippophaecolus*. *PLoS One*, 12(11), e0187105.

Cui, M., Wu, Y., Javal, M., Giguère, I., Roux, G., Andres, J. A., Keena, M., Shi, J., Wang, B., Braswell, E., Pfister, S. E., Hamelin, R., Roe, A., & Porth, I. (2022). Genome-scale phylogeography resolves the native population structure of the Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky). *Evolutionary Applications*, 15(March), 934–953. <https://doi.org/10.1111/eva.13381>

- Denlinger, D. L., Lu, Y. X., & Xu, W. H. (2013). Polycomb repressive complex 2 (PRC2) protein ESC regulates insect developmental timing by mediating H3K27me3 and activating prothoracicotropic hormone gene expression. *Journal of Biological Chemistry*, 288(32), 23554-23564.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. doi: 10.1038/ng.806
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379.
- Enriquez, T., & Colinet, H. (2019). Cold acclimation triggers major transcriptional changes in *Drosophila suzukii*. *BMC Genomics*, 20(1), 413. doi: 10.1186/s12864-019-5745-7
- Feder, M. E., & Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual review of physiology*, 61(1), 243-282.
- Ferretti, L., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2013). Population genomics from pool sequencing. *Molecular Ecology*, 22(22), 5561–5576. doi: 10.1111/mec.12522
- Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., ... Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, 22, 5594–5607.
- Frichot, E., & François, O. (2015). LEA: an R package for Landscape and Ecological Association studies. *Methods in Ecology and Evolution*, 6(8), 925-929.
- Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26, 5369–5406.
- Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., Morin, P. A., Narum, S. R., O'Brien, S. J., Roffler, G., Templin, W. D., Sunnucks, P., Strait, J., Warheit, K. I., Seamons, T. R., Wenburg, J., Olsen, J., Luikart, G., O'Brien, S. J., ... Luikart, G. (2016). Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. *Trends in Ecology & Evolution*, 31(2), 81–83. <https://doi.org/10.1016/j.tree.2015.10.009>
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, 201, 1555–1579.
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M.,

- Pudlo, P., Kerdelhué, C., & Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, 22, 3766–3779. <https://doi.org/10.1111/mec.12360>
- Gould, B. A., Chen, Y., & Lowry, D. B. (2017). Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Molecular ecology*, 26(1), 163-177.
- Günther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, 195, 205–220. doi: 10.1534/genetics.113.152462
- Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2017). Package ‘geosphere’. *Spherical trigonometry*, 1(7), 1-45.
- Hohenlohe, P. A., Hand, B. K., Andrews, K. R., & Luikart, G. (2018). Population Genomics Provides Key Insights in Ecology and Evolution. *Population Genomics: Concepts, Approaches and Applications*, 483–510. doi: 10.1007/13836_2018_20
- Holmstrup, M., Sørensen, J. G., Dai, W., Krogh, P. H., Schmelz, R. M., & Slotsbo, S. (2022). Analysis of heat and cold tolerance of a freeze-tolerant soil invertebrate distributed from temperate to Arctic regions: evidence of selection for extreme cold tolerance. *Journal of Comparative Physiology B*, 192(3–4), 435–445. doi: 10.1007/s00360-022-01433-w
- Huang, J., Qu, B., Fang, G., Li, X., & Zong, S. (2020). The drivers of the Asian longhorned beetle disaster show significant spatial heterogeneity. *Ecological Indicators*, 117, 106680.
- Jeffreys, H. (1961). Theory of probability, 3rd edn. Oxford: Oxford University Press.
- Keena, M. A. (2006). Effects of temperature on *Anoplophora glabripennis* (Coleoptera: Cerambycidae) adult survival, reproduction, and egg hatch. *Environmental Entomology*, 35, 912–921. <https://doi.org/10.1603/0046-225x-35.4.912>
- Keena, M. A., & Moore, P. M. (2010). Effects of Temperature on anoplophora glabripennis (Coleoptera: Cerambycidae) larvae and pupae. *Environmental Entomology*, 39, 1323–1335. <https://doi.org/10.1603/en09369>
- King, A. M., & MacRae, T. H. (2015). Insect heat shock proteins during stress and diapause. *Annual review of entomology*, 60, 59-75.
- Krueger, F. (2015). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2012; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- Koot, E., Arnst, E., Taane, M., Goldsmith, K., Thrimawithana, A., Reihana, K., ... & Chagné, D. (2022). Genome-wide patterns of genetic diversity, population structure and

demographic history in mānuka (*Leptospermum scoparium*) growing on indigenous Māori land. *Horticulture research*, 9.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987–2993.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Li, K., & Gong, Z. (2017). Feeling Hot and Cold: Thermal Sensation in *Drosophila*. *Neuroscience Bulletin*, 33(3), 317–322. doi: 10.1007/s12264-016-0087-9

Lischer, H., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. doi: 10.1093/bioinformatics/btr642

Lu, Y. X., Denlinger, D. L., & Xu, W. H. (2013). Polycomb repressive complex 2 (PRC2) protein ESC regulates insect developmental timing by mediating H3K27me3 and activating prothoracotropic hormone gene expression. *Journal of Biological Chemistry*, 288(32), 23554–23564.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, 4, 981–994.

Lv, B., Wang, J., Zhuo, J., Yang, H., Yang, S., Wang, Z., & Song, Q. (2020). Transcriptome sequencing reveals the effects of cadmium toxicity on the cold tolerance of the wolf spider *Pirata subpiraticus*. *Chemosphere*, 254, 126802. doi: 10.1016/j.chemosphere.2020.126802

McKenna, D. D., Scully, E. D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S. M., ... Richards, S. (2016). Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology*, 17, 1–18.

Montell, C. (2013). Gustatory Receptors: Not Just for Good Taste. *Current Biology*, 23(20), R929–R932. doi: 10.1016/j.cub.2013.09.026

Neess, D., Bek, S., Engelsby, H., Gallego, S. F., & Faergeman, N. J. (2015). Long-chain acyl-CoA esters in metabolism and signaling: role of acyl-CoA binding proteins. *Progress in Lipid Research*, 59, 1–25.

Nowak, D. J., Pasek, J. E., Sequeira, R. A., Crane, D. E., & Mastro, V. C. (2001). Potential effect of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) on urban trees in the United States. *Journal of Economic Entomology*, 94, 116–122. <https://doi.org/10.1603/0022-0493-94.1.116>

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggs, M. A. S. S. (2007). The vegan package. *Community ecology package*, 10(631-637), 719.

Pedlar, J. H., McKenney, D. W., Yemshanov, D., Hope, E. S., & Wallin, K. (2020). Potential Economic Impacts of the Asian Longhorned Beetle (Coleoptera: Cerambycidae) in Eastern Canada. *Journal of Economic Entomology*, 113, 839–850. <https://doi.org/10.1093/jee/toz317>

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135.

Porth, I., Klápková, J., McKown, A. D., Mantia, J. L., Guy, R. D., Ingvarsson, P. K., Hamelin, R., Mansfield, S. D., Ehlting, J., Douglas, C. J., & El-Kassaby, Y. A. (2015). Evolutionary Quantitative Genomics of *Populus trichocarpa*. *PLoS ONE*, 10(11), e0142864. <https://doi.org/10.1371/journal.pone.0142864>

Qin, M., Wang, H., Liu, Z., Wang, Y., Zhang, W., & Xu*, B. (2019). Changes in cold tolerance during the overwintering period in *Apis mellifera ligustica*. *Journal of Apicultural Research*, 58(5), 702–709. doi: 10.1080/00218839.2019.1634461

Ren, K., Mou, Y. N., Tong, S. M., Ying, S. H., & Feng, M. G. (2021). SET1/KMT2-governed histone H3K4 methylation coordinates the lifecycle in vivo and in vitro of the fungal insect pathogen *Beauveria bassiana*. *Environmental Microbiology*, 23(9), 5541-5554.

Roden, D. B., Haack, R. A., Keena, M. A., McKenney, D. W., Beall, F. D., & Roden, P. M. (2009). Potential northern distribution of Asian longhorned beetle in North America. *U.S. Department of Agriculture Interagency Research Forum on Invasive Species 2008*, 65–67. <https://www.nrs.fs.fed.us/pubs/gtr/gtr-nrs-p-36papers/48roden-p-36.pdf>

Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1839(12), 1362-1372.

Sakane, F., Imai, S. I., Kai, M., Yasuda, S., & Kanoh, H. (2007). Diacylglycerol kinases: why so many of them?. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 1771(7), 793-806.

- Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15, 749–763. <https://doi.org/10.1038/nrg3803>
- Schowalter, T. (2022). Insect ecology: an ecosystem approach. Academic press.
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, 68(1), 1–15. doi: 10.1111/evo.12258
- Shafer, A. B., Wolf, J. B., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., ... & Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in ecology & evolution*, 30(2), 78-87.
- Silva-Brandão, K. L., Silva, O. A. B. N. e., Brandão, M. M., Omoto, C., & Sperling, F. A. H. (2015). Genotyping-by-sequencing approach indicates geographic distance as the main factor affecting genetic structure and gene flow in Brazilian populations of *Grapholita molesta* (Lepidoptera, Tortricidae). *Evolutionary Applications*, 8, 476–485.
- Sod, S., Sharma, A., Sharma, N., & Kanwar, S. (2016). Carboxylesterases: sources, characterization and broader applications. *Insight Enzym Res*, 1(1).
- Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., & Neale, D. B. (2013). Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics and Genomes*, 9(4), 901–911. doi: 10.1007/s11295-013-0596-x
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. doi: 10.1093/bioinformatics/btu033
- Storey, K. B., & Storey, J. M. (2012). Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology*, 90(4), 456–475.
- Straw, N. A., Fielding, N. J., Tilbury, C., Williams, D. T., & Inward, D. (2015). Host plant selection and resource utilisation by Asian longhorn beetle *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in southern England. *Forestry*, 88, 84–95. <https://doi.org/10.1093/forestry/cpu037>
- Thomas, T., De, T. D., Sharma, P., Verma, S., Rohilla, S., Pandey, K. C., & Dixit, R. (2014). Structural and functional prediction analysis of mosquito Ninjurin protein: Implication in the innate immune responses in *Anopheles stephensi*. *Int. J. Mosq. Res*, 1, 60-65.
- Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29(12), 673–680. doi: 10.1016/j.tree.2014.10.004

- Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J. M., Pertoldi, C., Kristensen, T. N., Loeschke, V., Gregersen, V. R., & Bendixen, C. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, 103(4), 326–332. <https://doi.org/10.1038/hdy.2009.73>
- Torson, A. S., Zhang, M. L., Ong, K., Mohammad, L., Smith, A. J., Doucet, D., Roe, A. D., & Sinclair, B. J. (2021). Cold tolerance of laboratory-reared Asian longhorned beetles. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 257, 110957. <https://doi.org/10.1016/j.cbpa.2021.110957>
- Torson, A. S., Bowman, S., Doucet, D., Roe, A. D., & Sinclair, B. J. (2023). Molecular signatures of diapause in the Asian longhorned beetle: Gene expression. *Current Research in Insect Science*, 3, 100054.
- Turgeon, J. J., Gasman, B., Smith, M. T., Pedlar, J. H., Orr, M., Fournier, R. E., Doyle, J., Ric, J., & Scarr, T. (2022). Canada's response to invasion by Asian longhorned beetle (Coleoptera: Cerambycidae) in Ontario. *The Canadian Entomologist*, 154(s1), e1. <https://doi.org/10.4039/tce.2021.60>
- van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., ... Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5, 247–252.
- Vatanparast, M., & Park, Y. (2021). Comparative RNA-Seq Analyses of *Solenopsis japonica* (Hymenoptera: Formicidae) Reveal Gene in Response to Cold Stress. *Genes*, 12(10), 1610. doi: 10.3390/genes12101610
- Vatanparast, M., & Park, Y. (2022). Cold tolerance strategies of the fall armyworm, *Spodoptera frugiperda* (Smith) (Lepidoptera: Noctuidae). *Scientific Reports*, 12(1), 4129. doi: 10.1038/s41598-022-08174-4
- Vatanparast, M., Sajjadian, S. M., & Park, Y. (2022). Glycerol biosynthesis plays an essential role in mediating cold tolerance the red imported fire ant, *Solenopsis invicta*. *Archives of Insect Biochemistry and Physiology*, 109(2), e21861. doi: 10.1002/arch.21861
- Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, 67(12), 3403-3411.
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular ecology*, 23(23), 5649-5662.
- Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1), 323-354.
- Xu, Y., Li, Y., Wang, Q., Zheng, C., Zhao, D., Shi, F., ... Zong, S. (2021). Identification

of key genes associated with overwintering in *Anoplophora glabripennis* larva using gene co-expression network analysis. *Pest Management Science*, 77(2), 805–816. doi: 10.1002/ps.6082

Yocum, G. D., Rinehart, J. P., Horvath, D. P., Kemp, W. P., Bosch, J., Alroobi, R., & Salem, S. (2015). Key molecular processes of the diapause to post-diapause quiescence transition in the alfalfa leafcutting bee *Megachile rotundata* identified by comparative transcriptome analysis. *Physiological Entomology*, 40(2), 103-112.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. doi: 10.1111/2041-210x.12628

Zhao, D., Zheng, C., Shi, F., Xu, Y., Zong, S., & Tao, J. (2021). Expression analysis of genes related to cold tolerance in *Dendroctonus valens*. *PeerJ*, 9, e10864. doi: 10.7717/peerj.10864

Zhao, S.T., 2010. Vegetation in China, types and regionalization-Interpretation of Vegetation Map of the People's Republic of China. *Institute of Geographic Sciences and Natural Resources Research*, CAS.

Chapter 3

Tracking the Asian longhorned beetle's invasion history within North America with genomic evidence

3.1 Résumé

Les invasions biologiques par des nuisibles envahissants représentent une menace écologique et économique importante, comme c'est le cas avec le longicorne asiatique (ALB) qui cause des dommages considérables aux écosystèmes forestiers. Une gestion efficace des nuisibles repose sur une connaissance approfondie de la biologie de l'insecte et de son histoire invasive. Cette étude utilise la génomique pour combler certaines lacunes de connaissance et informer les systèmes de biosurveillance existants. Nous avons utilisé 2,768 polymorphismes mononucléotidiques à l'échelle du génome pour comparer les populations invasives d'ALB en Amérique du Nord. Cette variation génomique a été utilisée pour retracer les sources d'invasion et les patrons de propagation des populations étudiées, affinant ainsi notre compréhension de l'histoire invasive de l'ALB. Nous avons découvert que les populations d'ALB d'Amérique du Nord étaient distinctes, avec plusieurs introductions indépendantes depuis leur aire d'origine. Suite à leur introduction, toutes les populations invasives ont connu un goulot d'étranglement génétique et une expansion, certaines montrant également une propagation secondaire (e.g., Toronto). Les connaissances résultant de notre étude fournissent un support pour un outil de biosurveillance basé sur le génome qui peut être utilisé pour clarifier l'origine des individus interceptés, permettant aux agences de régulation de renforcer les mesures de biosécurité contre ce coléoptère envahissant.

3.2 Abstract

Biological invasions pose significant threats to ecological and economic stability, with invasive pests like the Asian longhorned beetle (ALB) causing substantial damage to forest ecosystems. Effective pest management relies on comprehensive knowledge of the insect's biology and invasion history. This study uses genomics to address these knowledge gaps and inform existing biosurveillance frameworks. We used 2,768 genome-wide single nucleotide polymorphisms to compare invasive ALB populations in North America, using genomic variation to trace their sources of invasion and spread patterns, thereby refining our understanding of ALB's invasion history. We found that the North American ALB populations were distinct, with multiple independent introductions from the native range. Following their introduction, all invasive populations experienced a genetic bottleneck and population expansion, with a few also showing secondary spread (e.g. Toronto). Knowledge from our study provides a foundation for a genome-based biosurveillance tool that can be used to clarify the origin of intercepted individuals, allowing regulatory agencies to strengthen biosecurity measures against this invasive beetle.

3.3 Introduction

Biological invasions pose a significant threat to the ecological and economic stability of our forests (Aukema et al., 2011; Pejchar & Mooney, 2009), and are considered one of the greatest threats to biodiversity (Schrader & Unger, 2003). Recent invasions in North America have highlighted the devastating long-term impacts of invasive pests once established in our forests (e.g. Emerald Ash Borer (*Agrilus planipennis*) (Flower et al., 2013); Hemlock Woolly Adelgid (*Adelges tsugae*) (Charles, Kathleen & Miquel, 2013); Chestnut blight (Anagnostakis, 1987)). Therefore, focus should be on preventing invasions and rapidly respond to new invasive pests to reduce the likelihood of their establishment (Epanchin-Niell & Liebhold, 2015). However, these proactive management approaches require detailed knowledge of the pest and its invasion pathway to establish effective controls that will reduce the likelihood of future introductions (Bilodeau et al., 2019; Roe et al., 2019).

Genomics, when integrated into a robust biosurveillance framework, can fill critical knowledge gaps and support proactive management of invasions which improves global biosecurity (Rees et al., 2022; Roe et al., 2019). Highly variable genomic markers, such as single nucleotide polymorphisms (SNPs) obtained through cost-effective genotyping-by-sequencing (Elshire et al., 2011) can provide detailed knowledge on invasive pest biology, including insights to the invasion history, pathways of introduction, and the regional sources of invasion (Hamelin & Roe, 2020). For example, genome-wide SNPs resolved invasion pathways for *Aedes aegypti* and showed that many individuals had signatures of insecticide resistance, highlighting the risk of relying solely on insecticides to prevent the spread of invasive insects (Smetana et al., 2019).

In another invasive mosquito, *Aedes albopictus*, genomics identified human-assisted transport and road corridors as important pathways for spread (Sherpa et al., 2019). Picq et al. (2018) showed that genome-wide markers could reliably trace the population sources for *Lymantria dispar asiatica* and *L. d. japonica*, which can be used to identify the sources of intercepted moths. These studies, among many others, highlight the breadth of knowledge that can be gained from genomic data on the invasion process and its ability to inform management responses to these threats.

The Asian longhorned beetle (Cerambycidae: *Anoplophora glabripennis* Motschulsky, ALB) is a polyphagous wood-boring beetle introduced to hardwood forests in North America and Europe (Meng et al., 2015; Blackburn et al., 2020). Recent work by Cui et al. (2022) described native ALB population variation using genome-wide SNPs and delineated distinct population structure among regional populations. The authors successfully assigned ALB individuals to regional groups with a limited number of SNPs, demonstrating the applicability of these genomic markers for tracing individuals to known source populations. Since its discovery in 1996 in Brooklyn, NY, there have been 16 distinct ALB infestations detected in North America (Figure 3.1). To date, the sources and relationships between these infestations are unknown. The source(s) of invasive ALB populations were previously explored by Carter et al. (2009, 2010) and Javal et al. (2019) using a combination of microsatellites and mitochondrial DNA. However, confidence in determining the sources for the invasive North American populations was hampered by complex population variation, limited information in the selected genetic markers, and low sample size (Carter et al. 2009, 2010; Javal et al. 2019).

Eradication has been used to control ALB populations in North America (Trotter & Hull-Sanders, 2015; Turgeon et al., 2010). Although this has been successful in several locations (e.g., Chicago, IL; Toronto, ON), control efforts are still on-going (Coyle et al., 2021). During these control efforts, new populations are frequently detected, and it is necessary to distinguish between new invasions and satellite populations caused by secondary spread. Tools such as genomic markers could provide a valuable insight into the history and spread of invasive ALB populations.

Here, using genomic data, we explored the invasion of ALB in North America. Genome-wide SNP markers and the native population structure described by Cui et al. (2022) provided us with the foundational knowledge required to trace the sources of invasion and assess the relationships between the North American ALB populations, allowing us to distinguish between new invasions and secondary spread. Knowledge gained from these genomic data can further help reduce uncertainty regarding spread dynamics and ensure the effective deployment of limited surveillance resources and guide the management responses to new detections (Melbourne & Hastings, 2009; Yemshanov et al., 2017; Browett et al., 2020; Darling, 2015).

3.4 Materials and Methods

3.4.1 Study Organism

Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky) (Coleoptera: Cerambycidae), native to China and the Korean Peninsula, was first detected in North America in 1996 and in Europe in 2001 (Lingafelter & Hoebeke, 2002; Poland, 1998). It is a polyphagous woodborer that has been recorded on >100 hardwood tree species with preferences for *Acer*, *Populus*, *Salix*, and *Ulmus* (Haack et al., 2006; MacLeod et al., 2002; Meng et al., 2015). Females lay individual eggs in the cambium and upon hatching the xylophagous larvae migrate into the heartwood (Meng et al., 2015) where it has been reported that a larva consumes 26 cm³ of wood on average (Tian et al., 1989). The average number of larvae per tree was found to be 23 based on a survey of 47 infested poplar trees (Tian et al., 1989). Adults typically emerge during the summer and feed on petioles, leaves and small branches (Yan, Qin & Xiao, 1992). These beetles were also thought to be relatively stationary and rarely dispersed in favourable habitat conditions (Zhou et al., 1984), with 98% of the individuals recaptured within 920 m of the release site in a mark-release study (Smith et al. 2004). However, this study also showed that a few individuals dispersed further than 2000 m, suggesting the potential of long-distance dispersal events. In China, ALB is considered a serious forest pest which causes nearly 12% of the total losses attributable to forest pests and diseases, costing an estimated \$1.5 billion annually (Hu et al., 2009).

3.4.2 Sampling and DNA preparation

To determine the source(s) of the invasive North American populations we sampled ALB throughout its native range (China and South Korea) to generate a reference collection and infer native population structure. As described by Cui et al. (2022), ALB samples came from four regions in China: 1) Northeast region (NE) which is bordered by the Greater Khingan Range to the west, 2) Northwest region (NW) bordered by the Helan Mountains to the east, 3) North Plain (divided into two regions, N1 & N2 based on the spatial structure), and 4) the South (S) bordered by the Huai River basin in the north. Cui et al. (2022) showed there was clear genetic structuring among these major regions in China. They also showed that although Shijiazhuang (SHI) is geographically located in the North Plain, it is genetically similar to the NW region. We also sampled from established invasive populations in North America over multiple years (Canada, United States of America) (Figure 3.1). In total, we sampled 266 ALB individuals from North America spanning the collection years 1999 to 2017 (Table 3.1, Table S3.1).

For the DNA preparation, we surface sterilized each sample using 95% ethanol. The tissue (a single leg or larval thoracic muscle) was flash frozen in liquid nitrogen, ground at 29 Hz for 1 min using a mixer mill (Retsch MM400, Germany). DNA was extracted from this homogenate

Table 3.1: Sampling location and genetic diversity of *A. glabripennis* in North America. N : number of individuals. Ho : observed heterozygosity. He : expected heterozygosity. π : nucleotide diversity. Standard error is presented for each value. The values of populations with sampling size less than four are not calculated. The indices for native populations are in Table S3.3.

Region	ID	Location	N	Collection year	Ho	He	π
IL	Chi	Chicago	18	1999	0.254 ± 0.005	0.235 ± 0.004	0.244 ± 0.004
MA	Bos	Boston	6	2010	0.204 ± 0.006	0.153 ± 0.004	0.174 ± 0.004
	Wor	Worcester	8	2008, 2009	0.223 ± 0.006	0.189 ± 0.004	0.203 ± 0.004
NJ	Jers	Jersey	2	2003	-	-	-
	Lin	Linden	4	2006	0.225 ± 0.006	0.165 ± 0.004	0.21 ± 0.005
NY	Far	Farmingdale	12	2013	0.27 ± 0.006	0.218 ± 0.004	0.227 ± 0.004
	Flush	Flushing	4	2000	0.211 ± 0.006	0.163 ± 0.004	0.193 ± 0.005
	Massap	Massapequa	21	2005-2007	0.248 ± 0.005	0.233 ± 0.004	0.241 ± 0.004
	NYC	NYC	19	2003-2005, 2008,2009	0.202 ± 0.005	0.227 ± 0.004	0.236 ± 0.004
	Qu	Queens	8	1999-2001	0.206 ± 0.005	0.211 ± 0.004	0.232 ± 0.004
OH	Beth	Bethel	16	2011-2015	0.264 ± 0.005	0.234 ± 0.004	0.242 ± 0.004
	Tate	Tate	1	2017	-	-	-
TOR	TOR1	Toronto	32	2004	0.233 ± 0.005	0.22 ± 0.004	0.223 ± 0.004
	TOR2	Toronto	5	2013	0.151 ± 0.005	0.132 ± 0.004	0.147 ± 0.004

using the DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany) with a RNAase treatment added to the procedure following the manufacturer's instructions. We measured DNA quality and quantity using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Wilmington, DE, USA).

3.4.3 Genotyping and bioinformatics processing

We prepared a total of $1\mu\text{g}$ DNA per sample for sequencing. Genotyping-by-Sequencing (GBS) libraries were prepared at the Ion Torrent platform (Institut de Biologie Intégrative et des Sys-

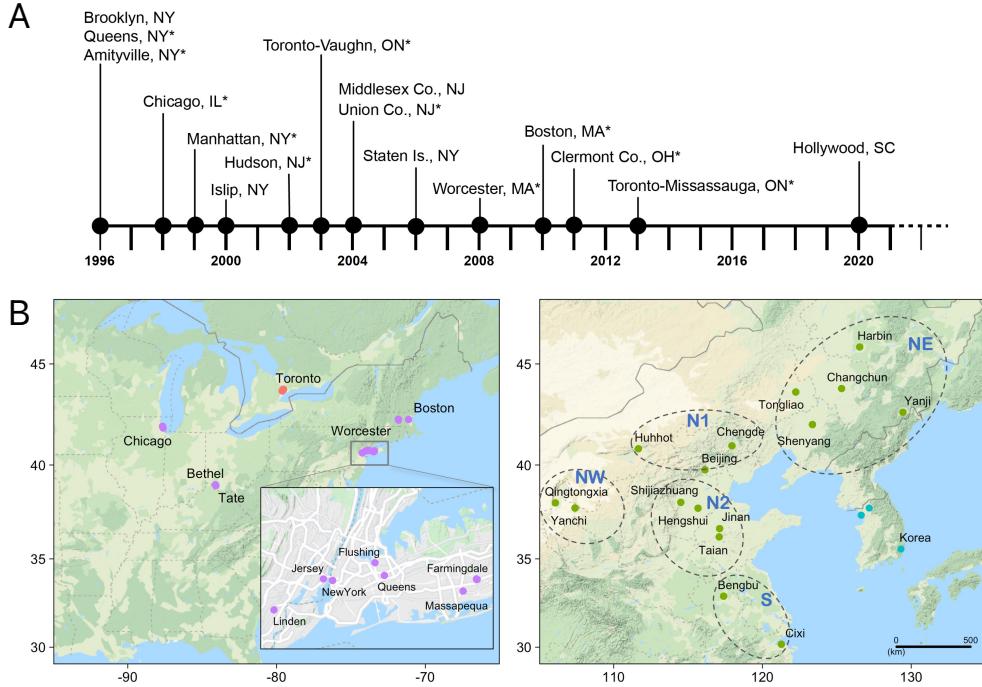


Figure 3.1: **A** Years of ALB occurrence in North America. Populations with an asterisk are the sample collection sites. **B** Sampling map. The left figure shows ALB sampling in the invasive range, Illinois (IL), Ohio (OH), Massachusetts (MA), New York (NY) and New Jersey (NJ) in the US (orchid) and Toronto, Canada (red). The inset shows the sampling in NY and NJ. The right figure shows ALB sampling in the native range, China (green) and Korea (cyan). The different regions within China are illustrated by dashed lines. N1: north region one. N2: north region two. NW: northwest region. NE: northeast region. S: south region.

tèmes (IBIS), Université Laval, Quebec City, Canada) following Abed et al., (2019) but modified by adding NsII to the PstI/MspI double digest to create a triple digest library. We adopted a double-size selection step to increase the number of fragments available for downstream analysis (Bayona-Vásquez et al., 2019; Peterson et al., 2012), and sequenced the resulting DNA fragments on the Ion Proton sequencer (Thermo Fisher Scientific, Waltham, MA, USA).

We used the Fast-GBS v1.0 pipeline to process raw sequencing reads (Torkamaneh et al., 2016; 2017). In this pipeline, SABRE v1.0 was used to demultiplex single-end, 150 bp barcoded reads (Joshi, 2011), and Cutadapt v2.1 (Martin, 2011) to remove the TruSeq Universal Adapter sequences. We aligned the trimmed reads to the ALB reference genome (GCA_000390285.1) (McKenna et al., 2016) using Burrows-Wheeler Aligner (BWA) v0.7.17 (Li, 2013). Samtools

v1.8 (Li et al., 2009) converted SAM files to BAM format for indexing, and we used PLATYPUS v0.8.1.1 to call SNP variants (Rimmer et al., 2014). We filtered the SNP variants using VCFtools v0.1.16 (Danecek et al., 2011) and PLINK v2.0 (Chang et al., 2015). We applied basic filters to retain biallelic SNPs and variants with a PASS flag, retaining loci with <50% missing data and individual samples with <20% missing data. Lastly, we retained loci with <10% missing data per site, a read depth >5, and minor allele frequency (MAF) > 0.05. We then pruned SNPs in a sliding window of 50 SNPs (advanced by 5 SNPs each time) with $r^2 < 0.4$. Finally, we removed samples based on relatedness using the KING method (Manichaikul et al., 2010) integrated in PLINK software with a cutoff value of 0.25 to remove full siblings.

3.4.4 Genetic diversity and Population Structure

We calculated the following genetic diversity indices in STACKs v1.48 (Catchen et al., 2013) for each native and invasive population: observed heterozygosity (H_o), expected heterozygosity (H_e), nucleotide diversity (π) and inbreeding coefficient (F_{IS}).

We performed principal component analyses (PCA) in adegenet v2.1.2 (Jombart, 2008) to characterize genetic differentiation among all samples. We applied the maximum likelihood estimation method Admixture v1.3.0 (Alexander et al., 2009) to estimate individual ancestry membership. This approach used a cross-validation method to identify the optimal K. We ran the models with ascending number of ancestral populations from 2 to 20 with a default 5-fold cross-validation and selected the optimal K based on the least cross-validation. We also calculated pairwise F_{ST} between populations in STACKs (Catchen et al., 2013).

3.4.5 Population assignment

To determine the evolutionary relationships among samples we computed a maximum likelihood (ML) phylogeny using 1,000 bootstrap replicates under the GTRGAMMA model in RAxML v8.2.9 (Stamatakis, 2014).

We applied a discriminant analysis of principal components (DAPC) in adegenet v2.1.2 (Jombart, 2008; Jombart et al., 2010) to assign invasive ALB to populations in the native range. We ran the "dapc" function for all native ALB samples with five principal components and retained three discriminate functions in the discriminant analysis. Based on the DAPC results obtained from the native dataset, we then performed the "predict.dapc" function to predict group memberships results for all the invasive ALB samples. We visualized the assignment results as a contingency table.

3.4.6 Invasion history and recent gene flow

To reconstruct the invasion history of the North American ALB, we compared different invasion scenarios using an approximate bayesian computation (ABC) method implemented in

DIYABC v2.1.0 (Cornuet et al., 2014). This approach simulates datasets for *a priori* invasion history models and compares these simulated datasets with the observed dataset. We conducted the ABC analysis using a step-by-step procedure described by Sherpa et al. (2019). We guided our scenario selection using previous molecular genetic studies on ALB (Javal et al. 2019, Cui et al. 2020) and our current results. The first three steps sequentially defined the divergence history of the native populations, while the next five steps identified the optimal invasion scenario for each invasive population (Figure S3.1, see Supplementary Note for further details of scenario design and definition of prior parameters). We only considered invasive populations with clear population structure; hence we did not examine populations from New York or New Jersey. We then estimated posterior parameters for a native *plus* North American evolutionary scenario which combined all populations. We simulated 20,000 data sets for each scenario. All summary statistics for SNPs in DIYABC were selected. We checked each model (scenarios and/or associated priors) using Principal Component Analysis (PCA). In this process, we projected both the observed dataset and the datasets simulated from the priors onto a PCA space. The PCA was performed using the genetic summary statistics as the components of the feature vector. We anticipated that the observed data would overlap with the simulated datasets, which would support the validity of our models. For model selection and estimated parameters, we used 1,000 trees simulated in DIYABC Random Forest v1.1.27 (Collin et al., 2021), an extended version of DIYABC which uses decision trees to facilitate model selection (Pudlo et al., 2016). This method then generates a classification vote for each scenario and an estimation of posterior probability for the best scenario.

We measured recent migration rates (migrants per generation) using BA3-SNPs v1.1 (Wilson & Rannala, 2003), a modified method based on BayesAss v3.04 (Mussmann et al., 2019) which estimates recent gene flow between populations using Bayesian Markov Chain Monte Carlo resampling. We only used populations that contained more than 15 individuals because estimates are more accurate with larger sample sizes since BayesAss assumes low migration rates which is hard to estimate with smaller sample sizes (Meirmans, 2014). First, we ran the analysis with the default value of each mixing parameter (i.e., 0.1). Based on the resulting acceptance rates, we adjusted the mixing parameters values until acceptance rates fell within a suggested range of 0.2 ~0.6 (Wilson & Rannala, 2003). Then, with the adjusted mixing parameters we performed 10 longer, independent runs for allele frequencies of 0.5 and inbreeding coefficient of 0.01, using 20 million iterations, a burn-in of five million, sampled every 1000 iterations, and produced trace files with the -t flag. We calculated the Bayesian deviance as suggested by Meirmans (2014) to select the best run. Run 5 (Figure S3.2) had the least deviance value and was selected for downstream parameter estimation of migration rates (the number of migrants per generation) between ALB populations.

3.5 Results

3.5.1 Genotyping results for ALB populations

We generated ~ 447 million reads for 651 ALB samples using GBS with a triple digest library and aligned ~ 359 million reads to the ALB reference genome. We obtained 969,515 SNP variants through the fastGBS pipeline. After variant filtering and quality control, we retained 2,768 SNPs and 490 individuals, including 331 reference individuals from China and 3 putative native samples from Korea (Cui et al., 2022), plus 156 invasive samples from North America. Numbers of individuals and SNPs before and after each filtering step are shown in Table S3.2.

3.5.2 Genetic diversity and population structure

H_o (observed) and H_e (expected) heterozygosity values as well as π (nucleotide diversity) were higher across ALB's native range (Table S3.3) compared to the invasive North American ALB populations (Table 3.1), with significantly highest inbreeding (F_{IS}) and lowest H_o in the invasive ALB populations ($p \leq 0.0001$, Figure 3.2).

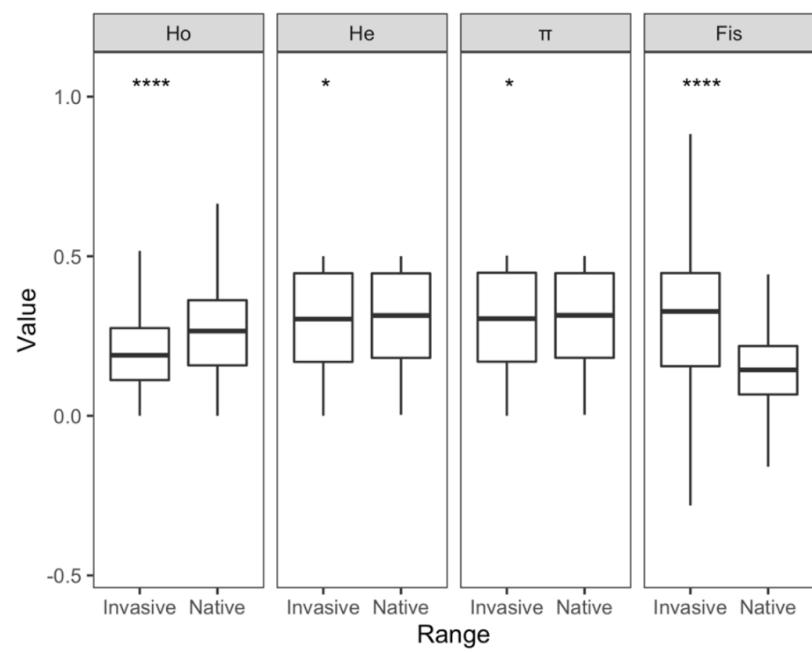


Figure 3.2: Genetic diversity of ALB in native and invasive ranges. H_o : observed heterozygosity. H_e : expected heterozygosity. π : nucleotide diversity. F_{IS} : inbreeding coefficient. Significance level is indicated on each panel. ‘*’: $p \leq 0.05$. ‘****’: $p \leq 0.0001$.

Notably, TOR2 had the lowest genetic diversity values relative to the other invasive populations. The genetic distance between populations is represented by pairwise F_{ST} values (Figure 3.3A). The average F_{ST} across the invasive populations was 0.1547 ± 0.0075 , whereas across

native reference populations it was only 0.0598 ± 0.0021 , with an average F_{ST} between native and invasive populations of 0.096 ± 0.0021 (Table S3.4). As previously shown in Cui et al. (2022), our native reference populations formed distinct regional clusters (Figure S3.3), with SHI grouping with NW populations (QI, YC) rather than with geographically proximate populations in the North Plain. When we included the invasive populations (Figures 3.3B, S3.4), Toronto samples formed a single distinct cluster, USA samples formed groups nested within Chinese regions, and no North American samples were associated with native Korean populations (Figures S3.5, S3.6).

In our admixture analysis we selected K=14 as the optimal value of K for the combined data set of native and invasive populations (see Figure S3.7 for K=3, K=4). Our admixture results (Figure 3.3C) and the PCA clustering (Figure S3.6) were largely congruent for the invasive populations, with a few exceptions. Three of our invasive populations (TOR, OH, IL) formed distinct and uniform admixture plots and PCA clusters across sites and sampling year (Figures 3.3, S3.6). Our MA population showed a distinct, uniform admixture plot; however we observed temporal variation among individuals in the PCA plots (Figure S3.6C). For example, we found that all 2008 individuals collected from Worcester, MA, formed a tight group nested within the native South cluster, along with two 2010 individuals from Boston, MA. The remaining samples (Worcester 2009 and Boston 2010) formed a loose group with no clear source association (Figure S3.6C). We also observed genetic complexity in our NY and NJ collections, compared to the other invasive populations. The NY infestation was the oldest and most extensive, with samples spanning 14 years (1999-2013), and we observed spatial and temporal genetic variation among these collection locations and time points. For example, the early samples from New York City (1999-2009) were admixed and formed a large mixed cluster in the PCA (Figures 3.3, S3.6B). Farmingdale (2013), however, formed a distinct group, separate from the remaining New York City samples in both the admixture plot and PCA (Figures 3.3, S3.6B). Massapequa (2005-2007), Queens (1999-2001), Flushing (2004) and New Jersey (2003, 2006) showed variable levels of admixture and clustering within the PCA analyses (Figures 3.3, S3.6B,C).

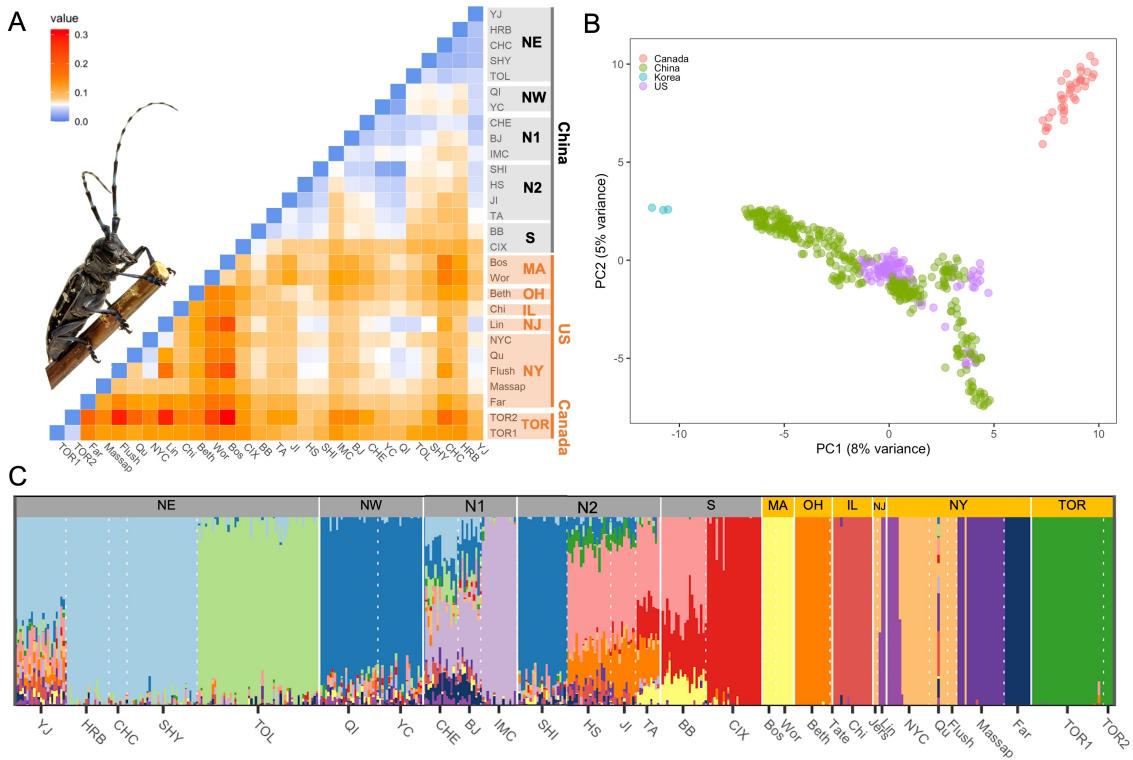


Figure 3.3: Population structure of global ALBs. **A** Principal components analysis of global ALB. **B** Admixture bar plots for the proportion of genetic membership ancestry ($K=14$, the optimal number according to the least cross validation, $K=3$ and $K=4$ – see Figure S3.7). Individuals are shown as vertical bars colored in proportion to their estimated ancestry within each cluster. **C** Pairwise F_{ST} between populations. Populations with sampling size less than four are not shown. The picture of *A. glabripennis* adult is kindly provided by Dr. Brent Sinclair.

3.5.3 Population assignment

To further refine our population assignments, we assigned invasive individuals to native reference population clusters using DAPC discriminant functions derived from a reference DAPC model (Figures 3.4A, S3.8). We summarized these individual assignments in a contingency table (Figure 3.4B), with assignment results for each individual and their posterior membership probabilities shown in Table S3.5. Most invasive individuals were assigned to the N2 cluster. This included all individuals from TOR, OH, and IL, as well as some individuals from MA, NJ, and NY. The remaining NY individuals were assigned to N1, including all individuals from Farmingdale, NY. We observed a similar pattern in NJ, with individuals assigned to both N1 and N2 clusters, but with the addition of one individual assigned to the NE cluster, the only North American individual associated with this region. The MA infestation also had individuals assigning to N2, but six individuals were assigned to S. This is consistent with the

PCA results reported earlier (Figure S3.4C). Notably, no individuals were assigned to the NW cluster.

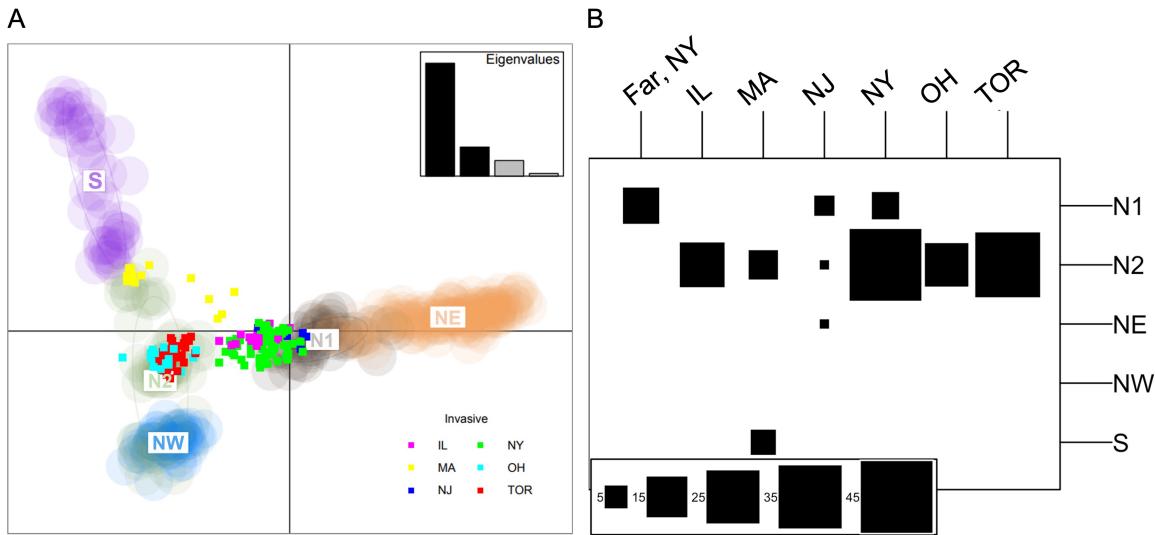


Figure 3.4: Population assignment in DAPC. **A** Scatterplot in two dimensions of discriminant functions showing clusters of native ALBs (light dots) with adding the invasive individuals (solid squares). **B** DAPC population assignment contingency table. The table indicates the distribution of the number of invasive individuals (columns) assigned to each cluster across native populations (rows). There is one sample from NJ assigned to N2 and NE, respectively. Farmingdale (Far) was indicated with a separate population from NY in the table.

We constructed an ML tree using our SNP data set (2,768 SNPs) to further clarify the relationships between invasive individuals and native populations (Figure 3.5). A number of invasive populations were nested within native lineages with high support. Individuals from TOR and OH were well supported within the N2 lineage and MA individuals nested within the S clade. Individuals from Farmingdale nested within the N1 clade. The remaining individuals from NY, IL, and NJ formed a single clade with low support values, and there was no clear, well-supported relationship of these samples with a native lineage.

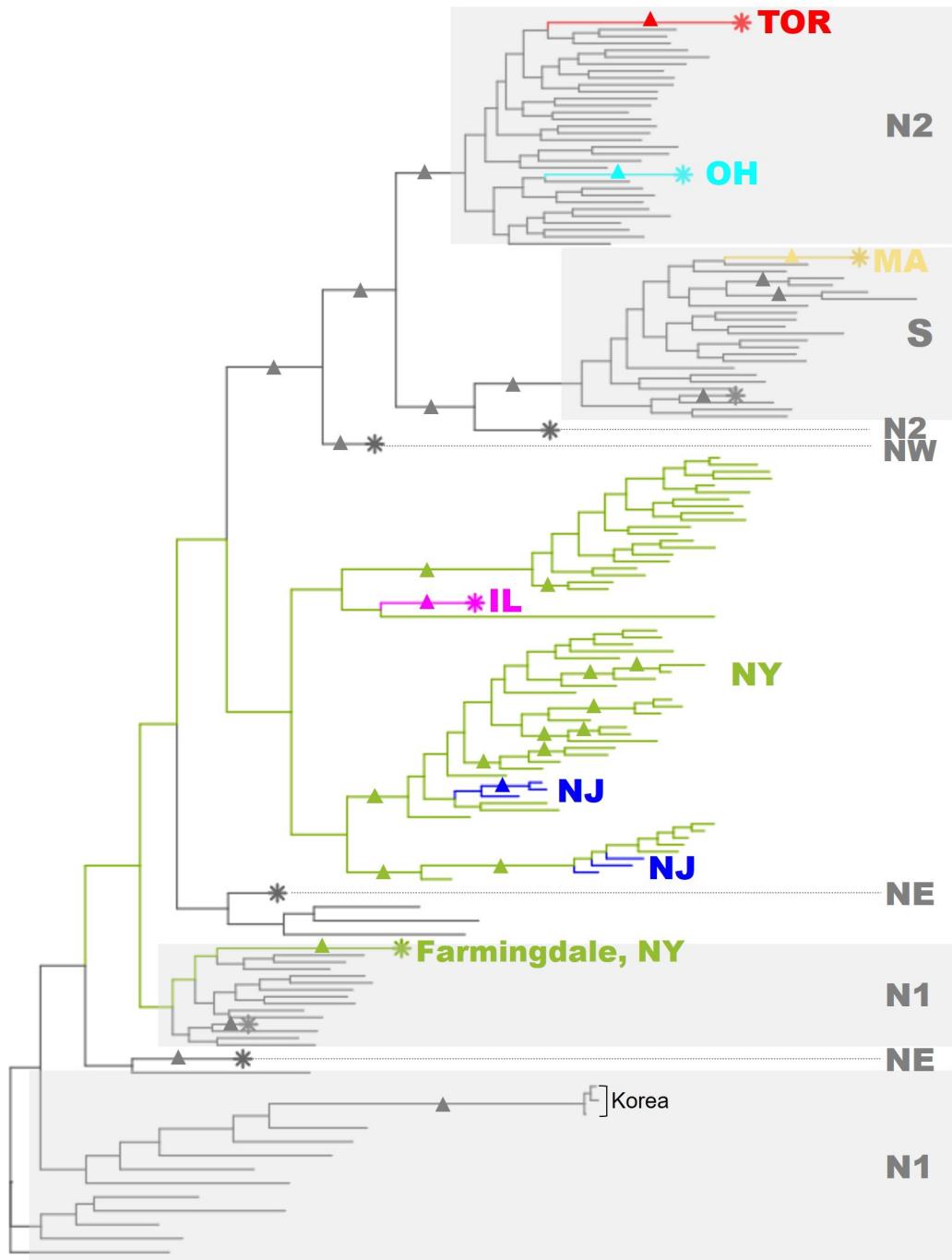


Figure 3.5: Maximum likelihood phylogenetic tree (unrooted). Native ALBs are color-coded in grey, whereas invasive populations are colored distinctly. Branches with a '*' symbol are collapsed clades. Branches with bootstrap value $> 80\%$ are indicated with '▲' symbol. For the native regions in China, the collapsed clade N2 (top) includes populations HA and JI; S includes BB and CIX; N2 (bottom) includes TA; NW includes YC, QI and SHI; NE (top) includes SHY, HRB, CHC and YJ; N1 (top) includes BJ, IMC and CHE; NE (bottom) includes TOL; N1 (bottom) includes CHE.

3.5.4 Invasion history and recent gene flow

We used a DIYABC approach to examine the invasion history and population demographics of five invasive populations (MA, OH, IL, TOR, Far). Using an eight-step approach to DIYABC model selection (see Supplementary Note), we selected a single optimal population history model for each population (Figure S3.1) and a combined native + North American invasion scenario (Figure 3.6). The observed data set overlapped with the simulated data sets in the PCA (Figure S3.9). We selected scenarios in each step which had the highest classification votes and the posterior probability for the selected native + NA invasion models were above 0.90 (steps 4-8).

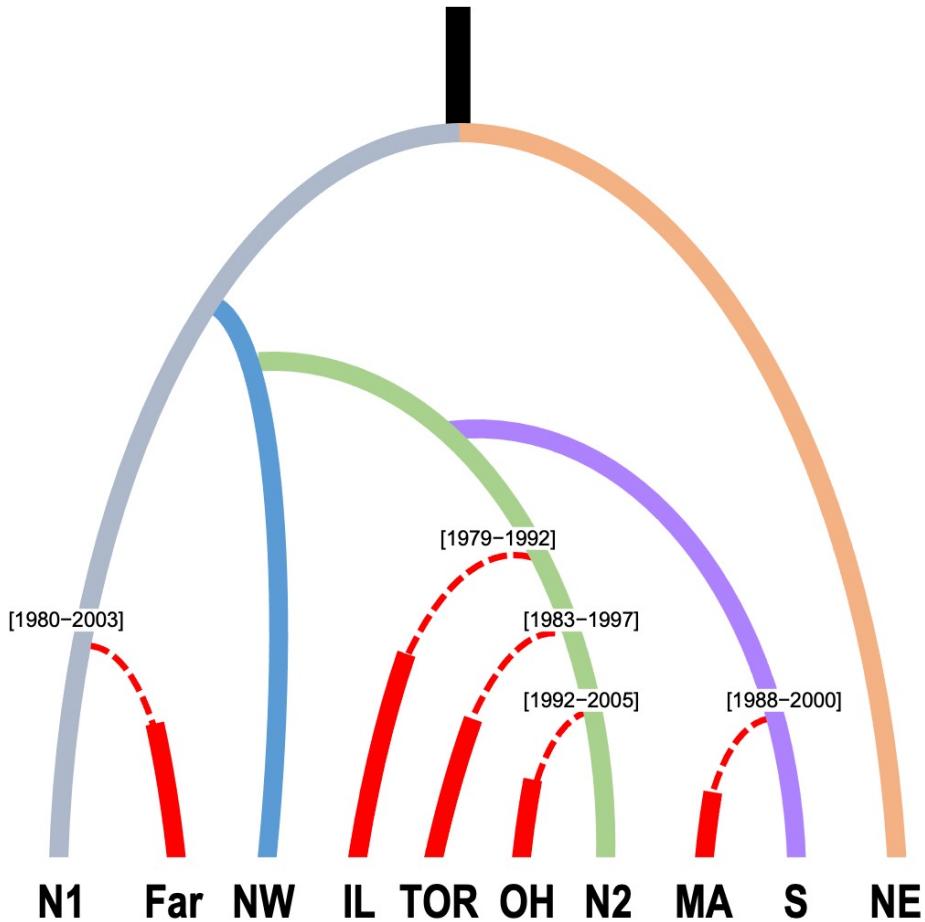


Figure 3.6: Invasion history of North American ALB populations. N1: North region 1 (grey); N2: North region 2 (green); NW: Northwest (skyblue); NE: Northeast (orange); S: South (purple). IL: Illinois; TOR: Toronto; OH: Ohio; MA: Massachusetts; Far: Farmingdale. Invasive populations are colored in red, with thin dashed red lines indicating bottleneck events, followed by population expansion (solid red line). The unsampled ancestry population is colored in black (top). The mean value and 95% confidence interval (CI, in exact years) is indicated for the introduction date of the invasive populations (time is not to scale).

The average effective population size for each contemporary native population at the time of sampling varied from 3,672 (S, 95% CI: 740-8,018) to 6,847 (N1, 95% CI: 3,054-9,659) (Table S3.6). The average estimation of the divergence time for the native populations varied from 324 (95% CI: 86-692) to 1,549 (95% CI: 428-2,969) years (Table S3.6). Notably, in the DIYABC model, the units of time are either years or generations, with the assumption of one year as the life cycle for ALB.

The effective population size for invasive populations ranged from 4,900 (OH, 95% CI: 403-9,468) to 5,089 (Farmingdale, 95% CI: 384-9,491), and the variability in the estimates of the invasive populations was higher compared to native ones. This indicates a greater spread or dispersion in the estimated population sizes for invasive species. The average introduction time of invasive populations spans from 13 to 24 years prior to the time they were sampled, which falls between the years 1985 and 1998 (95% confidence interval: 1979-2005) (Table S3.6). The mean predicted founding population for the invasive populations ranged from 34 (MA, 95% CI: 10-183) to 87 individuals (IL, 95% CI: 13-193) (Table S3.6). Each introduction experienced a bottleneck which lasted between 11-19 years (i.e. mean bottleneck duration) but could have lasted as little as one year (TOR) or as long as 33 years (Farmingdale). Between the time of introduction and the sampling date, the invasive populations expanded as the populations grew and spread.

We measured recent migration between populations in both the native and invasive ranges. Notably, BayesAss estimates contemporary migration, and the resulting values can be regarded as indicators of recent colonization. We selected Run 5 (Figure S3.2) based on its Bayesian deviance to generate point estimates and calculated the number of contemporary migrants per generation (Nm) within and among native and invasive populations (Figure 3.7). Within the invasive range, we observed more frequent gene flow between populations (Worcester to Boston, 0.0339 Nm ; New York to Linden, 0.0206 Nm ; the 1st infestation of Toronto to the 2nd, 0.0196 Nm). Overall, we saw low contemporary migration from the native to invasive range and no migration from the invasive to the native range.

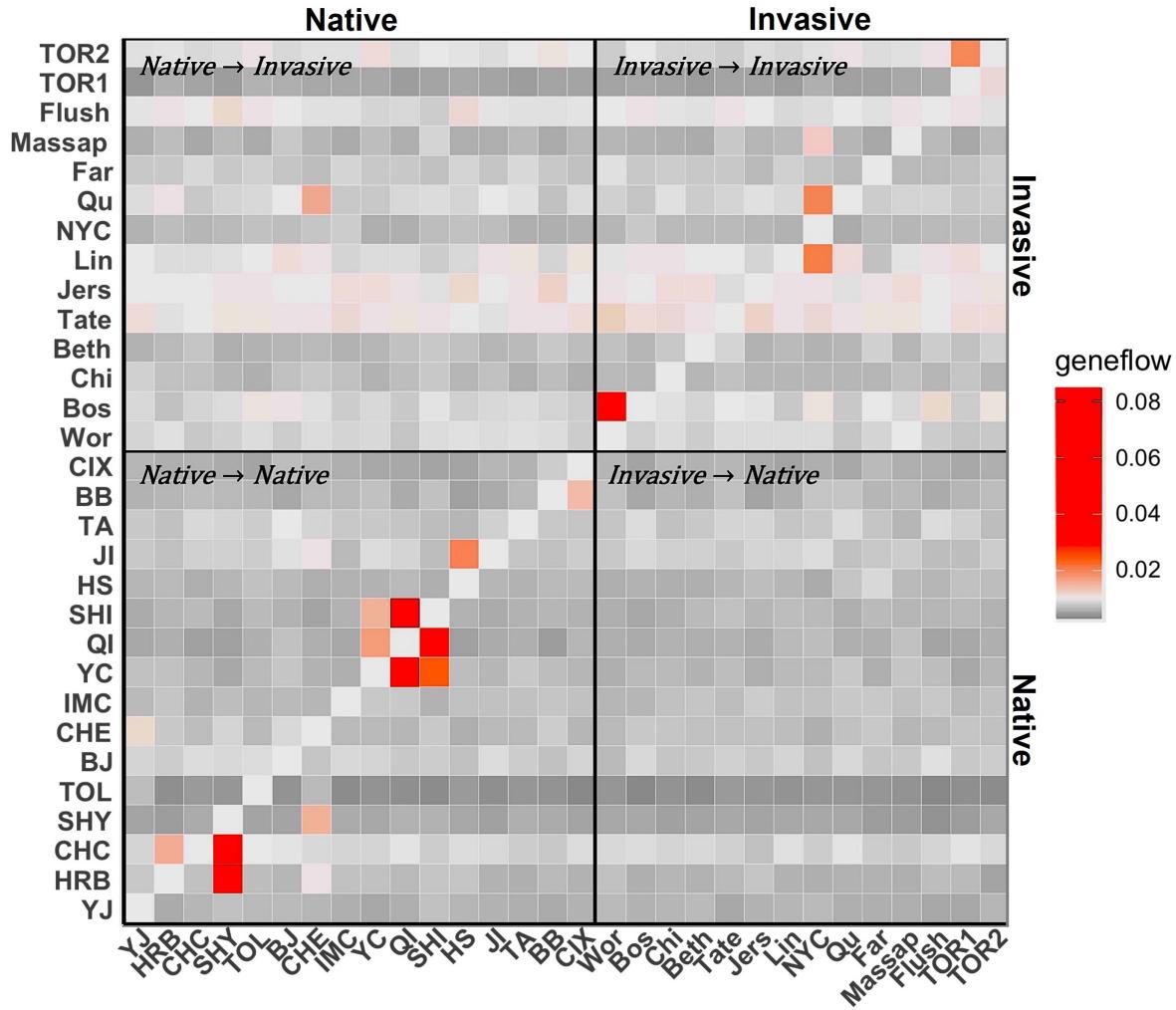


Figure 3.7: Gene flow between and within ranges. The migration direction reads from populations in vertical to populations in horizontal order. Grey to orange color represents the migration rates from low to high.

3.6 Discussion

The Asian longhorned beetle is an invasive, highly destructive insect pest that poses a significant threat to hardwood forests throughout North America and Europe. Genome-wide markers obtained from reduced representation libraries provide insight to the population variation among invasive ALB populations in North America. We showed that North American populations were structured, with multiple independent introductions from four sources in the native range. We also showed that secondary spread from primary infestations created new satellite populations, which were likely human-assisted given the relatively poor dispersal abilities of ALB. These populations experienced genetic bottlenecks followed by population expansion, demonstrating their resilience to founder effects. Collectively, these data fill im-

portant knowledge gaps about the invasion of ALB in North America and can help inform future biosurveillance approaches and eradication efforts for other active infestations.

3.6.1 Genetic structure and bottlenecks in North American ALB invasion

Based on 2,768 SNPs, we showed that introduced ALB populations were highly structured, with distinct genomic signatures, reduced heterozygosity and nucleotide diversity (Figure 3.2; Table 3.1). We detected significant inbreeding within these populations, as we would expect among populations with low numbers of founders and recent genetic bottlenecks. This was supported by our demographic modeling where we predicted that the invasive populations were founded with relatively few individuals (<100 individuals, Table S3.6). Following establishment, all ALB populations experienced an increase in their effective population sizes, suggesting population growth in the new habitat. However, results highly rely on the priors chosen. Our evidence suggests that the invasive populations of ALB went through bottleneck events, lasting on average 11 to 20 years. It is known that invasive species face inbreeding pressure and reduced genetic diversity due to bottleneck events, which can limit their ability to adapt to new environments. However, despite these challenges, studies have revealed that invasive species often exhibit a high level of fitness and are exceptionally successful at establishing and spreading in new environments (Mergeay, Verschuren & Meester, 2006; Gatto-Almeida et al., 2022). This phenomenon represents a genetic paradox associated with invasive species (Estoup et al., 2016; Schrieber & Lachmuth, 2017). The occurrence of multiple introductions has been one key explanation of invasion success despite genetic bottleneck in the founder (Kolbe et al., 2004), which are common among invasive insects (Garnas et al., 2016), with recent examples in termites (Blumenfeld et al., 2021; Eyer et al., 2021), *Spodoptera frugiperda* (Tay et al., 2022), *Trichocorixa verticalis* (Ortego et al., 2021), *Lycorma delicatula* (Kim et al., 2021), *Roeseliana roeselii* (Kaňuch et al., 2022) and *Bemisia argentifolii* (Wongnikong et al., 2021). Repeated introductions can potentially mitigate inbreeding and increase genetic diversity, however, most of our studied invasive populations appear to have resulted from independent introduction events. Insects are renowned for their rapid evolutionary rates (Loxdale, 2010), which can contribute to their adaptability and resilience in the face of challenges from inbreeding pressure. Furthermore, pre-adaptation, where the invasive populations already possess traits that are favorable in the invaded region, is likely to contribute to invasion success (Estoup et al., 2016).

3.6.2 Invasion source of North American ALBs

Tracing the source of an invasion provides useful knowledge on the history and the frequency of introductions from specific source regions, i.e., indicating propagule pressure (Simberloff, 2009; Hamelin & Roe, 2020). Successful identification of an invasion source is dependent on levels of population structure in the native range (Hamelin & Roe, 2020; Roe et al., 2019) and the ability for genetic markers to resolve this structure. Cui et al. (2022) effectively resolved

distinct regional structure among ALB populations in China and Korea using genome-wide SNP markers, providing an opportunity to reconstruct the invasion history of North American populations. Since 1996, 16 ALB infestations have been detected in North America. Human assisted dispersal is likely key to the movement of ALB (Haack et al., 2006; Turgeon et al., 2022) and many other invasive insects (Ladin et al., 2023; Short et al., 2020). The primary pathway for ALB introduction is solid wood packaging material, such as pallets, dunnage, and spools, as for other wood-boring invasive insects (Greenwood et al., 2023). Port inspections continue to discover infested wood packaging material (Krishnankutty et al., 2020), despite phytosanitary measures established to curb the spread of wood-boring invasives (Aukema et al., 2010; Sjöman et al., 2014). As the invasions of ALB persist, the genetic structure and invasion history of these introductions were unresolved (Javal et al 2019; Carter et al. 2010). Here, our genomic evidence shows that the invasive populations had unique genomic signatures (Figure 3.3) where genetic distance between invasive populations was more pronounced than both between native and invasive populations and within the native range (Figure 3.3A), which represented independent introductions from the native range (Figures 3.4, 3.5). A similar situation was found in the Ciosi et al. (2008) study of the invasive western corn rootworm *Diabrotica virgifera*, which found that in the invasive range, each individual population became less diverse after invasion, but greater genetic differences persisted among the different invasive populations. The large genetic variation can be due to different source(s), random genetic drift, as well as differences in selection pressure. Our results are consistent with Javal et al. (2019) who also showed that multiple native regions acted as sources for European ALB populations, however our genome-wide markers provided greater resolution on the invasive population structure.

Previously, Cui et al. (2022) delineated six native ALB populations using genome-wide SNP markers. Earlier approaches were unable to achieve the same resolution due to marker variability and low sample size (Carter et al. 2009, 2010; Javal et al. 2019). Using this defined genomic map, we were able to determine that four native ALB populations acted as sources for the North American populations (N1, N2, NE, S; Figures 3.4-3.6 phylogenetic reconstruction, DAPC assignment, and DIY-ABC analyses). Samples from NY infestations, including Farmingdale, were the only populations derived from the N1 region (Figure 3.6, S3.8). The N2 region (Figure 3.1; represented by four sites south of Beijing) was the dominant source of our invasive populations (Figures 3.4, 3.5). Native populations of ALB have expanded rapidly since the 1980s (Yan, 1985) and which was linked to afforestation efforts in northern China associated with the Three North Shelterbelt Forest program (Luo et al., 2000) during the 1980s-1990s. Population outbreaks have expanded beyond this region and ALB is now considered a widespread pest throughout temperate China (Luo et al., 2000). The N2 source region (outside of the Three North Shelterbelt Forest) has been the primary area of population outbreaks (Yan, 1985) and continues to grow (Luo et al., 2000), contributing to the subsequent discovery of additional invasions within North America. The presence of high numbers

of a species in its native range elevates the risk of its importation to other areas, leading to increased propagule pressure (Lockwood, 2005). Colautti et al. (2006) have demonstrated a positive correlation between propagule pressure and the successful establishment of invasive populations and suggested by factoring in propagule pressure, we are better equipped to develop and implement more effective strategies to mitigate the unwanted effects of nonnative invasions. While the risk of future ALB introductions remains high given our result and increasing global connectivity and trade (Garnas et al., 2012), our data can help identify high risk source regions and guide targeted port inspections and surveillance protocols; however, it is still challenging to develop a biosurveillance approach that targets a particular source region, even with the knowledge provided here, given that multiple native sources have contributed to the North America invasion.

Despite multiple introductions, there is limited evidence of admixture among the invasive North American populations (Figures 3.3, 3.4). Admixture can increase genetic diversity and create new allele combinations that could affect the evolutionary trajectory of an invasive population (Dlugosch & Parker, 2008; Rius & Darling, 2014). A prerequisite for admixture to occur is contact between different genetic lineages, and although we have not detected this for most of our populations, we have detected complex genetic variation within New York State (NY), such as New York, Queens, Flushing, and Massapequa (Figure 3.3C). Notably, nucleotide diversity in these admixed NY populations was not significantly higher than in populations from single origins, suggesting only a moderate admixture level. As also shown in Figure 3.3C, not all individuals from NJ and NY demonstrate a pattern of admixture. Although our DAPC assignment indicated that among all our sampled native regions, these NY populations appear to originate from source region N2, the source(s) contributing to the admixture remain unclear as they form a single branch on the phylogenetic tree. Therefore, the origins of the majority of NY populations remain largely uncertain.

3.6.3 Spread of ALB

Following introduction, establishment and expansion of an invasive species is driven by secondary spread into neighbouring habitats (Blackburn et al., 2011). Limiting the secondary spread is often a critical component of invasive species control and management (Garnas et al., 2016; Pyšek & Richardson, 2010). Unlike many invasive species, eradication of ALB is possible and due primarily to its limited dispersal capacity (Smith et al., 2001, 2004) and its relatively low rate of development and low reproductive rate (Keena, 2002) in the outbreak area (Turgeon et al., 2022). Human-assisted movement, however, can undermine these control activities and cause regulated pests to breach regulated quarantine zones (Hulme, 2009; Rassati et al., 2018). Thus, eliminating secondary spread is crucial for effective control of invasive pests. Distinguishing between new introductions and satellite populations due to secondary spread helps regulatory agencies evaluate the success of eradication efforts, establish

appropriate quarantine zones, and identify high risk pathways of movement (Garnas et al., 2016). Here, we quantified gene flow between individual infestations and verified a number of existing hypotheses around the history of spread in North America. For instance, Brooklyn is considered the original infestation in the New York area where it was first detected outside its native range in 1996 (Sawyer et al., 2006). Central Long Island includes several infestations which were predicted to be a result of secondary spread from the original Brooklyn population (Haack et al., 1997). Our genetic results support secondary spread, namely gene flow from NYC to Massapequa (Long Island region) (Figure 3.7); Boston was a satellite population of Worcester, a population well outside the dispersal range previously recorded in ALB (Smith et al., 2001, 2004). Worcester is the largest infestation in North America (Santos & Cole, 2012; Danko et al., 2016), and it is plausible that the spread was via human activity, particularly the movement of infested wood; further, the Toronto infestation in 2013 has been known as a satellite from the 2003 infestation (Turgeon et al., 2015). Our results also confirmed this finding.

Considering that secondary spread is a relatively common occurrence, it is plausible that these populations could serve as sources for other invasions, potentially demonstrating the so-called bridgehead effect (Lombaert et al., 2010) where a successful invasive lineage acts as a source for new invasive populations, a phenomenon frequently reported in other invasive groups (Garnas et al., 2016; Lombaert et al., 2010; Rius & Darling, 2014). Given the distinct genomic signatures of each invasive population, we need to be aware that secondary dispersal between infestations could potentially lead to more mixing within the invasive range and alter the evolutionary potential and inherent risk posed by these populations. Admixed invasive populations have been identified as important drivers of global invasions and are frequently detected in bridgehead invasion scenarios (Lombaert et al., 2010). A potential bridgehead scenario from invasive USA populations (Illinois) was already predicted for ALB introduction at Gien, France (Javal et al. 2019). Our genomic data provide a baseline to assess whether genetic structure and diversity in invasive ALB populations is changing over time. These data could then be used to detect if/when admixture occurred.

Moreover, as we now know, it is extremely difficult to predict spatial spread of invasive species, because it requires accounting for lots of uncertainties (Melbourne & Hastings, 2009). Notably, based on our result, long distance satellites of ALB (tens of kilometers) appears to be not uncommon (within New York City; NY to NJ; Worcester to Boston; within Toronto), which is mainly due to human activity as the distance between source and satellite are beyond the natural dispersal distance capacity ($<=2.6$ km) (Smith et al., 2004). Therefore, it is important to maintain long distance survey sites outside of quarantine zones, considering the distance of successful satellites relative to the quarantine zone, as incorporated in the Yemshanov (2017) model for designing optimal invasive species surveys. As we gather more information on the invasion processes of various species, it is possible to reduce uncertainty and achieve more

accurate error estimation in ecological predictions.

3.7 Conclusion

The Asian longhorned beetle is a high-risk invasive species that continues to threaten temperate forests in North America, Europe and other parts of Asia. There have been over 30 outbreaks of ALB in North America and Europe (Haack et al., 2010), with the most recent outbreak discovered in South Carolina in 2020 (Coyle et al., 2021) and also in Japan (Akita et al., 2021). While eradication is possible, active infestations are still being controlled and live ALB are still intercepted along trade pathways (Wu et al., 2017; Turgeon et al., 2022). Although high-risk wood packaging material (WPM) is targeted for inspection, it has not led to a significant reduction in the infestation rate of such WPM and live insects are occasionally found in ISPM15-marked WPM (Haack et al., 2014). As shown here, knowledge derived from genomic data can elucidate the invasion pathways for ALB, providing information that can further guide surveillance and management efforts. Our SNP markers can be translated into a target-enriched screening tool (Altmüller et al., 2014; Diepenbroek et al., 2020) that will be able to rapidly reconstruct the invasion history of new or existing ALB populations or an intercepted individual. With such tool in hand, we can also re-examine global invasion scenarios for ALB populations (e.g., Javal et al., 2009; Lee et al., 2020) to refine our understanding of the invasion history of this important pest. By harnessing the power of genomic data and our refined understanding of ALB invasions, we pave the way towards innovative genomic biosurveillance and informed management strategies, thereby managing the threat posed by invasive insects more effectively.

3.8 Supplementary materials

The supplementary tables can be found in Dryad: <https://datadryad.org/stash/share/fCbWSHakbd2HBPpTFYfbVTBo3UNjnxKGNofJ1wkukm0>.

3.8.1 Supplementary Note

Scenario design for DIYABC

To investigate the invasion history of North American ALB populations, we followed an 8-step analysis (Figure S3.1). We limited our analyses to five invasive populations (MA, OH, IL, Far, TOR) due to the limited sample size in NJ and evidence of admixture in NY.

Here, we first tested the relationships among three native populations N1, N2 and NE (step 1). The optimal scenario for each step is selected based on classification votes for each scenario, which represents the number of times a scenario is selected in a forest of 1000 simulated trees. The scenario with the most classification votes predicted was selected (Table S3.7).

In step 2, we added NW to the scenario selected in step 1. Here, we tested four scenarios. In the scenario 1-3, we hypothesized that NW could originate from N1, NE or N2 and the last scenario we hypothesized N2 diverged from NW and was indirectly diverged from N1.

In step 3, we identified the optimal evolutionary history for all five native source populations. Here, we added S population to the scenario selected in step 3. We tested two scenarios to examine whether S originated from N2 or NW, based on the phylogenetic tree where S, NW and N2 form a well-supported monophyletic clade.

In steps 4-8 we identified the invasion history for each invasive population. We incorporated individual invasive populations into the optimal native population scenario selected in step 3. We tested the invasion history for IL (step 4), TOR (step 5), OH (step 6), MA (step 7) and Far (step 8), respectively. In each step, we iteratively tested the invasive origin from all five native regions. Each invasive population experienced a genetic bottleneck and reduction in effective population size, followed by a population expansion. We selected the optimal invasion history model in this step.

Prior parameters

We considered two types of prior parameters, population size (N) and time (T). N includes effective population size for populations at the time of sampling and the population size of the invasive founder population, while T includes divergence timepoint and bottleneck duration. We kept the prior parameters of the historical divergence time, the effective population size of the contemporary populations as broad range 10-10,000 as default (Table S3.6). We chose a smaller range for the effective population size of founder populations as 10-200. For the prior intervals of the introduction date and bottleneck duration, we assumed a single generation per year since this has been observed in many regions (Haack et al., 2010; Hu et al., 2009), although this may vary in parts of the native and invasive range (Coyle et al., 2021; Hu et al., 2009; Straw et al., 2015). We assumed a single introduction in each invasive location in our tested scenarios based on our STRUCTURE and DAPC results (Figures 3.3,3.4; Table S3.1). Therefore, the priors for invasion events were defined based on the first sampling year. We adjusted the prior interval accordingly based on their first official record and our first sampling date, with taking into account of a lag phase of 20 years based on Morimoto's finding that lag time takes 4.4-23.2 years for the invasive insects they studied (Morimoto et al., 2019) (see Table S3.6 for details).

3.8.2 Supplementary Figures

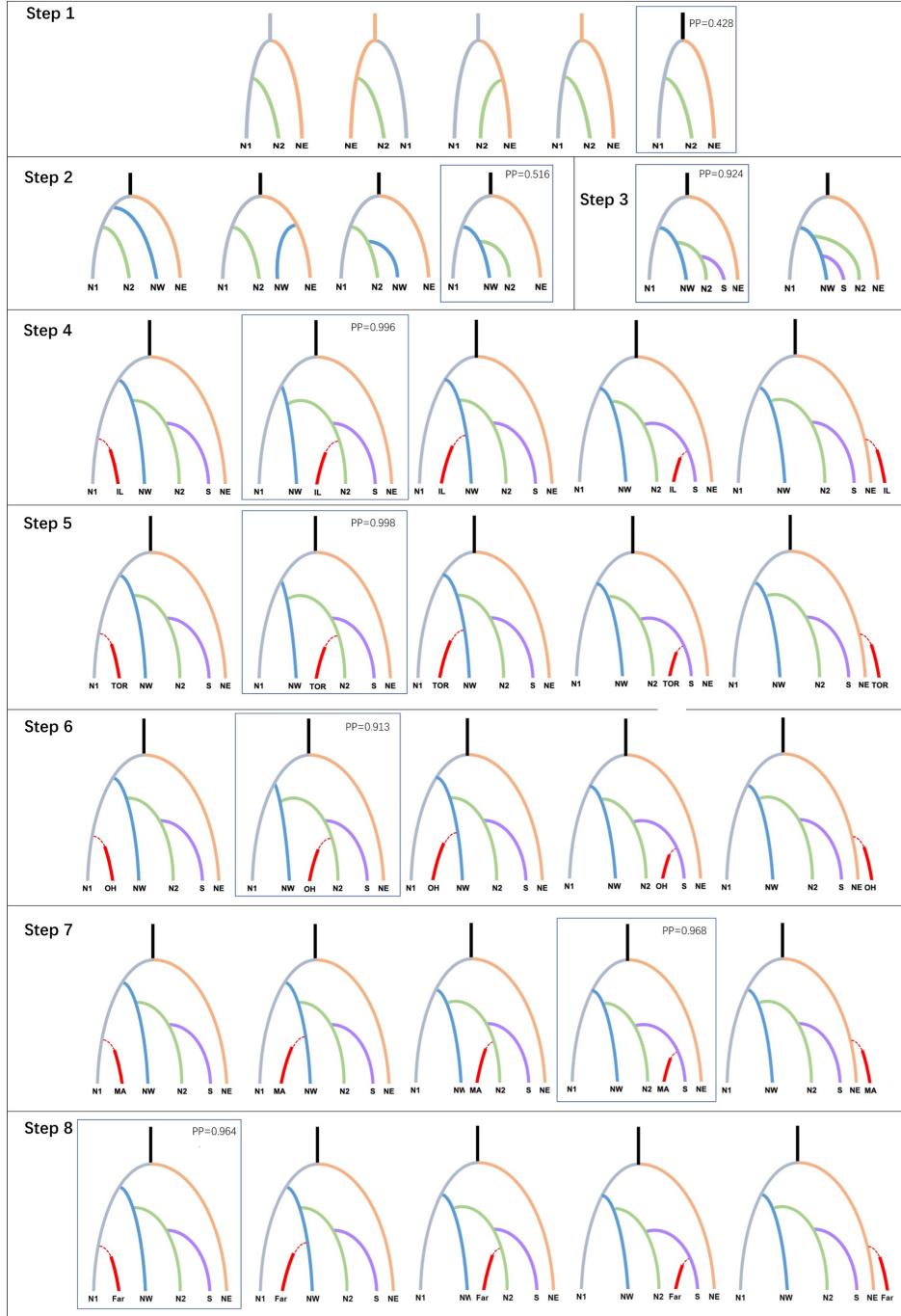


Figure S3.1: Schematic representation of DIYABC analysis and model selection. N1: North region 1 (grey). N2: North region 2 (green). NW: Northwest (blue). NE: Northeast (yellow). S: South (purple). IL: Illinois. TOR: Toronto. OH: Ohio. MA: Massachusetts. Far: Farmingdale. Invasive population in step 4-8 is colored in red. The dashed line (red) indicates bottleneck event. The unsampled ancestry population is colored in black. The classification votes are indicated under each scenario. PP: Posterior probability for the selected model with the most votes.

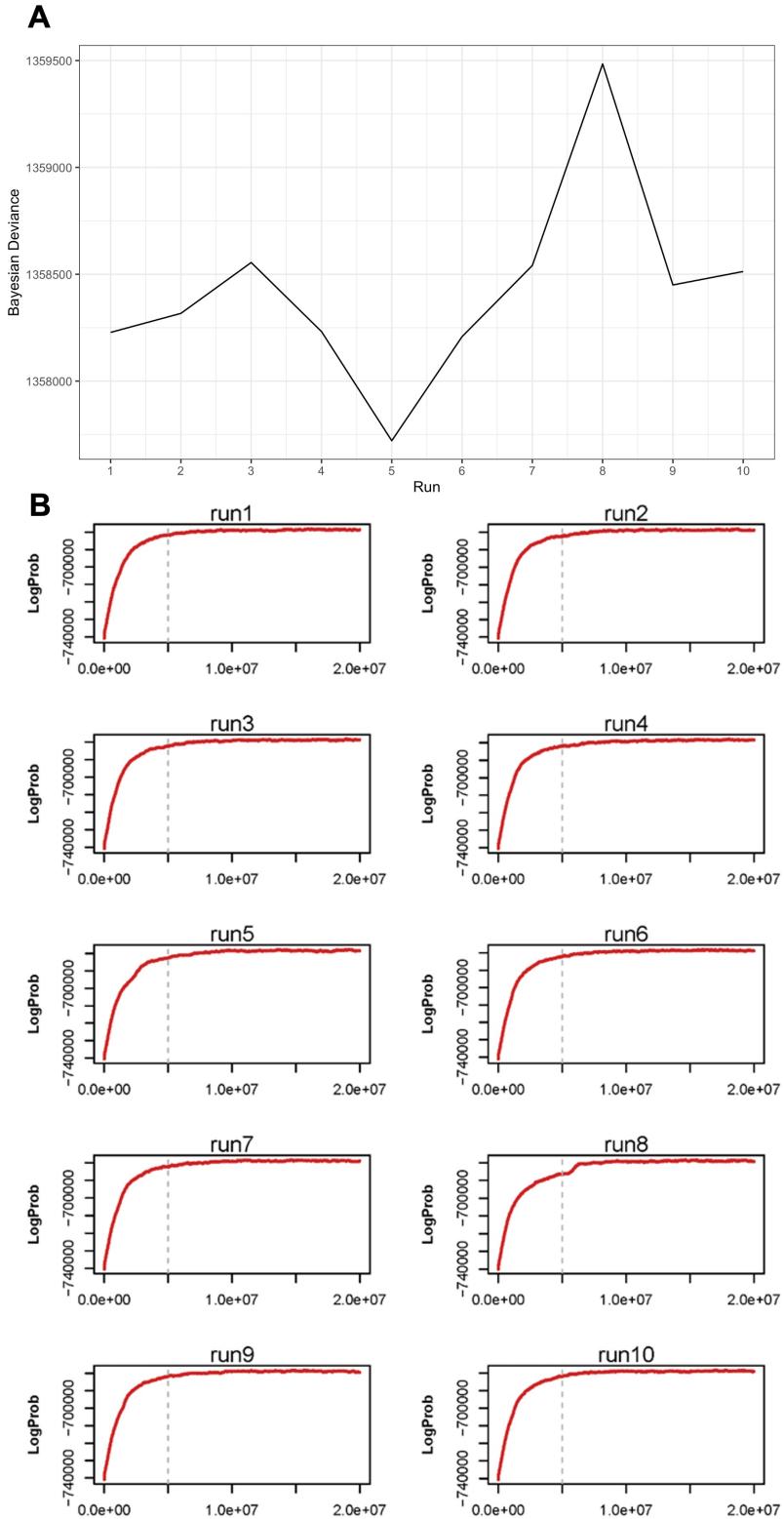


Figure S3.2: Ten replicate runs in BayesAss. **A** Bayesian deviance plot of ten runs. **B** Trace plots. X-axes are the state of the iterations. The vertical lines indicate the end of the burn-in period (i.e., 5,000,000 iterations). Run 5 (A) showed the lowest Bayesian deviance and was selected for parameter estimation.

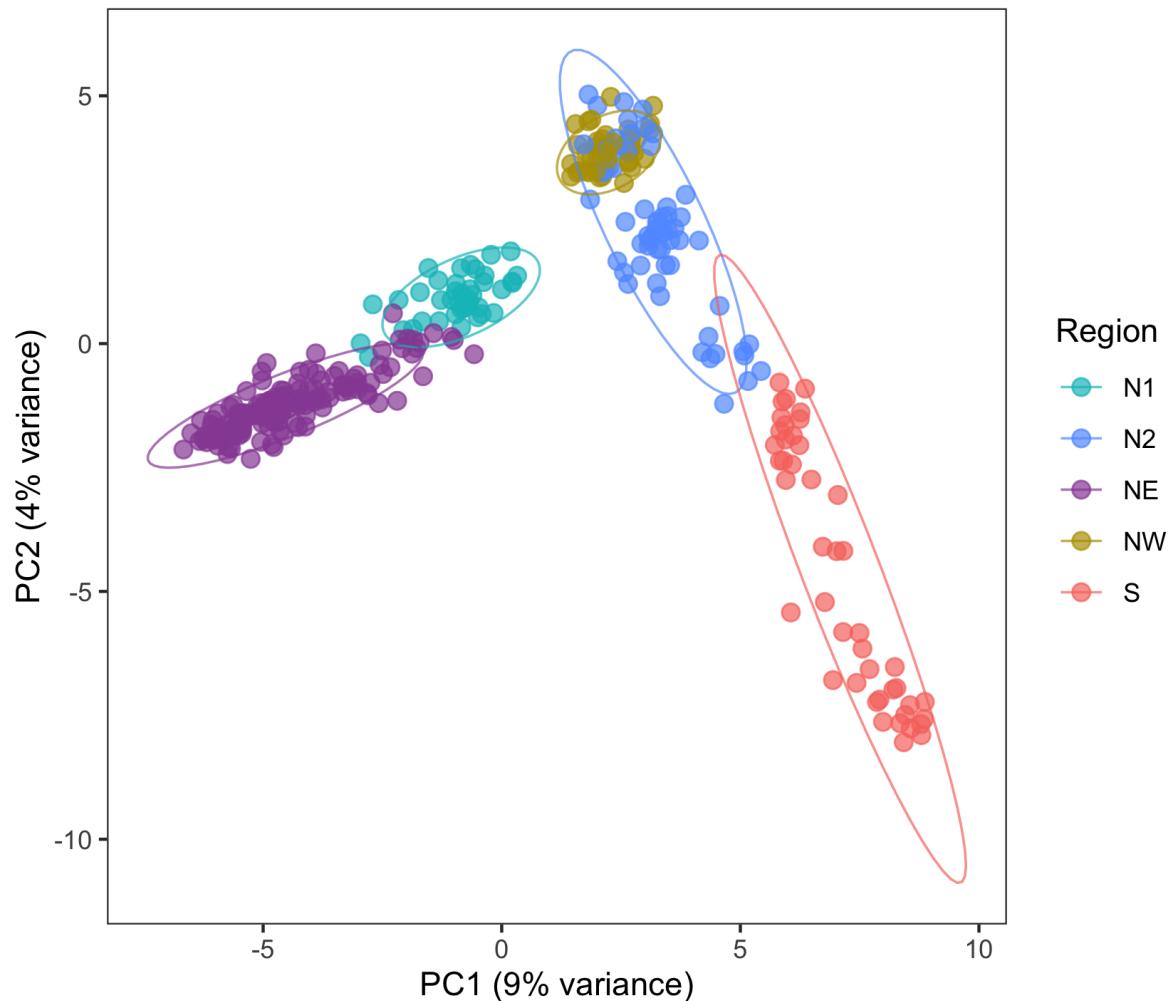


Figure S3.3: Principal component analysis for native ALBs. Colored groups represent different regions in China, i.e., N1 (North 1), N2 (North 2), NE (Northeast), NW (Northwest) and S (South).

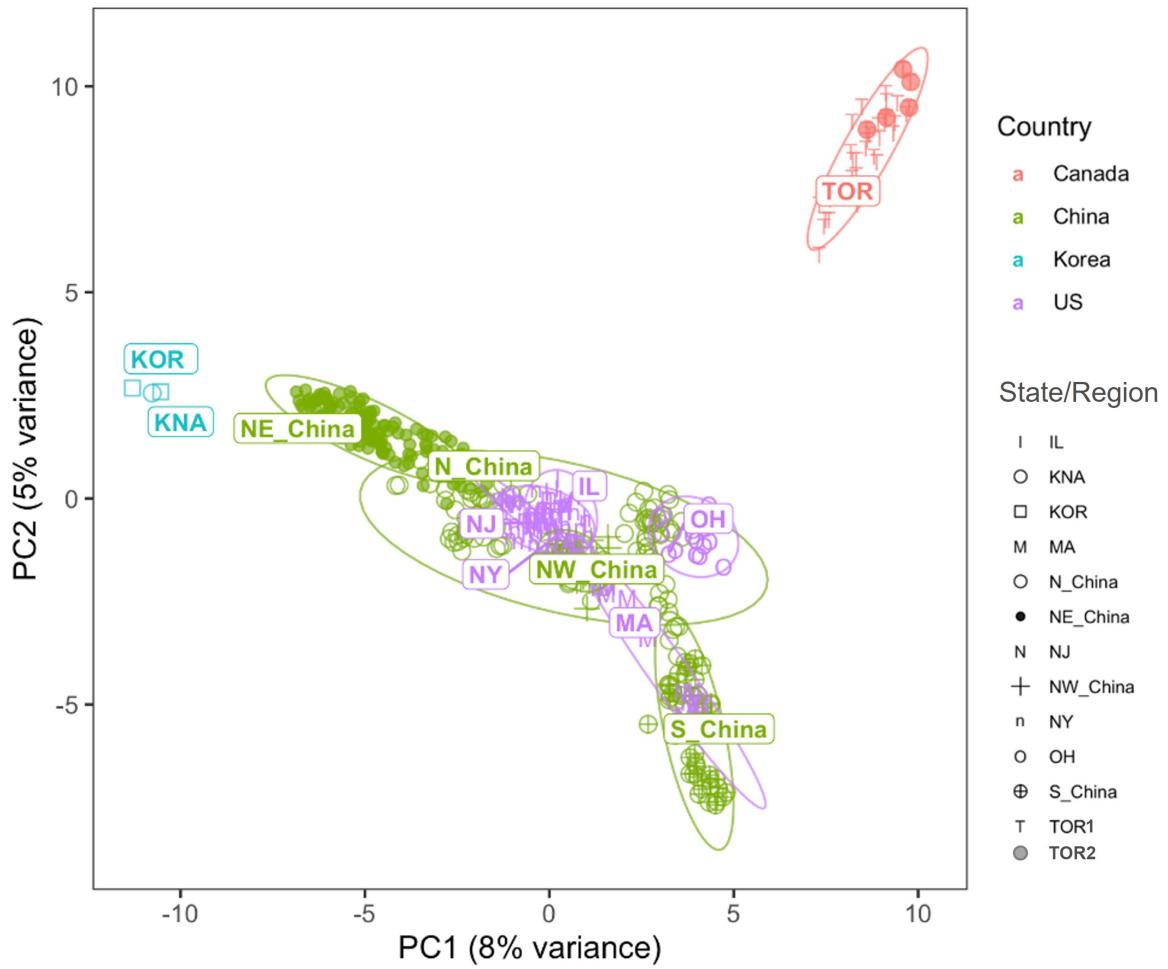


Figure S3.4: Principal component analysis for global ALBs with each population colored and shaped differently. Within Toronto (TOR), the 'T' shaped symbol indicates ALB sampled from the first infestation and the circular symbol denotes the second infestation.

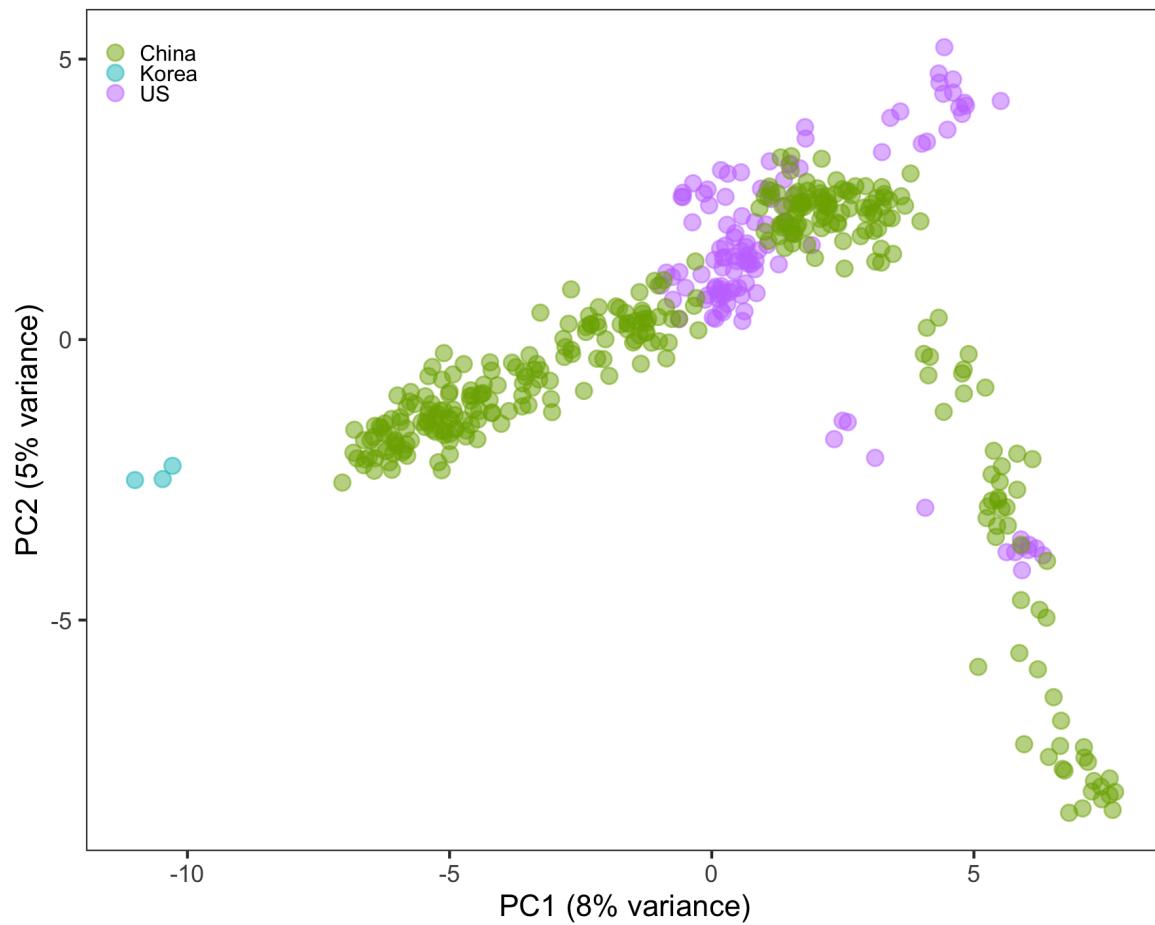


Figure S3.5: Principal component analysis for global ALBs without Toronto samples.

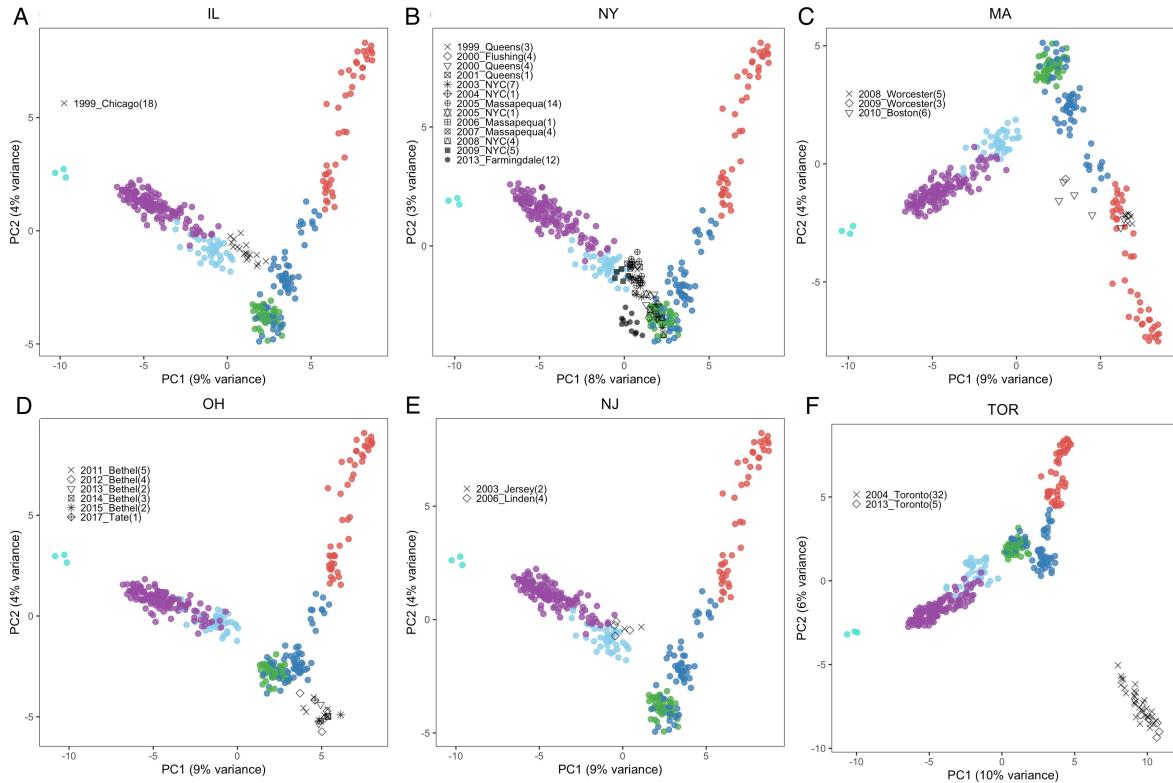


Figure S3.6: Principal component analysis for each invasive sampling locality projecting onto the native ALBs. The collection year is indicated. **A** Illinois samples. **B** New York samples. **C** Massachusetts samples. **D** Ohio samples. **E** New Jersey samples. **F** Toronto samples. The colored symbols represent native populations: North 1 (sky blue), North 2 (blue), Northeast (mauve), South (red), Northwest (green) and Korea samples (light turquoise). Invasive samples are in black.

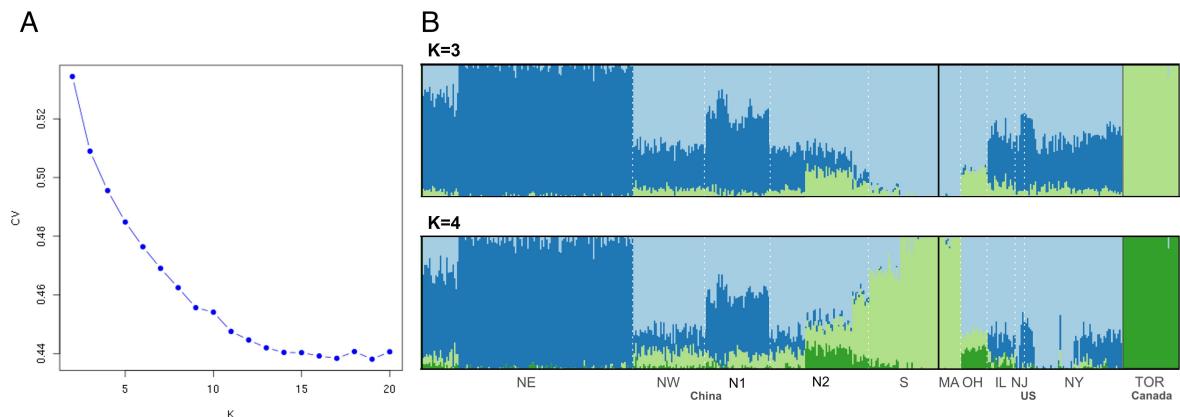


Figure S3.7: Admixture analysis. **A** Cross validation (CV) error plot. X-axis shows numbers of K and y-axis shows CV error. **B** Bar plot when K = 3 and K = 4.

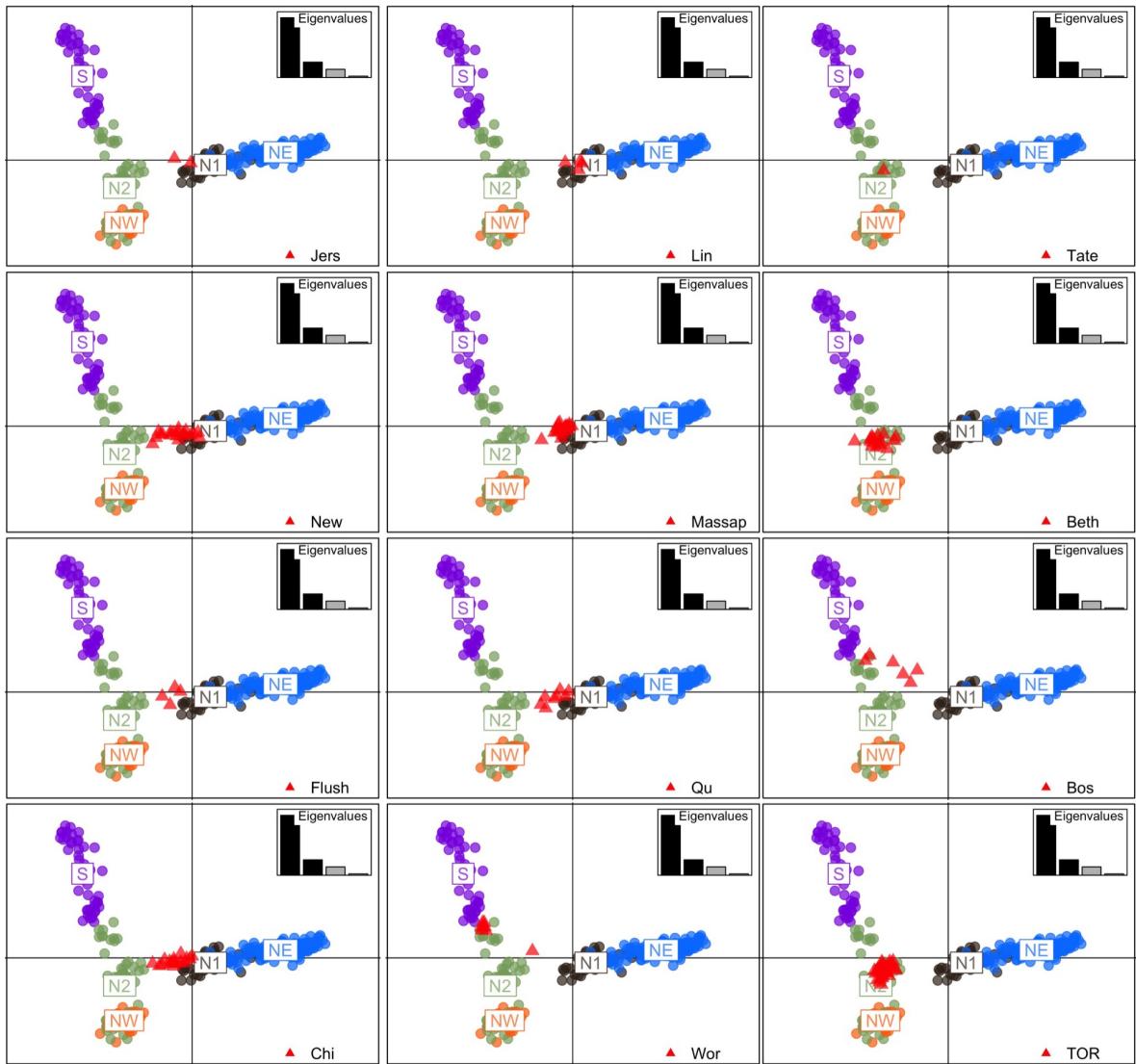


Figure S3.8: DAPC assignment plot for each invasive population. On each plot, the respective invasive population individuals are shown with red triangle symbols. The five lineages of the native ALB populations are directly named on the plots and are colored differently (dots). Jers: Jersey City, New Jersey. Lin: Linden, New Jersey. Tate: Tate, Ohio. New: New York, New York. Massap: Massapequa, New York. Beth: Bethel, Ohio. Flush: Flushing, New York. Qu: Queens, New York. Bos: Boston, Massachusetts. Chi: Chicago, Illinois. Wor: Worcester, Massachusetts. TOR: Toronto, Ontario.

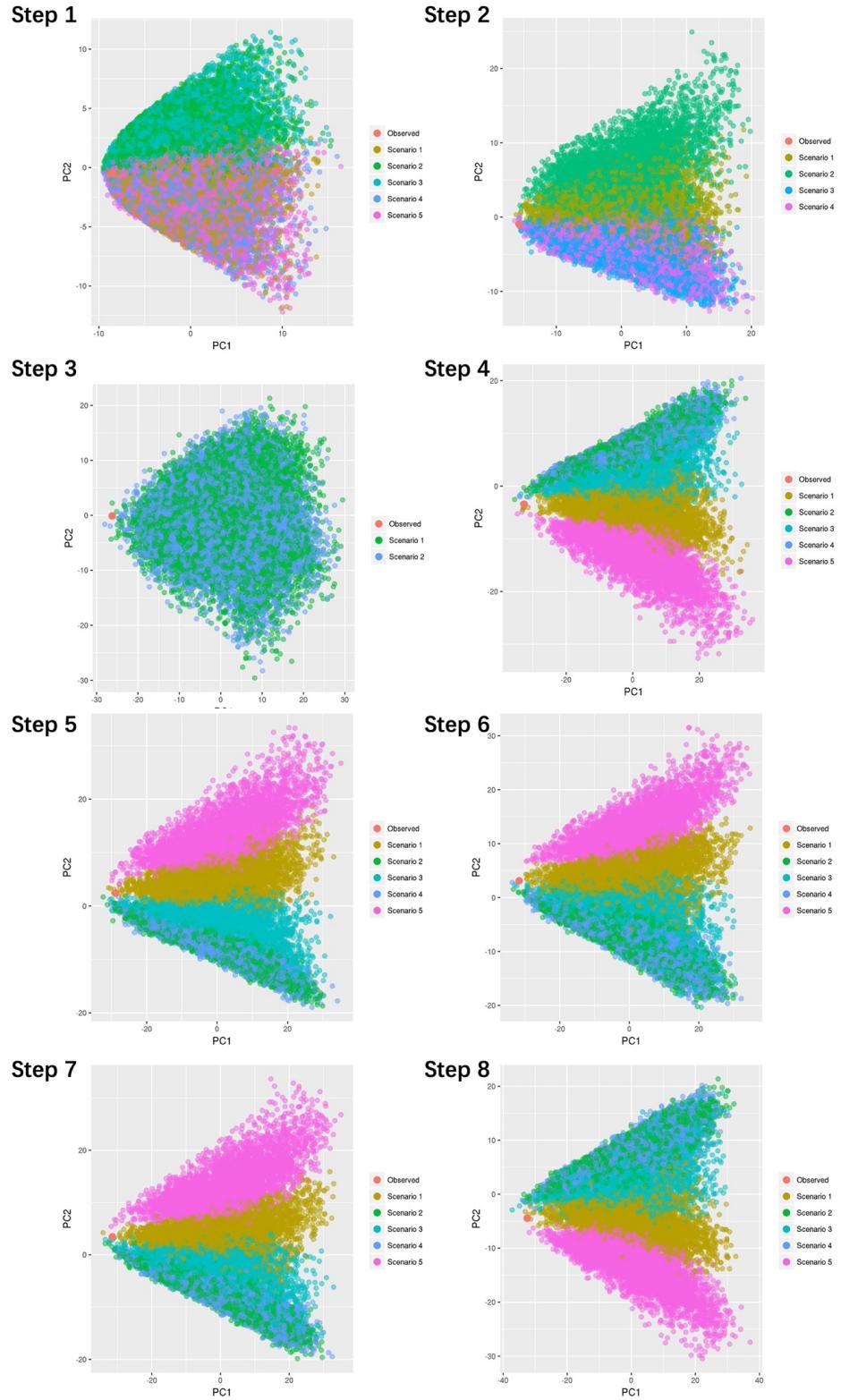


Figure S3.9: PCA for model checking of each step (1-8) in the DIYABC. The corresponding scenarios are shown in Figure S3.1.

3.9 References

- Abed, A., Légaré, G., Pomerleau, S., St-Cyr, J., Boyle, B., & Belzile, F. J. (2019). Genotyping-by-sequencing on the ion torrent platform in barley. *Methods in Molecular Biology*, 1900, 233–252.
- Akita A., Katô T., Yanagi T., & Kubota K. (2021) Reports of the alien species *Anoplophora glabripennis* (Motschulsky, 1853) (Coleoptera, Cerambycidae) found in Hyogo pref., Japan. *Gekkan-Mushi* 601:41–45 (in Japanese)
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.
- Altmüller, J., Budde, B. S., & Nürnberg, P. (2014). Enrichment of target sequences for next-generation sequencing applications in research and diagnostics. *Biological Chemistry*, 395, 231–237. <https://doi.org/10.1515/hsz-2013-0199>
- Anagnostakis, S. L. (1987). Chestnut Blight: The Classical Problem of an Introduced Pathogen. *Mycologia*, 79(1), 23–37. <https://doi.org/10.1080/00275514.1987.12025367>
- Aukema, J. E., Leung, B., Kovacs, K., Chivers, C., Britton, K. O., Englin, J., Frankel, S. J., Haight, R. G., Holmes, T. P., Liebhold, A. M., McCullough, D. G., & Holle, B. V. (2011). Economic Impacts of Non-Native Forest Insects in the Continental United States. *PLoS ONE*, 6(9), e24587. <https://doi.org/10.1371/journal.pone.0024587>
- Aukema, J. E., McCullough, D. G., Holle, B. V., Liebhold, A. M., Britton, K., & Frankel, S. J. (2010). Historical Accumulation of Nonindigenous Forest Pests in the Continental United States. *BioScience*, 60(11), 886–897. <https://doi.org/10.1525/bio.2010.60.11.5>
- Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, 2019, e7724.
- Bilodeau, P., Roe, A. D., Bilodeau, G., Blackburn, G. S., Cui, M., Cusson, M., Doucet, D., Griess, V. C., Lafond, V. M. A., Nilausen, C., Paradis, G., Porth, I., Prunier, J., Srivastava, V., Stewart, D., Torson, A. S., Tremblay, E., Uzunovic, A., Yemshanov, D., & Hamelin, R. C. (2019). Biosurveillance of forest insects: part II—adoption of genomic tools by end user communities and barriers to integration. *Journal of Pest Science*, 92, 71–82. <https://doi.org/10.1007/s10340-018-1001-1>
- Blackburn, T. M., Pyšek, P., Bacher, S., Carlton, J. T., Duncan, R. P., Jarošík, V., Wilson, J. R. U., & Richardson, D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology and Evolution*, 26, 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>

Blackburn, G. S., Bilodeau, P., Cooke, T., Cui, M., Cusson, M., Hamelin, R. C., ... & Porth, I. (2020). An applied empirical framework for invasion science: confronting biological invasion through collaborative research aimed at tool production. *Annals of the Entomological Society of America*, 113(4), 230-245.

Blumenfeld, A. J., Eyer, P.-A., Husseneder, C., Mo, J., Johnson, L. N. L., Wang, C., Grace, J. K., Chouvinc, T., Wang, S., & Vargo, E. L. (2021). Bridgehead effect and multiple introductions shape the global invasion history of a termite. *Communications Biology*, 4(1), 196. <https://doi.org/10.1038/s42003-021-01725-x>

Browett, S. S., O'Meara, D. B., & McDevitt, A. D. (2020). Genetic tools in the management of invasive mammals: recent trends and future perspectives. *Mammal Review*, 50(2), 200–210. <https://doi.org/10.1111/mam.12189>

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38, 5825–5829.

Carter, M., Smith, M., Turgeon, J., & Harrison, R. (2009) Analysis of genetic diversity in an invasive population of Asian long-horned beetles in Ontario, Canada. *Canadian Entomologist*, 141, 582–594.

Carter, M., Smith, M., & Harrison, R. (2010). Genetic analyses of the Asian longhorned beetle (Coleoptera, Cerambycidae, *Anoplophora glabripennis*), in North America, Europe and Asia. *Biological Invasions*, 12, 1165–1182.

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

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4. <https://doi.org/10.1186/s13742-015-0047-8>

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6, 80–92.

Ciosi, M., Miller, N. J., Kim, K. S., Giordano, R., Estoup, A., & Guillemaud, T. (2008). Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology*, 17(16), 3614-3627.

Colautti, R. I., Grigorovich, I. A., & MacIsaac, H. J. (2006). Propagule pressure: a null model

- for biological invasions. *Biological invasions*, 8, 1023-1037.
- Collin, F. D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., ... Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources*, 21, 2598–2613.
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... & Estoup, A. (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30(8), 1187-1189.
- Coyle, D. R., Trotter, R. T., Bean, M. S., & Pfister, S. E. (2021). First Recorded Asian Longhorned Beetle (Coleoptera: Cerambycidae) Infestation in the Southern United States. *Journal of Integrated Pest Management*, 12(1). <https://doi.org/10.1093/jipm/pma007>
- Cui, M., Wu, Y., Javal, M., Giguère, I., Roux, G., Andres, J. A., ... Porth, I. (2022). Genome-scale phylogeography resolves the native population structure of the Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky). *Evolutionary Applications*, 15(6), 934.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- Danko, J. J. V., Delauer, D. G. Martin, and J. Rogan. 2016. Asian longhorned beetle outbreak in Worcester, Massachusetts: A history of urban forestry practices. *The Northeastern Geographer* 8:1–21
- Darling, J. A. (2015). Genetic studies of aquatic biological invasions: closing the gap between research and management. *Biological Invasions*, 17(3), 951–971. <https://doi.org/10.1007/s10530-014-0726-x>
- Diepenbroek, M., Bayer, B., Schwender, K., Schiller, R., Lim, J., Lagacé, R., & Anslinger, K. (2020). Evaluation of the ion AmpliseqTM phenotriuum panel: MPS-based assay for ancestry and phenotype predictions challenged by casework samples. *Genes*, 11, 1–24. <https://doi.org/10.3390/genes11121398>
- Slugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, 17(1), 431–449. <https://doi.org/10.1111/j.1365-294x.2007.03538.x>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Epanchin-Niell, R. S., & Liebhold, A. M. (2015). Benefits of invasion prevention: Effect

of time lags, spread rates, and damage persistence. *Ecological Economics*, 116, 146–153. <https://doi.org/10.1016/j.ecolecon.2015.04.014>

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is There a Genetic Paradox of Biological Invasion? *Annual Review of Ecology, Evolution, and Systematics*, 47, 51–72. <https://doi.org/10.1146/annurev-ecolsys-121415-032116>

Eyer, P., Blumenfeld, A. J., Johnson, L. N. L., Perdereau, E., Shults, P., Wang, S., Dedeine, F., Dupont, S., Bagnères, A., & Vargo, E. L. (2021). Extensive human-mediated jump dispersal within and across the native and introduced ranges of the invasive termite *Reticulitermes flavipes*. *Molecular Ecology*, 30(16), 3948–3964. <https://doi.org/10.1111/mec.16022>

Facon, B., Hufbauer, R. A., Tayeh, A., Loiseau, A., Lombaert, E., Vitalis, R., Guillemaud, T., Lundgren, J. G., & Estoup, A. (2011). Inbreeding Depression Is Purged in the Invasive Insect *Harmonia axyridis*. *Current Biology*, 21(5), 424–427. <https://doi.org/10.1016/j.cub.2011.01.068>

Flanagan, S. P., & Jones, A. G. (2017). Constraints on the FST -Heterozygosity Outlier Approach. *Journal of Heredity*, 108, 561–573.

Flower, C. E., Knight, K. S., & Gonzalez-Meler, M. A. (2013). Impacts of the emerald ash borer (*Agrilus planipennis* Fairmaire) induced ash (*Fraxinus* spp.) mortality on forest carbon cycling and successional dynamics in the eastern United States. *Biological Invasions*, 15(4), 931–944. <https://doi.org/10.1007/s10530-012-0341-7>

Garnas, J. R., Auger-Rozenberg, M.-A., Roques, A., Bertelsmeier, C., Wingfield, M. J., Saccaggi, D. L., Roy, H. E., & Slippers, B. (2016). Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biological Invasions*, 18(4), 935–952. <https://doi.org/10.1007/s10530-016-1082-9>

Garnas, J. R., Hurley, B. P., Slippers, B., & Wingfield, M. J. (2012). Biological control of forest plantation pests in an interconnected world requires greater international focus. *International Journal of Pest Management*, 58(3), 211–223. <https://doi.org/10.1080/09670874.2012.698764>

Gatto-Almeida, F., Pichlmüller, F., Bodey, T. W., Samaniego, A., & Russell, J. C. (2022). The tails of two invasive species: genetic responses to acute and chronic bottlenecks. *Biological Invasions*, 24(10), 3263–3273.

Glémén, S. (2003). How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution*, 57(12), 2678–2687.

Greenwood, L. F., Coyle, D. R., Guerrero, M. E., Hernández, G., MacQuarrie, C. J. K., Trejo, O., & Noseworthy, M. K. (2023). Exploring pest mitigation research and management

associated with the global wood packaging supply chain: What and where are the weak links? *Biological Invasions*, 1–27. <https://doi.org/10.1007/s10530-023-03058-8>

Haack, R. A., Bauer, L. S., Gao, R. T., McCarthy, J. J., Miller, D. L., Petrice, T. R., & Poland, T. M. (2006). *Anoplophora glabripennis* within-tree distribution, seasonal development, and host suitability in China and Chicago. *Great Lakes Entomologist*, 39, 169–183.

Haack, R. A., Britton, K. O., Brockerhoff, E. G., Cavey, J. F., Garrett, L. J., Kimberley, M., ... & Vasilaky, K. N. (2014). Effectiveness of the International Phytosanitary Standard ISPM No. 15 on reducing wood borer infestation rates in wood packaging material entering the United States. *PLoS One*, 9(5), e96611.

Haack, R. A., Hérard, F., Sun, J., & Turgeon, J. J. (2010). Managing invasive populations of asian longhorned beetle and citrus longhorned beetle: A worldwide perspective. *Annual Review of Entomology*, 55, 521–546. <https://doi.org/10.1146/annurev-ento-112408-085427>

Haack, R. A., Law, K. R., Mastro, V. C., Ossenbruggen, H. S., & Raimo, B. J. (1997). New York's battle with the Asian long-horned beetle. *Journal of Forestry*, 95(12), 11-15.

Hamelin, R. C., & Roe, A. D. (2020). Genomic biosurveillance of forest invasive alien enemies: A story written in code. *Evolutionary Applications*, 13, 95–115. <https://doi.org/10.1111/eva.12853>

Hu, J., Angeli, S., Schuetz, S., Luo, Y., & Hajek, A. E. (2009). Ecology and management of exotic and endemic Asian longhorned beetle *Anoplophora glabripennis*. *Agricultural and Forest Entomology*, 11, 359–375.

Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>

Javal, M., Lombaert, E., Tsykun, T., Courtin, C., Kerdelhué, C., Prospero, S., ... & Roux, G. (2019). Deciphering the worldwide invasion of the Asian long-horned beetle: A recurrent invasion process from the native area together with a bridgehead effect. *Molecular ecology*, 28(5), 951-967.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

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

Joshi, N. (2011). Sabre: a barcode demultiplexing and trimming tool for FastQ files. *GitHub: San Francisco, CA, USA*.

Kaňuch, P., Cassel-Lundhagen, A., Preuss, S., Nordlander, G., & Berggren, Å. (2022). Parapatric Genetic Lineages Persist in a Multiply Introduced Non-native Bush-Cricket. *Frontiers in Ecology and Evolution*, 10, 812079. <https://doi.org/10.3389/fevo.2022.812079>

Keena, M. A. (2002). *Anoplophora glabripennis* (Coleoptera: Cerambycidae) fecundity and longevity under laboratory conditions: Comparison of populations from New York and Illinois on *Acer saccharum*. *Environmental Entomology*, 31, 490–498.

Keena, M. A., & Moore, P. M. (2010). Effects of Temperature on *Anoplophora glabripennis* (Coleoptera: Cerambycidae) larvae and pupae. *Environmental Entomology*, 39, 1323–1335.

Kim, H., Kim, S., Lee, Y., Lee, H.-S., Lee, S.-J., & Lee, J.-H. (2021). Tracing the Origin of Korean Invasive Populations of the Spotted Lanternfly, *Lycorma delicatula* (Hemiptera: Fulgoridae). *Insects*, 12(6), 539. <https://doi.org/10.3390/insects12060539>

Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, 431(7005), 177–181. <https://doi.org/10.1038/nature02807>

Krishnankutty, S., Nadel, H., Taylor, A. M., Wiemann, M. C., Wu, Y., Lingafelter, S. W., Myers, S. W., & Ray, A. M. (2020). Identification of Tree Genera Used in the Construction of Solid Wood-Packaging Materials That Arrived at U.S. Ports Infested With Live Wood-Boring Insects. *Journal of Economic Entomology*, 113(3), 1183–1194. <https://doi.org/10.1093/jee/toaa060>

Ladin, Z. S., Eggen, D. A., Trammell, T. L. E., & D'Amico, V. (2023). Human-mediated dispersal drives the spread of the spotted lanternfly (*Lycorma delicatula*). *Scientific Reports*, 13(1), 1098. <https://doi.org/10.1038/s41598-022-25989-3>

Lee, S., Lee, Y., & Lee, S. (2020). Population genetic structure of *Anoplophora glabripennis* in South Korea: Invasive populations in the native range? *Journal of Pest Science*, 93, 1181–1196. <https://doi.org/10.1007/s10340-020-01245-3>

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Lingafelter, S. W., & Hoebeke, R. E. (2002). Revision of the Genus *Anoplophora* (Coleoptera: Cerambycidae). *Entomological Society of Washington*, Washington DC.

Lockwood, J. L., Cassey, P., & Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends in ecology & evolution*, 20(5), 223-228.

- Lombaert, E., Guillemaud, T., Cornuet, J.-M., Malausa, T., Facon, B., & Estoup, A. (2010). Bridgehead Effect in the Worldwide Invasion of the Biocontrol Harlequin Ladybird. *PLoS ONE*, 5(3), e9743. <https://doi.org/10.1371/journal.pone.0009743>
- Loxdale, H. D. (2010). Rapid genetic changes in natural insect populations. *Ecological Entomology*, 35, 155–164. <https://doi.org/10.1111/j.1365-2311.2009.01141.x>
- Luo, Y., Huang, J., & Li, J. (2000). The main achievements, problems and prospects in research of long-horned beetles in China. *Entomological Knowledge*, 37, 116–121.
- MacLeod, A., Evans, H. F., & Baker, R. H. A. (2002). An analysis of pest risk from an Asian longhorn beetle (*Anoplophora glabripennis*) to hardwood trees in the European community. *Crop Protection*, 21, 635–645.
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26, 2867–2873.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), 10-12.
- Mergeay, J., Verschuren, D., & Meester, L. D. (2006). Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603), 2839-2844.
- McKenna, D. D., Scully, E. D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S. M., ... Richards, S. (2016). Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology*, 17, 1–18.
- Meirmans, P. G. (2014). Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources*, 14, 726–733.
- Melbourne, B. A., & Hastings, A. (2009). Highly Variable Spread Rates in Replicated Biological Invasions: Fundamental Limits to Predictability. *Science*, 325(5947), 1536–1539. <https://doi.org/10.1126/science.1176138>
- Meng, P. S., Hoover, K., & Keena, M. A. (2015). Asian longhorned beetle (Coleoptera: Cerambycidae), an introduced pest of maple and other hardwood trees in North America and Europe. *Journal of Integrated Pest Management*, 6, 1–13.
- Morimoto, N., Kiritani, K., Yamamura, K., & Yamanaka, T. (2019). Finding indications of lag time, saturation and trading inflow in the emergence record of exotic agricultural insect pests in Japan. *Applied Entomology and Zoology*, 54(4), 437–450. <https://doi.org/10.1007/s13355-019-00640-2>

Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods in Ecology and Evolution*, 10, 1808–1813.

Nowak, D. J., Pasek, J. E., Sequeira, R. A., Crane, D. E., & Mastro, V. C. (2001). Potential effect of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) on urban trees in the United States. *Journal of Economic Entomology*, 94, 116–122.

Ortego, J., Céspedes, V., Millán, A., & Green, A. J. (2021). Genomic data support multiple introductions and explosive demographic expansions in a highly invasive aquatic insect. *Molecular Ecology*, 30(17), 4189–4203. <https://doi.org/10.1111/mec.16050>

Pejchar, L., & Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, 24(9), 497–504. <https://doi.org/10.1016/j.tree.2009.03.016>

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7, e37135.

Picq, S., Keena, M., Havill, N., Stewart, D., Pouliot, E., Boyle, B., Levesque, R. C., Hamelin, R. C., & Cusson, M. (2018). Assessing the potential of genotyping-by-sequencing-derived single nucleotide polymorphisms to identify the geographic origins of intercepted gypsy moth (*Lymantria dispar*) specimens: A proof-of-concept study. *Evolutionary Applications*, 11, 325–339. <https://doi.org/10.1111/eva.12559>

Poland, T. (1998). Chicago joins New York in battle with the Asian longhorned beetle. *Newsl. Mich. Entomol.* 43, 15–17.

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, 32(6), 859–866. <https://doi.org/10.1093/bioinformatics/btv684>

Pyšek, P., & Richardson, D. M. (2010). Invasive Species, Environmental Change and Management, and Health. *Annual Review of Environment and Resources*, 35(1), 25–55. <https://doi.org/10.1146/annurev-environ-033009-095548>

Rassati, D., Haack, R. A., Knížek, M., & Faccoli, M. (2018). National Trade can drive range expansion of bark- and wood-boring beetles. *Journal of Economic Entomology*, 111, 260–268. <https://doi.org/10.1093/jee/tox308>

Rees, C. B., Hand, B. K., Carter, S. C., Bargeron, C., Cline, T. J., Daniel, W., Ferrante, J. A., Gaddis, K., Hunter, M. E., ... Luikart, G. (2022). A framework to integrate innovations in invasion science for proactive management. *Biological Reviews*, 97(4), 1712–1735. <https://doi.org/10.1111/br.12750>

[//doi.org/10.1111/brv.12859](https://doi.org/10.1111/brv.12859)

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., Wilkie, A. O., ... Lunter, G. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, 46, 912–918.

Rius, M., & Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, 29(4), 233–242. <https://doi.org/10.1016/j.tree.2014.02.003>

Roe, A. D., Torson, A. S., Bilodeau, G., Bilodeau, P., Blackburn, G. S., Cui, M., Cusson, M., Doucet, D., Griess, V. C., Lafond, V., Paradis, G., Porth, I., Prunier, J., Srivastava, V., Tremblay, E., Uzunovic, A., Yemshanov, D., & Hamelin, R. C. (2019). Biosurveillance of forest insects: part I—integration and application of genomic tools to the surveillance of non-native forest insects. *Journal of Pest Science*, 92, 51–70. <https://doi.org/10.1007/s10340-018-1027-4>

Santos, R., and L. Cole. 2012. Massachusetts Asian longhorned beetle cooperative eradication program determines next steps for eradication efforts in Worcester County. Washington, DC: U.S. Department of Agriculture.

Sawyer, A., Mastro, V., Emens, B. C., & Denholm, T. (2006). Infestation dynamics of the Asian longhorned beetle in the United States. In *Proceedings, 17th US Department of Agriculture interagency research forum on gypsy moth and other invasive species* (Vol. 2006, pp. 10-13).

Schrader, G., & Unger, J. G. (2003). Plant quarantine as a measure against invasive alien species: the framework of the International Plant Protection Convention and the plant health regulations in the European Union. *Biological Invasions*, 5, 357-364.

Schrieber, K., & Lachmuth, S. (2017). The genetic paradox of invasions revisited: The potential role of inbreeding × environment interactions in invasion success. *Biological Reviews*, 92, 939–952. <https://doi.org/10.1111/brv.12263>

Sherpa, S., Blum, M. G. B., Capblancq, T., Cumer, T., Rioux, D., & Després, L. (2019). Unravelling the invasion history of the Asian tiger mosquito in Europe. *Molecular Ecology*, 28(9), 2360–2377. <https://doi.org/10.1111/mec.15071>

Short, M. T., Chase, K. D., Feeley, T. E., Kees, A. M., Wittman, J. T., & Aukema, B. H. (2020). Rail transport as a vector of emerald ash borer. *Agricultural and Forest Entomology*, 22(1), 92–97. <https://doi.org/10.1111/afe.12360>

Simberloff, D. (2009). The Role of Propagule Pressure in Biological Invasions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 81–102.

Sjöman, H., Östberg, J., Nilsson, J., Forestry, J. N.-A. & U., 2014, undefined, & Nilsson, J. (2014). Review of host trees for the wood-boring pests *Anoplophora glabripennis* and *Anoplophora chinensis*: an urban forest perspective. *Arboriculture & Urban Forestry*, 40(3), 143-164.

Smetana, S., Schmitt, E., & Mathys, A. (2019). Sustainable use of *Hermetia illucens* insect biomass for feed and food: Attributional and consequential life cycle assessment. *Resources, Conservation and Recycling*, 144, 285–296. <https://doi.org/10.1016/j.resconrec.2019.01.042>

Smith, M. T., Bancroft, J., Li, G., Gao, R., & Teale, S. (2001). Dispersal of *Anoplophora glabripennis* (Cerambycidae). *Environmental Entomology*, 30, 1036–1040. <https://doi.org/10.1603/0046-225x-30.6.1036>

Smith, M. T., Tobin, P. C., Bancroft, J., Li, G., & Gao, R. (2004). Dispersal and spatiotemporal dynamics of Asian longhorned beetle (Coleoptera: Cerambycidae) in China. *Environmental Entomology*, 33, 435–442.

Spaulding, H. L., & Rieske, L. K. (2010). The aftermath of an invasion: Structure and composition of Central Appalachian hemlock forests following establishment of the hemlock woolly adelgid, *Adelges tsugae*. *Biological Invasions*, 12, 3135-3143.]

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

Straw, N. A., Tilbury, C., Fielding, N. J., Williams, D. T., & Cull, T. (2015). Timing and duration of the life cycle of Asian longhorn beetle *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in southern England. *Agricultural and Forest Entomology*, 17(4), 400–411. <https://doi.org/10.1111/afe.12120>

Tay, W. T., Jr., R. L. M., Czepak, C., & Groot, A. T. (2022). *Spodoptera frugiperda*: Ecology, Evolution, and Management Options of an Invasive Species. *Annual Review of Entomology*, 68(1), 299–317. <https://doi.org/10.1146/annurev-ento-120220-102548>

Teixeira, J. C., & Huber, C. D. (2021). Authors' Reply to Letter to the Editor: Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics*, 22(4), 547-549.

Tian Y., Liu R., Niu X., and Sheng Z. (1989). Study on the Loss of Poplar Wood Caused by *Anoplophora glabripennis*. *Journal of Shandong Forestry Science and Technology*, 1(3), 92-93. (In Chinese with English title)

Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLoS ONE*, 11, 1–14.

Torkamaneh, D., Laroche, J., Bastien, M., Abed, A., & Belzile, F. (2017). Fast-GBS: A new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. *BMC Bioinformatics*, 18, 1–7.

Trotter, R. T., & Hull-Sanders, H. M. (2015). Quantifying Dispersal of the Asian longhorned beetle (*Anoplophora glabripennis*, Coleoptera) with incomplete data and behavioral knowledge. *Biological Invasions*, 17, 3359–3369. <https://doi.org/10.1007/s10530-015-0961-9>

Turgeon, J. J., Gasman, B., Smith, M. T., Pedlar, J. H., Orr, M., Fournier, R. E., Doyle, J., Ric, J., & Scarr, T. (2022). Canada's response to invasion by Asian longhorned beetle (Coleoptera: Cerambycidae) in Ontario. *The Canadian Entomologist*, 154(s1), e1. <https://doi.org/10.4039/tce.2021.60>

Turgeon, J. J., Orr, M., Grant, C., Wu, Y., & Gasman, B. (2015). Decade-old satellite infestation of *Anoplophora glabripennis* Motschulsky (Coleoptera: Cerambycidae) found in Ontario, Canada outside regulated area of founder population. *The Coleopterists Bulletin*, 69(4), 674–678.

Turgeon, J. J., Pedlar, J., Groot, P. D., Smith, M. T., Jones, C., Orr, M., & Gasman, B. (2010). Density and location of simulated signs of injury affect efficacy of ground surveys for Asian longhorned beetle. *Canadian Entomologist*, 142, 80–96. <https://doi.org/10.4039/n09-049>

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177–1191.

Wongnikong, W., Hereward, J. P., Brunschot, S. L. van, & Walter, G. H. (2021). Multiple invasions of *Bemisia argentifolii* into Australia and its current genetic connectivity across space. *Journal of Pest Science*, 94(4), 1331–1343.

Wu, Y., Trepanowski, N. F., Molongoski, J. J., Reagel, P. F., Lingafelter, S. W., Nadel, H., ... & Ray, A. M. (2017). Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in trade-associated solid wood packaging material using DNA barcoding and morphology. *Scientific Reports*, 7(1), 1–12.

Yan, J. (1985). Study on the distribution area of *Anoplophora glabripennis* in East China. *Journal of North-Eastern Forestry College*, 13.

Yan, J., Qin, X., & Xiao, G. (1992). *Anoplophora glabripennis* (Motsch.). *Forest insects of China* (2nd ed., pp. 455–457). Beijing: China forestry publishing house.

Yemshanov, D., Haight, R. G., Koch, F. H., Lu, B., Venette, R., Fournier, R. E., & Turgeon, J. J. (2017). Robust surveillance and control of invasive species using a scenario optimization approach. *Ecological Economics*, 133, 86–98.

Zhou, J., Zhang, K., & Lu, Y. (1984). Study on Adult Activity and Behavioral Mechanism of

Anoplophora nobilis Ganglbauer. *Scientia Silvae Sinicae*, 20, 372–379.

Conclusion

In this study, we conducted an in-depth investigation of the Asian longhorned beetle (ALB), a destructive invasive species causing considerable ecological and economic damage worldwide, with the goal to understand its population genetic characteristics to contribute to genomic biosurveillance. Previous studies using traditional markers, such as microsatellites, has been conducted to investigate the population genetics of ALB populations (Javal et al., 2019). We found that the SNP markers worked well in defining genetic groups and even with a smaller set of these markers we were able to accurately assign individuals to populations. While their results laid the groundwork for future studies, they were not particularly designed for biosurveillance purposes. Consequently, we utilized next generation sequencing technology to sequence and obtain genome-wide markers, allowing us to conduct a comprehensive population genomics study.

In **Chapter One**, we address questions regarding the performance of genome-wide SNP markers in determining the population structure of native ALB populations. Using thousands of markers, we clearly saw different groups that were region-specific, with six distinct clusters and a noticeable divide between the North and South. We observed limited gene flow (signs of regional dispersal, likely human-mediated). We found that the SNP markers worked well in defining genetic groups and even with a smaller set of these markers we were still able to accurately assign populations.

In **Chapter Two**, we aim to examine significant signatures related to natural selection to gain a better understanding of local adaptation in ALB. By performing a whole-genome selection scan and gene-environment association analysis on pooled native ALB data, we identified over 5,000 potentially selected genes. Temperature was confirmed to play a key role in creating genetic differences. However, although we found common cold tolerance genes in our selection scan, their significance was not confirmed in the gene-environment analysis.

In **Chapter Three**, we aim to answer questions about the sources of North American ALB populations and reconstruct their invasion history. We found that most invasive populations were distinct populations and multiple independent introductions were prominent. Recurrent introductions from parts of the North Region were observed. By testing different invasion scenarios, we were able to illuminate the historical trajectory of these invasions, identifying

the most probable scenario that shaped the current invasive populations.

The findings we have gathered throughout the three chapters could pave the way for innovative approaches in biosurveillance. However, there are also concerns that arise from these findings that need to be addressed. Throughout this thesis, we recognize the limitations and challenges inherent in studying the genomic biosurveillance of ALB, and we propose potential directions for future research to address these issues.

Genomic Biosurveillance and assessment standard Genomic biosurveillance has proven to be a valuable tool for species identification, offering the advantages of not requiring intact specimens and being effective across all life stages. While the idea of employing a toolkit to pinpoint the geographic source of invasion remains somewhat conceptual and in its infancy, numerous challenges must be addressed before practical implementation. For example, when assessing the geographical source of an IAS, it is important to account for a margin of error, and a standard should be established for normalizing the accuracy of such methods.

Evolution Speed and Climate Change Impact on Insect Population The evolution speed of insect populations should be considered. Insects tend to have short generation times and high reproductive rates, which can lead to increased genetic variation and rapid evolutionary changes. With ongoing climate change, will it play a role in altering the allele frequency? How much should we expect? Although in our study, for example, the sampling year's span a few decades, the genetic makeup examined by thousands of SNPs markers remains unchanged.

Sampling Site Limitations and Integration of New Data Despite our broad sampling sites, we did not cover all ALB distributions. With a larger number of sampling sites, it is possible to obtain a somewhat different picture of the population history in ALB. For instance, we currently lack samples from European populations, which could be a valuable addition to our dataset. Acquiring samples from these regions would allow us to test for the bridgehead effect among worldwide ALB invasions and expand our findings. Furthermore, we should focus on improving reproducibility to facilitate the integration of new data as we obtain additional samples.

Integration Issues As discussed in depth in Roe et al. (2019) and Bilodeau et al. (2019), factors such as end user knowledge, efficiency, cost-effectiveness, economic feasibility, and a defined policy framework influence adoption and integration of these new tools and slow their real-world applications.

Wet lab validation To confirm the function and relevance of candidate genes identified through computational analyses, it is crucial to perform wet lab experiments. These experiments may include gene expression analysis, gene knockout or knockdown studies, and functional assays. By validating the candidate genes in a wet lab setting, we can establish a more solid link between the genetic variations and the observed phenotypes or traits.

Improved reference genome annotation The quality and completeness of a reference genome assembly and annotation substantially influence the ability to detect selection signatures with depth and accuracy. Annotated genomes provide information about the locations and functions of genes, regulatory elements, and other functional sequences. A more complete and accurate reference genome allows for better identification of genes under selection and aids in understanding the molecular mechanisms underlying adaptive traits.

Non-consistent outlier sets from different methods When using various methods to identify genomic regions under selection, it is common to observe inconsistencies in the sets of outlier loci. This discrepancy may arise due to differences in the underlying assumptions, algorithms, and statistical power of each method. To increase confidence in the results, we can employ complementary approaches, compare the findings, and focus on the regions consistently identified as outliers by multiple methods.

Genetic basis of traits In order to use genome-wide association methods between genetic variations and phenotypic traits or environmental variables to identify determining genomic signatures, it is essential to account for confounding factors such as population structure, complex interactions between multiple genes, and epigenetic modifications may influence the relationship between genetic variations and traits of interest, which can make it challenging to accurately identify the underlying genetic mechanisms. To address these challenges, we must carefully design such studies, ensuring a sufficient sample size, appropriate statistical methods, and consideration of potential confounding factors to accurately identify the genomic signatures.

Despite the numerous challenges, our work continues as the Asian longhorned beetle (ALB) remains one of the "100 worst invasive species" and our battle against invasive species persists in the face of relentless biological invasion amid globalization. The insights acquired in this study will not only aid in the application of genomic biosurveillance for managing invasive species and mitigating their impacts on ecosystems and human well-being, but also hold far-reaching implications for the fields of invasion biology and evolutionary ecology, promoting a deeper understanding of the complex challenges posed by invasive species.

Bibliography

- [1] Abdi, H., and Williams, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2, 4 (2010), 433–459.
- [2] Aebersold, R., and Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 7620 (2016), 347–355.
- [3] Alachiotis, N., and Pavlidis, P. Raisd detects positive selection based on multiple signatures of a selective sweep and snp vectors. *Communications biology* 1, 1 (2018), 79.
- [4] Alexander, D. H., Novembre, J., and Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19 (2009), 1655–1664.
- [5] An, Y., Wang, B., Yang, X., Lin, X., Chen, J., Huang, X., and Mastro, V. Characterizing populations of anoplophora glabripennis and related taxa with rapd. *Acta Entomologica Sinica* 47 (2004), 229–235.
- [6] Balkenhol, N., Dudaniec, R., Krutovsky, K., Johnson, J., Cairns, D., Segelbacher, G., Selkoe, K., von der Heyden, S., Wang, I., Selmoni, O., et al. Population genomics: Concepts, approaches and applications, 2017.
- [7] Batista, P. D., Janes, J. K., Boone, C. K., Murray, B. W., and Sperling, F. A. Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*dendroctonus ponderosae*). *Ecology and Evolution* 6, 16 (2016), 6292–6300.
- [8] Bilodeau, P., Roe, A. D., Bilodeau, G., Blackburn, G. S., Cui, M., Cusson, M., Doucet, D., Griess, V. C., Lafond, V. M., Nilausen, et al. Biosurveillance of forest insects: part II—adoption of genomic tools by end user communities and barriers to integration. *Journal of Pest Science* 92 (2019), 71–82.
- [9] Blackburn, G. S., Bilodeau, P., Cooke, T., Cui, M., Cusson, M., Hamelin, R. C., Keena, M. A., Picq, S., Roe, A. D., Shi, J., et al. An applied empirical framework for invasion science: confronting biological invasion through collaborative research aimed at tool production. *Annals of the Entomological Society of America* 113, 4 (2020), 230–245.
- [10] Blackburn, T., Pyšek, P., Bacher, S., Carlton, J., Duncan, R., Jarošík, V., Wilson, J., and Richardson, D. A proposed unified framework for biological invasions. *Trends in Ecology and Evolution* 26 (2011), 333–339.
- [11] Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., Turner, K. G., Whitney, K. D., and Rieseberg, L. H. What we still don't know about invasion genetics. *Invasion genetics: The baker and stebbins legacy* (2016), 346–370.

- [12] Bradshaw, C. J. A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., and Courchamp, F. Massive yet grossly underestimated global costs of invasive insects. *Nature Communications* 7, 1 (2016), 12986.
- [13] Brockerhoff, E., Liebhold, A., Richardson, B., and Suckling, D. Eradication of invasive forest insects: concepts, methods, costs and benefits. *New Zealand Journal of Forestry Science* 40, Suppl (2010), 117– 135.
- [14] Burnett, K., Kaiser, B., and Roumasset, J. Economic lessons from control efforts for an invasive species: Miconia calvescens in hawaii. *Journal of forest economics* 13, 2-3 (2007), 151–167.
- [15] Capinha, C., Essl, F., Seebens, H., Moser, D., and Pereira, H. The dispersal of alien species redefines biogeography in the anthropocene. *Science* 348 (2015), 1248–1251.
- [16] Carter, M., Smith, M., and Harrison, R. Patterns of genetic variation among populations of the Asian longhorned beetle (Coleoptera: Cerambycidae) in China and Korea. *Annals of the Entomological Society of America* 102 (2009), 895–905.
- [17] Charlesworth, B. Molecular population genomics: a short history. *Genetics research* 92, 5-6 (2010), 397–411.
- [18] Colautti, R. I., Grigorovich, I. A., and MacIsaac, H. J. Propagule pressure: a null model for biological invasions. *Biological invasions* 8 (2006), 1023–1037.
- [19] Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M., and Estoup, A. Diyabc v2. 0: a software to make approximate bayesian computation inferences about population history using single nucleotide polymorphism, dna sequence and microsatellite data. *Bioinformatics* 30, 8 (2014), 1187–1189.
- [20] Coyle, D. R., Trotter, R. T., Bean, M. S., and Pfister, S. E. First recorded asian longhorned beetle (coleoptera: Cerambycidae) infestation in the southern united states. *Journal of Integrated Pest Management* 12, 1 (2021), 10.
- [21] Cristescu, M. E. Genetic reconstructions of invasion history. *Molecular Ecology* 24 (2015), 2212–2225.
- [22] Csilléry, K., Blum, M. G., Gaggiotti, O. E., and François, O. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution* 25, 7 (2010), 410–418.
- [23] Cuthbert, R. N., Diagne, C., Haubrock, P. J., Turbelin, A. J., and Courchamp, F. Are the “100 of the world’s worst” invasive species also the costliest? *Biological Invasions* 24, 7 (2022), 1895–1904.

- [24] de Villemereuil, P., Frichot, É., Bazin, É., François, O., and Gaggiotti, O. E. Genome scan methods against more complex models: when and how much should we trust them? *Molecular ecology* 23, 8 (2014), 2006–2019.
- [25] Diagne, C., Leroy, B., Gozlan, R., Vaissière, A.-C., Assailly, C., Nuninger, L., Roiz, D., Jourdain, F., Jarić, I., and Courchamp, F. Invacost, a public database of the economic costs of biological invasions worldwide. *Scientific Data* 7, 1 (2020), 277.
- [26] Dodds, K. J., Hull-Sanders, H. M., Siegert, N. W., and Bohne, M. J. Colonization of three maple species by asian longhorned beetle, anoplophora glabripennis, in two mixed-hardwood forest stands. *Insects* 5, 1 (2013), 105–119.
- [27] Estoup, A., and Guillemaud, T. Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular ecology* 19, 19 (2010), 4113–4130.
- [28] FAO. International Standards for Phytosanitary Measures Integrated Measures, 2002.
- [29] Feng, Y., Tursun, R., Xu, Z., Ouyang, F., and Zong, S. Effect of three species of host tree on the cold hardiness of overwintering larvae of anoplophora glabripennis (coleoptera: Cerambycidae). *European Journal of Entomology* 113 (2016), 212–216.
- [30] Feng, Y., Xu, L., Li, W., Xu, Z., Cao, M., Wang, J., Tao, J., and Zong, S. Seasonal changes in supercooling capacity and major cryoprotectants of overwintering asian longhorned beetle (anoplophora glabripennis) larvae. *Agricultural and Forest Entomology* 18, 3 (2016), 302–312.
- [31] Feng, Y., Xu, L., Tian, B., Tao, J., Wang, J., and Zong, S. Cold hardiness of asian longhorned beetle (coleoptera: Cerambycidae) larvae in different populations. *Environmental Entomology* 43 (2014), 1419–1426.
- [32] Forster, B., Wermelinger, B., et al. First records and reproductions of the asian longhorned beetle anoplophora glabripennis (motschulsky)(coleoptera, cerambycidae) in switzerland. *Mitteilungen der Schweizerischen Entomologischen Gesellschaft* 85, 3/4 (2012), 267–275.
- [33] Gao, R., and Li, G. Review and prospect of research on anoplophora glabripennis. In *Proceedings, US Department of Agriculture Interagency Research Forum on Gypsy Moth and Other Invasive Species, 2000: January 16-19, 2001, Loews Annapolis Hotel, Annapolis, Maryland* (2001), vol. 285, US Department of Agriculture, Forest Service, Northeastern Research Station, p. 59.
- [34] Grgac, R., Rozsypal, J., Marteaux, L., Štětina, T., and Koštál, V. Stabilization of insect cell membranes and soluble enzymes by accumulated cryoprotectants during freezing stress. *Proceedings of the National Academy of Sciences* 119, 41 (2022), 2211744119.

- [35] Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS genetics* 5, 10 (2009), e1000695.
- [36] Haack, R., Cavey, J. F., Hoebeke, E. R., and Law, K. Anoplophora glabripennis: a new tree-infesting exotic cerambycid invades new york. *Newsletter of the Michigan Entomological Society* 41, 2-3 (1996), 1–3.
- [37] Haack, R. A., Bauer, L. S., Gao, R. T., McCarthy, J. J., Miller, D. L., Petrice, T. R., and Poland, T. M. Anoplophora glabripennis within-tree distribution, seasonal development, and host suitability in China and Chicago. *Great Lakes Entomologist* 39 (2006), 169–183.
- [38] Hamelin, R. C., and Roe, A. D. Genomic biosurveillance of forest invasive alien enemies: A story written in code. *Evolutionary Applications* 13 (2020), 95–115.
- [39] Hebert, P. D., Cywinski, A., Ball, S. L., and DeWaard, J. R. Biological identifications through dna barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, 1512 (2003), 313–321.
- [40] Hérard, F., Ciampitti, M., Maspero, M., Krehan, H., Benker, U., Boegel, C., Schrage, R., Bouhot-Delduc, L., and Bialooki, P. Anoplophora species in europe: infestations and management processes 1. *EPPO Bulletin* 36, 3 (2006), 470–474.
- [41] Holzapfel, E., and Harrell, J. Transoceanic dispersal studies of insects. *Pacific Insects* 10, 1 (1968), 115–153.
- [42] Hopkin, A., Scarr, T., and Marie, S. Status of important forest pests in ontario 2003. *Online: www. cfs. nrcan. gc. ca/index/foresthealth* (2003).
- [43] Hua, L., Li, S., and Zhang, X. Iconography of forest insects in hunan, 1992. 467.
- [44] Hulme, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of applied ecology* 46, 1 (2009), 10–18.
- [45] Hulme, P. E. Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth* 4, 5 (2021), 666–679.
- [46] Javal, M., Lombaert, E., Tsykun, T., Courtin, C., Kerdelhué, C., Prospero, S., Roques, A., and Roux, G. Deciphering the worldwide invasion of the Asian long-horned beetle: A recurrent invasion process from the native area together with a bridgehead effect. *Molecular Ecology* 28 (3 2019), 951–967.
- [47] Johnson, C. H., Ivanisevic, J., and Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology* 17, 7 (2016), 451–459.

- [48] Kalinowski, S. T. The computer program structure does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* 106, 4 (2011), 625–632.
- [49] Keena, M. Anoplophora glabripennis (coleoptera: Cerambycidae) fecundity and longevity under laboratory conditions: Comparison of populations from new york and illinois on acer saccharum. *Environmental Entomology* 31 (2002), 490–498.
- [50] Keena, M., and Sánchez, V. Reproductive behaviors of asian longhorned beetle. *Proceedings, Emerald Ash Borer Asian Longhorned Beetle Res. Technol. Dev. Meet. US Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team FHTET-2007-04, Cincinnati, OH* (2006), 123–124.
- [51] Keena, M. A. Effects of temperature on anoplophora glabripennis (coleoptera: Cerambycidae) adult survival, reproduction, and egg hatch. *Environmental Entomology* 35, 4 (2006), 912–921.
- [52] Keizi, K., and Yamamura, K. Exotic insects and their pathways for invasion. In *Invasive species: vectors and management strategies*. Island Press, 2003, p. 44–67.
- [53] Kethidi, D., Roden, D., Ladd, T., Krell, P., Retnakaran, A., and Feng, Q. Development of scar markers for the dna-based detection of the asian long-horned beetle, anoplophora glabripennis (motschulsky). *Archives of Insect Biochemistry and Physiology* 52 (2003), 193–204.
- [54] Kimura, M. *The neutral theory of molecular evolution*. Cambridge University Press, 1983.
- [55] Kowarik, I., et al. Time lags in biological invasions with regard to the success and failure of alien species. *Plant invasions: general aspects and special problems* (1995), 15–38.
- [56] Lawson, D. J., Van Dorp, L., and Falush, D. A tutorial on how not to over-interpret structure and admixture bar plots. *Nature communications* 9, 1 (2018), 3258.
- [57] Lee, S., Lee, Y., and Lee, S. Population genetic structure of anoplophora glabripennis in south korea: Invasive populations in the native range? *Journal of Pest Science* 93 (2020), 1181–1196.
- [58] Lee Jr, R. E. Principles of insect low temperature tolerance. In *Insects at low temperature*. Springer, 1991, pp. 17–46.
- [59] Leung, B., Lodge, D., Finnoff, D., Shogren, J., Lewis, M., and Lamberti, G. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269, 1508 (2002), 2407–2413.

- [60] Lingafelter, S., and Hoebeke, R. *Revision of the Genus Anoplophora (Coleoptera: Cerambycidae)*. Entomological Society of Washington, Washington, DC, 2002.
- [61] Lowe, S., Browne, M., Boudjelas, S., and Poorter, M. *100 of the world's worst invasive alien species: a selection from the global invasive species database*, vol. 12. Invasive Species Specialist Group Auckland, 2000.
- [62] Luque, G., Bellard, C., Bertelsmeier, C., Bonnaud, E., Genovesi, P., Simberloff, D., and Courchamp, F. The 100th of the world's worst invasive alien species. *Biological Invasions* 16, 5 (2014), 981–985.
- [63] Mackay, T. F., Stone, E. A., and Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10, 8 (2009), 565–577.
- [64] Manel, S., Gaggiotti, O. E., and Waples, R. S. Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution* 20 (2005), 136–142.
- [65] McKenna, D. D., Scully, E. D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S. M., Mitchell, R. F., Waterhouse, R. M., Ahn, S.-J., Arsala, D., et al. Genome of the asian longhorned beetle (anoplophora glabripennis), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome biology* 17, 1 (2016), 1–18.
- [66] Memmott, J., Craze, P., Harman, H., Syrett, P., and Fowler, S. The effect of propagule size on the invasion of an alien insect. *Journal of Animal Ecology* 74, 1 (2005), 50–62.
- [67] Meng, P. S., Hoover, K., and Keena, M. A. Asian longhorned beetle (Coleoptera: Cerambycidae), an introduced pest of maple and other hardwood trees in North America and Europe. *Journal of Integrated Pest Management* 6 (2015), 1–13.
- [68] Meurisse, N., Rassati, D., Hurley, B., Brockerhoff, E., and Haack, R. Common pathways by which non-native forest insects move internationally and domestically. *Journal of Pest Science* 92 (2019), 13–27.
- [69] Morimoto, N., Kiritani, K., Yamamura, K., and Yamanaka, T. Finding indications of lag time, saturation and trading inflow in the emergence record of exotic agricultural insect pests in japan. *Applied Entomology and Zoology* 54, 4 (2019), 437–450.
- [70] Motschulsky, V. d. Diagnoses de coleopteres nouveaux, trouves par mm tararinoff et guschkewitsch aux environs de pekin. *Etud. ent.* 2 (1853), 44–51.
- [71] Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A. G. Recent and ongoing selection in the human genome. *Nature Reviews Genetics* 8, 11 (2007), 857–868.

- [72] Nowak, D., Pasek, J., Sequeira, R., Crane, D., and Mastro, V. Potential effect of anoplophora glabripennis (coleoptera: Cerambycidae) on urban trees in the united states. *Journal of Economic Entomology* 94 (2001), 116–122.
- [73] Poland, T., and Rassati, D. Improved biosecurity surveillance of non-native forest insects: a review of current methods. *Journal of Pest Science* 92 (2019), 37–49.
- [74] Porth, I., Klápště, J., McKown, A. D., La Mantia, J., Guy, R. D., Ingvarsson, P. K., Hamelin, R., Mansfield, S. D., Ehlting, J., Douglas, C. J., et al. Evolutionary quantitative genomics of populus trichocarpa. *PloS one* 10, 11 (2015), e0142864.
- [75] Pritchard, J. K., Stephens, M., and Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* 155, 2 (2000), 945–959.
- [76] Pyšek, P., and Prach, K. Plant invasions and the role of riparian habitats: A comparison of four species alien to central europe on jstor. *Journal of Biogeography* (1993).
- [77] Pyšek, P., and Richardson, D. M. Invasive species, environmental change and management, and health. *Annual review of environment and resources* 35 (2010), 25–55.
- [78] Rizzo, D., Taddei, A., Lio, D. D., Bruscoli, T., Cappellini, G., Bartolini, L., Salemi, C., Luchi, N., Pennacchio, F., and Rossi, E. Molecular Identification of Anoplophora glabripennis (Coleoptera: Cerambycidae) From Frass by Loop-Mediated Isothermal Amplification. *Journal of Economic Entomology* 113, 6 (2020), 2911–2919.
- [79] Roden, D., Haack, R., Keena, M., McKenney, D., Beall, F., and Roden, P. Potential northern distribution of asian longhorned beetle in north america. *U.S. Department of Agriculture Interagency Research Forum on Invasive Species* (2009), 65–67.
- [80] Roe, A., Torson, A., Bilodeau, G., Bilodeau, P., Blackburn, G., Cui, M., Cusson, M., Doucet, D., Griess, V., Lafond, V., Paradis, G., Porth, I., Prunier, J., Srivastava, V., Tremblay, E., Uzunovic, A., Yemshanov, D., and Hamelin, R. Biosurveillance of forest insects: part i—integration and application of genomic tools to the surveillance of non-native forest insects. *Journal of Pest Science* 92 (2019), 51–70.
- [81] Roques, A., Auger-Rozenberg, M.-A., Blackburn, T., Garnas, J., Pyšek, P., Rabitsch, W., Richardson, D., Wingfield, M., Liebhold, A., and Duncan, R. Temporal and interspecific variation in rates of spread for insect species invading europe during the last 200 years. *Biological Invasions* 18, 4 (2016), 907–920.
- [82] Samarasekera, G. D. N., Bartell, N. V., Lindgren, B. S., Cooke, J. E. K., Davis, C. S., James, P. M. A., Krokene, P., and Coltman, D. W. Spatial genetic structure of the mountain pine beetle (dendroctonus ponderosae) outbreak in western canada: historical patterns and contemporary dispersal. *Molecular Ecology* 21, 12 (2012), 2931–2948.

- [83] Schröder, T., Hoyer-Tomiczek, U., Bögel, C., Schrage, R., et al. Asian longhorn beetle in germany. *AFZ/Der Wald, Allgemeine Forst Zeitschrift für Waldwirtschaft und Umweltvorsorge* 61, 16 (2006), 888–890.
- [84] Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., et al. No saturation in the accumulation of alien species worldwide. *Nature communications* 8, 1 (2017), 14435.
- [85] Shatz, A. J., Rogan, J., Sangermano, F., Ogneva-Himmelberger, Y., and Chen, H. Characterizing the potential distribution of the invasive asian longhorned beetle (anoplophora glabripennis) in worcester county, massachusetts. *Applied Geography* 45 (2013), 259–268.
- [86] Simberloff, D. How much information on population biology is needed to manage introduced species? *Conservation Biology* 17, 1 (2003), 83–92.
- [87] Smith, M., Bancroft, J., Li, G., Gao, R., and Teale, S. Dispersal of anoplophora glabripennis (cerambycidae). *Environmental Entomology* 30 (2001), 1036–1040.
- [88] Smith, M., Bancroft, J., and Tropp, J. Comparison of the reproductive potential of anoplophora glabripennis (motsch.) among host tree species. In *Proceedings, U.S. Department of Agriculture Interagency Research Forum on Gypsy Moth and Other Invasive Species 2001., Gen. Tech* (2001), p. 123–128.
- [89] Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., and Neale, D. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* 9 (2013), 901–911.
- [90] Stefan, M., Markham, C., Benjamin, R., and Coath, J. Case study. invasive insects in plant biosecurity: The asian longhorned beetle eradication program. In *The Handbook of Plant Biosecurity*. Springer, 2014, p. 485–517.
- [91] Storey, K. B., and Storey, J. M. Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology* 90, 4 (2012), 456–475.
- [92] Tomiczek, C., Hoyer-Tomiczek, U., et al. Asian longhorned beetle (anoplophora glabripennis) and citrus longhorned beetle (anoplophora chinensis) in europe-actual situation. *Forstschutz aktuell*, 38 (2007), 2–5.
- [93] Torson, A., Zhang, M., Ong, K., Mohammad, L., Smith, A., Doucet, D., Roe, A., and Sinclair, B. Cold tolerance of laboratory-reared asian longhorned beetles. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 257 (2021), 110957.

- [94] Toxopeus, J., Koštál, V., and Sinclair, B. Evidence for non-colligative function of small cryoprotectants in a freeze-tolerant insect. *Proceedings of the Royal Society B* 286, 1899 (2019), 20190050.
- [95] Toxopeus, J., Lebenzon, J., McKinnon, A., and Sinclair, B. Freeze tolerance of cyphoderis monstrosa (orthoptera: Prophalangopsidae). *The Canadian Entomologist* 148, 6 (2016), 668–672.
- [96] Toxopeus, J., and Sinclair, B. Mechanisms underlying insect freeze tolerance. *Biological Reviews* 93 (2018), 1891–1914.
- [97] Turgeon, J. J., Orr, M., Grant, C., Wu, Y., and Gasman, B. Decade-old satellite infestation of anoplophora glabripennis motschulsky (coleoptera: Cerambycidae) found in ontario, canada outside regulated area of founder population. *The Coleopterists Bulletin* 69, 4 (2015), 674–678.
- [98] Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics* 101, 1 (2017), 5–22.
- [99] Wan, F.-H., and Yang, N.-W. Invasion and management of agricultural alien insects in china. *Annual Review of Entomology* 61, 1 (2015), 1–22.
- [100] Wang, J. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular ecology resources* 17, 5 (2017), 981–990.
- [101] Wang, Z., Gerstein, M., and Snyder, M. Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* 10, 1 (2009), 57–63.
- [102] Weigand, H., and Leese, F. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society* 184, 2 (2018), 528–583.
- [103] Wu, Y., Trepanowski, N. F., Molongoski, J. J., Reagel, P. F., Lingafelter, S. W., Nadel, H., Myers, S. W., and Ray, A. M. Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in tradeassociated solid wood packaging material using DNA barcoding and morphology. *Scientific Reports* 7 (2017), 1–12.
- [104] Yan, J., Qin, X., and Xiao, G. Anoplophora glabripennis (motsch.). *Forest insects of China* (1992), 455–457.

The section above contains the references for the Introduction and Conclusion of the thesis.