European population genomic differentiation and dispersal pattern of the invasive beetle Anoplophora glabripennis

Dissertation to obtain the doctoral degree of Agricultural Sciences (Dr. sc. agr.)

Faculty of Agricultural Sciences

University of Hohenheim

Institute of Animal Science

submitted by

Iris Hanna Häussermann

from *Schorndorf*

2023

This thesis was accepted as doctoral dissertation at 2nd of September 2022 in fulfilment of the requirements for the doctoral degree of Agricultural Sciences (Dr. sc. agr.) by the Faculty of Agricultural Sciences at University of Hohenheim.

Date of the oral examination: 17th of October 2022

Examination Committee

Chairperson of the committee: Prof. Dr. Jörn Bennewitz

1st Examiner: Prof. Dr. Martin Hasselmann

2nd Examiner: Prof. Dr. Lars Krogmann

3rd Examiner: Prof. Dr. em. Claus P. W. Zebitz

Index

List of figures

List of tables

List of acronyms

ABC	approximate Bayesian computation
AIAI	anthropogenically induced adaptation to invade
ALB	asian long-horned beetle (*Anoplophora glabripennis*); Motschulsky 1853
AMOVA	analysis of molecular variance
Amp	Ampicillin
approx.	approximately
BIC	Bayesian Information Criterion
BLAST	basic local alignment search tool
BLE	Federal Office for Agriculture and Food (German: Bundesanstalt für Landwirtschaft und Ernährung)
BMEL	Federal Ministry of Food and Agriculture (German: Bundesministerium für Ernährung und Landwirtschaft)
BOLD	Barcode of Life Data System
CDS	coding sequence
ch.	chapter
CLB	chinese or citrus long-horned beetle (*Anoplophora chinensis*); Forster 1771
COI	Cytochrome oxidase subunit 1 (mitochondrial gene)
COII	Cytochrome oxidase subunit 2 (mitochondrial gene)
CTAB	Cetyltrimethylammonium bromide
ddH_2O	ultrapure water
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylendiamintetraacetat
e.g.	latin *exempli gratia*, for example
EPPO	European and Mediterranean Plant Protection Organization
EU	European Union
F_{ST}	fixation index, measure of population differentiation due to population structure, S = subpopulation, T = total population
GBS	Genotype-by-Sequencing
GDP	Gross Domestic Product
GTR	substitution model, General Time Reversible model
HKY	substitution model, Hasegawa-Kishino-Yano model
HTS	high throughput sequencing
HWE	Hardy-Weinberg equilibrium

i5K	initiative to guide, organize and analysis the sequencing of 5000 arthopod genomes; provide communitication and workshops for improved sequencing, assembly, annotation, and data management standards
IAS	invasive alien species
ID	identification
i.e.	latin *id est*, that is
Inc.	corporation
IPPC	international plant potection convention
IPTG	Isopropyl-ß-D-Thiogalactopyranosid
K2	substitution model, Kimura-2-parameter model
Kap.	chapter, German: Kapitel
Lat Long	latitude and longitude, geographical coordinates in decimal degrees
LB	lysogeny broth, culture medium
LfL	plant protection office of the German state Bavaria; German: Bayerische Landesanstalt für Landwirtschaft
LTS	long term support
LTZ	plant protection office of the German state Baden-Wuerttemberg; German: Landwirtschaftliches Technologiezentrum Augustenberg
MAF	minor allelle frequency
MRCA	most recent common ancestor
MOPS	3-(N-morpholino)propanesulfonic acid
mt	mitochondrial
My	million years
N/A	not available
NCBI	National Center for Biotechnology Information
nGBS	normalised Genotype-by-Sequencing
NGS	next-generation sequencing
NPPO	national plant protection organizations
NUMTs	nuclear mitochondrial-like sequence
NUPTs	nuclear plastid-like sequence
OD_{600}	optical density at 600 nm
ORF	open reading frame
OS	operating system
OXPHOS	oxidative phosphorylation
PCA	principal component analysis
pers. comm.	personal communication
PHID-Coleo	plant health identification of Coleoptera, BLE funded project

Pop-ID	population identification of collection site
RAD-seq	restriction site associated DNA sequencing
RE	restriction enzyme
rpm	revolutions per minute
RT	room temperature
SNP	single nucleotide polymorphism
SOC	super optimal broth with catabolite repression, culture medium
spp.	species (plural) from the same genus
T92	substitution model, Tamura-3-parameter model
T_A	annealing temperature
TBE	electrophoresis buffer containing Tris, boric acid and EDTA
TFB	transformation buffer
Tris	Tris(hydroxymethyl)aminomethan
UHO	University of Hohenheim
UV	ultraviolet light
v/v	volume per volume, ml in 100 ml
v.	version
VCF	variant call format
w/v	weight per volume, g in 100 ml
WGS	whole genome sequencing
WPM	wood packing material
WSL	Windows Subsystem for Linux
X-β-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

“Let's think the unthinkable, let's do the undoable. Let us prepare to grapple with the ineffable itself and see if we may not eff it after all.”

— Douglas Adams, Dirk Gently's Holistic Detective Agency

Anoplophora glabripennis, picture provided by Olaf Zimmermann (LTZ, Germany)

1 Summary

Biological invasions are among the most threatening processes in the world, endangering biodiversity, food chains and nutrient cycling. Anthropogenic activities (e.g. homogenized habitats, trade) are the main factors to facilitate the increasing rates of invasive alien species. In this study, the invasion of the **Asian long-horned beetle** (ALB, *Anoplophora glabripennis*) was examined. Its native distribution is eastern Asia (China, Korean peninsula), but by extensive trade, this beetle was introduced via wood packing materials to North-America (1996) and Europe (2001). ALB attacks healthy broadleaved trees (e.g. *Acer* spp.*, Salix* spp.*, Populus* spp.), which can become lethal due to larval feeding. In the EU this beetle is classified as a quarantine organism, hence eradication programs are conducted according to EU's regulations. **This study aims** to detect genetic differences and kinship between the European infestation sites in Germany, Switzerland and Italy, from which the introduction and dispersal patterns can be deviated. Therefore, mitochondrial (mt) DNA-markers of the Cytochrome oxidase subunits I and II genes (COI and II) were used (ch. A and B, Sanger sequencing), as well as genome wide single nucleotide polymorphisms (SNPs), which were obtained by a Genotype-by-sequencing (GBS) approach (ch. C, Illumina sequencing).

Chapter A: The 658 bp DNA marker COI is primarily used for specimen identification. This DNA marker was applied in multiple population genetic ALB studies, which enabled the comparison with global ALB and *Anoplophora* spp. COI-sequences from public databases (BOLD, NCBI). The COI-haplotypes (17 SNPs) found in the European dataset showed no geographic pattern. Some shared common haplotypes diminished the possibility to infer relationships between infestation sites. Only a few unique haplotypes within a small spatial scale could indicate some independent introductions and secondary dispersals. The global COI datasets revealed several limitations for its usage in population genetic research of ALB. Source populations cannot be found by COI-sequences with the indistinct population genetic pattern in the native range found in previous studies. This blurred appearance is caused by multiple haplotypes per site without geographic pattern, while the at most 1-2 haplotypes per site in Europe could therefore indicate population bottlenecks. The phylogeny with other *Anoplophora* spp. showed, that COI can be used to separate ALB from other species of this genus, but some morphological misidentifications which have been introduced into databases, strongly underline the importance of integrative approaches. **Chapter B:** During sequencing of a 1271 bp COI-II segment, indels, stop codons and extreme high nucleotide differences assume the presence of nuclear mitochondrial sequence copies in the nuclear genome (NUMTs), which was not reported for ALB so far. Screening with BLAST revealed several alleged NUMTs in ALB's genome. Three of them were Sanger-sequenced to validate if they are genuine NUMTs and to determine the genetic diversity. All three NUMTs were proven to

be genuine with insertion times into the nuclear genomes millions of years ago. Typical for old NUMTs, signs of multiple copies were detected, which were most likely co-amplified. Hence, it could be assumed that the draft genome of ALB does not cover all existing NUMTs yet, suggesting higher numbers of undetected or undetectable NUMTs. The presence of NUMTs can be seen as a chance to improve the identification and phylogenetic approaches on ALB, by avoiding them in mtDNA marker amplification, as well as to develop a NUMT sequence DNA marker that includes genome and mitogenome evolution in one go. But, such a marker still has to be found, since the three sequenced ones were too disturbed by co-amplification of other NUMTs themselves. **Chapter C:** The genome-wide SNPs discovered with GBS detected several signs for small, bottlenecked and highly structured populations within the European dataset, like high inbreeding, reduced heterozygosity, a very slow decay of linkage disequilibrium (LD) beside a very low nucleotide diversity. The very high population differentiation, which was measured among most European populations is presumably derived from multiple independent introductions to Europe, which are spatially restricted in mating. By clustering analyses (PCA, Admixture, 7810 SNPs) and phylogeny (32432 SNPs), also some cases of secondary dispersal in a small spatial scale were confirmed. Additionally, some populations indicated admixture, which might have been originated by either multiple introductions from different sources or recurrent introductions from an admixed source population. Either way, these populations might be most capable to persist and establish, since propagule-pressure and maintained diversity can help overcome adverse effects of genetic bottlenecks. However, the populations showing strong founder effects most likely originated from the same introduction event and were shaped by genetic drift.

Conclusions: The results of this population genomic study of invasive European ALB populations showed very complex introduction patterns into Europe. Anthropogenically induced adaptation to invade (AIAI) is very likely, explaining why ALB is not much affected by hostile effects from genetic bottlenecks. In Europe, ALB does not face adaptive challenges, because the climate conditions and host trees are comparable with introduction sources from (sub)urban areas in Asia. **Outlook:** Future research on ALB's global invasion relies on a global cooperation framework, including large-scale sampling of natural populations from China and Korea to trace back the introduction routes and global invasion history. GBS generated SNPs could furthermore be compared with the closely related invasive Citrus long-horned beetle (CLB, *Anoplophora chinensis*) in Europe, to check for introgression. However, the advances in high-throughput sequencing methods and pioneer studies like these can strongly enhance phytosanitary measures and early responses in invasion science.

2 Zusammenfassung

Biologische Invasionen gehören zu den bedrohlichsten Prozessen der Welt, die Biodiversität, Nahrungsketten und Nährstoffzyklen gefährden. Anthropogene Aktiviäten (z.B. homogenisierte Habitate, Handel) sind die Hauptfaktoren, die ein erhöhtes Vorkommen invasiver Arten begünstigen. Diese Studie untersuchte die Invasion des **Asiatischen Laubholzbockkäfers** (ALB, *Anoplophora glabripennis*). Sein natürlicher Lebensraum liegt in China und der koreanischen Halbinsel und konnte durch intensiven Handel über Holzverpackungsmaterial nach Nordamerika (1996) und Europa (2001) eingeschleppt werden. ALB befällt gesunde Laubbäume (z.B. *Acer* spp.*, Salix* spp.*, Populus* spp.) was durch den Larvenfraß oft tödlich enden kann. In der EU gilt ALB als Quarantäneschaderreger, weshalb Tilgungsprogramme nach EU-Richtlinien durchgeführt werden. **Ziel dieser Arbeit** war es, genetische Unterschiede von Befallsgebieten in Deutschland, der Schweiz und Italien zu finden, um Verwandschaften und damit Einschleppungs- und Verbreitungswege abzuleiten. Dafür wurden mtDNA-Marker der Cytochromoxidase Untereinheiten I und II (Kap. A und B, Sanger Sequenzierung,) und durch Genotype-by-sequencing (GBS) gewonnene genomweite Einzelnukleotidpolymorphismen (SNPs) verwendet (Kap. C, Illumina Sequenzierung).

Kapitel A: Der COI DNA-Marker (658 bp) wird hauptsächlich zur Artbestimmung verwendet. Der häufige Einsatz in populationsgenetischen Studien des ALBs ermöglichte den Abgleich mit öffentlichen Datenbanken (BOLD, NCBI) globaler ALB und *Anoplophora* spp. COI-Sequenzen. Die COI-Haplotypen (17 SNPs) aus dem europäischen Datensatz zeigten kein geographisches Muster. Einige geteilte Haplotypen reduzierten mögliche Rückschlüsse auf Verwandschaft zwischen Befallsgebieten. Lediglich einzeln vorkommende Haplotypen deuteten auf unabhängige Einschleppunngen und sekundäre Verbringungen. Der globale COI Datensatz zeigte Nutzungslimitierungen des Markers in populationsgenetischen ALB Studien. Wegen unklaren geographischen Mustern im nativen Raum vorheriger Studien, konnten keine Originalpopulationen bestimmt werden. Während in Asien viele Haplotypen pro Ort vorhanden sind, liegen in Europa max. 1-2 Haplotypen pro Ort vor, was auf einen genetischen Flaschenhals hindeutet. Die Phylogenie mit anderen *Anoplophora* spp. zeigte, dass COI-Sequenzen zwar ALB von anderen *Anoplophora* Arten abgrenzen können, jedoch auch morphologische Fehlbestimmungen in Datenbanken integriert wurden, was die Wichtigkeit integrativer Ansätze untermauert. **Kapitel B:** Während der Sequenzierung eines 1271 bp COI-II Abschnitts ließen Indels, Stopcodons und hohe Nukleotidunterschiede die Anwesenheit von nukleären mitochondrialen Sequenzkopien (NUMTs) vermuten, was für ALB bisher nicht berichtet wurde. Über BLAST konnten mehrere vermeintliche NUMTs im ALB Genom gefunden werden. Drei davon wurden Sanger-sequenziert, um ihre Echtheit und genetische Diversität zu untersuchen. Alle drei wurden als Millionen Jahre alte, echte Insertionen ins

nukleäre Genom erkannt. Typisch für ältere NUMTs zeigten sich Anzeichen für mehrere Kopien, welche wahrscheinlich mit amplifiziert wurden. Es lässt sich vermuten, dass das vorläufige ALB Genom bisher nicht alle existierenden NUMTs abdeckt, was auf mehrere unendeckte oder undetektierbare NUMTs deutet. Das Vorhandensein von NUMTs kann als Chance gesehen werden, um Ansätze zur Identifizierung und Phylogenie zu verbessern, indem sie bei mtDNA-Marker Ansätzen vermieden werden, aber auch NUMT-Sequenz DNA-Marker, die mitochondriale und nukleäre Evolution abdecken, entwickelt werden können. Solche Marker müssen jedoch erst noch entdeckt werden, da die drei sequenzierten NUMTs selbst durch Co-Amplifikation von anderen NUMTs gestört wurden. **Kapitel C:** Die genomweiten SNPs die durch einen GBS Ansatz gewonnen wurden, fanden mehrere Anzeichen für kleine Populationen mit genetischen Flaschenhälsen und starker Populationsstruktur im europäischen Datensatz, wie z.B. hohe Inzucht, reduzierte Heterozygotie, eine langsame Abnahme des Kopplungsungleichgewichts und einer sehr niedrigen Nukleotiddiversität. Die gemessene starke Populationsstruktur ist vermutlich durch viele unabhängige Einschleppungen nach Europa entstanden, die sich räumlich getrennt nicht verpaaren konnten. Mit Cluster-Analysen (PCA, Admixture, 7810 SNPs) und Phylogenie (32432 SNPs), wurden ein paar Fälle sekundärer Verbringung bestätigt. Zusätzlich zeigten einige Populationen genetische Vermischung, die entweder von mehrfachen Einschleppungen diverser Quellen oder wiederholten Einschleppungungen aus genetisch vermischten Ursprungspopulationen stammen. Solche Populationen könnten sich leichter etablieren, da Nachschub und erhaltene Diversität helfen negative Effekte genetischer Verarmung zu umgehen. Die Populationen mit starkem Gründereffekt lassen jedoch auf eine einzelne Einschleppung und genetischen Drift schließen.

Schlussfolgerungen: Die Ergebnisse der populationsgenomischen Untersuchung invasiver ALB Populationen in Europa zeigen sehr komplexe Einschleppungswege auf. Eine anthropogen induzierte Adaption zur Invasion (AIAI) ist sehr wahrscheinlich und erklärt, wie ALB die schädlichen Effekte einer gentischen Verarmung umgehen kann. In Europa gibt es keine adaptive Herausforderung für ALB, da Wirtsbäume und klimatische Bedingungen den (sub)urbanen Gebieten aus Asien ähneln. **Ausblick:** Zukünftige Forschung der ALB Invasionen benötigt ein verlässliches Kooperations-Netz zur großflächigen Probenakquise aus natürlichen Populationen Chinas und Koreas, um die Einschleppungswege und globale Invasionsgeschichte zurückzuverfolgen. Um mögliche Introgression nachzuweisen, könnten durch GBS generierte SNPs mit dem verwandten invasiven Citrusbockkäfer (CLB, *Anoplophora chinensis*) in Europa verglichen werden. Fortschritte in Hochdurchsatz-Sequenzierungsmethoden und Pionierstudien wie diese können die pflanzengesundheitliche Praxis in der Invasionsbiologie unterstützen.

3 Introduction

This study is part of the federal collaborative project PHID-Coleo (plant health identification of Coleoptera) and was supported by funds of the Federal Ministry of Food and Agriculture (BMEL), based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme. The project aims to establish new methods for integrated identification, i.e. morphological and molecular identification of wood-boring beetles associated with wood packing materials (WPM) and therein for the early detection of potentially invasive species. The collaboration partners from the plant protection office of the German state Baden-Wuerttemberg (LTZ Augustenberg, Karlsruhe) had their focus on morphological identification keys and COI barcoding for interspecific characterization of species from the families Bostrichidae, Lyctidae and Cerambycidae. In addition, this study part described here focused on the intraspecific characterization of one of the most important invasive species worldwide of the Cerambycidae family, the Asian long-horn beetle (ALB) *Anoplophora glabripennis* (ch. 3.2).

3.1 Biological Invasions

In the last century, the occurrence of biological invasions has increased tremendously (Seebens et al. 2017) and they pose one of the major threats for biodiversity (Early et al. 2016), natural (Budde et al. 2016) and urban environments (Haack et al. 1996; Haack et al. 2010). A biological invasion process encompasses translocation, establishment and spread of a species into a new habitat with subsequent harmful effects on the invaded ecosystem (Blackburn et al. 2011; Budde et al. 2016). The threats on the invaded ecosystems are severe and complex. They can range from altered functions in ecosystems, changed habitat structures to nutrient cycle shifts (Asner and Vitousek 2005; Hänel and Chown 1998; O'Dowd et al. 2003; Pyšek et al. 2020) or new invasions with other species can be facilitated due to the eventually weakened environment (O'Dowd et al. 2003). Some definitions used to describe biological invasions are defined in **Box 1**.

For the range expansion to their new habitat, invasive alien species (IAS) require dispersal properties, either by direct movement (animals) or indirect movement (plants). Other dispersal possibilities are transport by soil, water, or by human-mediated diversions. (Chown et al. 2015) Usually, species are restricted in a geographical range by natural dispersal barriers (Connallon and Sgrò 2018), but in the last centuries humans facilitated movements beyond their natural distribution (Hulme et al. 2008; Liebhold and Tobin 2008). Compared to natural invasions, those which are facilitated by anthropogenic factors occur in much higher rates, with a higher chance of secondary dispersal (Cristescu 2015; Early et al. 2016). Human-mediated biological invasions can happen for multiple different reasons, like e.g. disturbances of environments,

international trade, inadequate prevention measures, resource exploitation, urbanization, atmospheric pollution, climate change or public unawareness. The vectors of transport eliminate ecological and physical barriers and enable vast numbers of often multiple introductions of species beyond their natural barriers. (Cristescu 2015; Hulme 2021; Pyšek et al. 2020; Venette and Morey 2019)

Box 1. Definitions related to biological invasions (FAO 2021; Javal 2017)

Native species: a species encountered in their native geographical area, that does not extend beyond the geographical barriers and limits of natural dispersal

Introduction: intentional or unintentional displacement by anthropogenic actions of an individual or a population from the natural to a new area that cannot be reached naturally

Introduced species: presence of a species outside its natural geographic area linked to human activities, that does not necessarily expand their range

Alien species: a species outside its natural distribution (alien i.e. non-native)

Invasive species: introduced species which maintains and reproduces by its own in the invaded area that can spread and/or modify its ecosystem

Anthropogenic factors in biological invasions

Due to globalization and a growing world population, the frequency of appearances of biological invasions is irresistibly rising together with global trade of commodities and travel (Bullock et al. 2018; Hulme 2015; Seebens et al. 2017; **Figure 1**). For instance, Hulme (2021) compared the vast increase of the global number of alien species first records with the increase of Gross Domestic Product (GDP) representing the imports, showing merchandise imports have more than tripled, while occurrences of IAS increased almost 20-fold since the early 19th century. The tremendous increase of IAS in non-indigenous areas poses risks for human health, food supplies and economic losses by the necessary prevention and management measures. Additionally, IAS are among the common causes of recent species extinctions, just behind biological resource usage. (Bellard et al. 2016; FAO 2021; Hulme 2009; Pejchar and Mooney 2009; Pyšek and Richardson 2010) The multitude of introduction pathways range from intentional release of e.g. biological control agents, over accidental introductions e.g. via shipping containers, to human movement of infested material e.g. plants, wood-packing materials (Bullock et al. 2018; Gippet et al. 2019; Leal et al. 2010). Since a huge proportion of all global invasive alien species (IAS) is provided by invasive insects (Seebens et al. 2017), the following description of the phases and pathways are based on insects. During the process of human-mediated transport, the IAS pass through the phases departure, transport and arrival which are all influenced by human activities (Bullock et al. 2018).

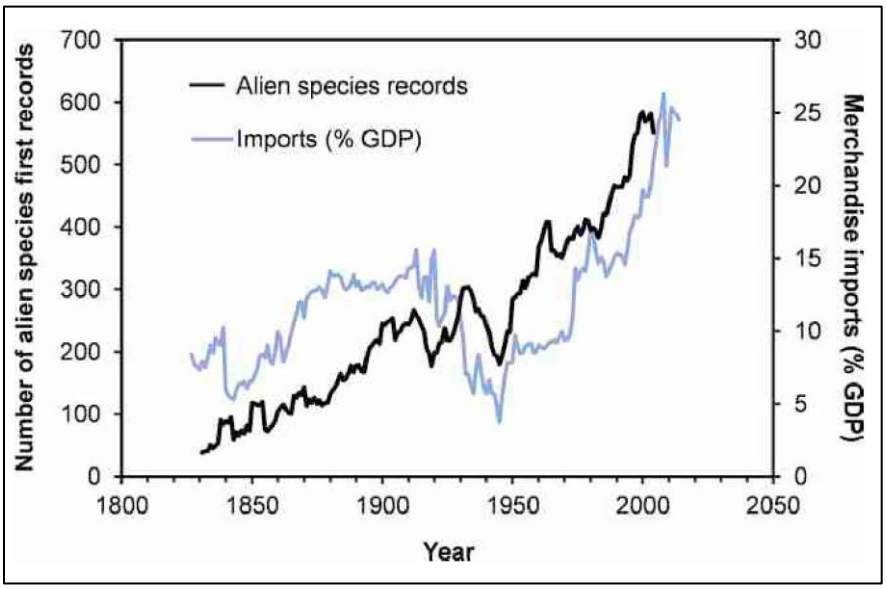


Figure 1. Comparison of imports and alien species first records over time (figure by Hulme 2021).
Changes since the early 19th century of Gross Domestic Product (GDP) representing the imports and the global number of alien species first records (Hulme 2021; Seebens et al. 2017).

For the departure, insects attach to a transport vehicle or are intentionally captured. The movement phase can be any vector, such as cars, trains, ships or airplanes. The last phase of human-mediated dispersal is the arrival when insects detach from their transport vehicle, when they are intentionally released or accidentally escape. (Gippet et al. 2019) The dispersal of insects facilitated by human activities can be either contamination, hitchhiking or harvesting (**Figure 2**). On the contamination pathway a commodity (e.g. plants, mammals, soil) is contaminated with any life stage of insects (Haack et al. 2010; Meurisse et al. 2019; Migliorini et al. 2015). With the hitchhiking pathway, insects attach to vehicles (e.g. shipping container, cars), which they use as shelter. The harvesting pathway represents intentional capture and transport, for instance as biological control agents. (Gippet et al. 2019; e.g. *Harmonia axyridis* Roy et al. 2012) Several factors of the human-mediated dispersal can impact the invasiveness and invasion success of the IAS. For instance, the quantity of the commodities as well as the traffic volume and frequency can determine the introduced population size and success by propagule pressure (Eritja et al. 2017; Simberloff 2009). The propagule pressure (influenced by individual number and release events) is higher in areas which are highly connected (Morel-Journel et al. 2019), hence transportation hubs can promote so called bridgehead effects (Bertelsmeier and Keller 2018; Bertelsmeier and Ollier 2021). In addition, the chances for a successful establishment in the new habitat are elevated, when the IAS is introduced multiple times (Simberloff 2009).

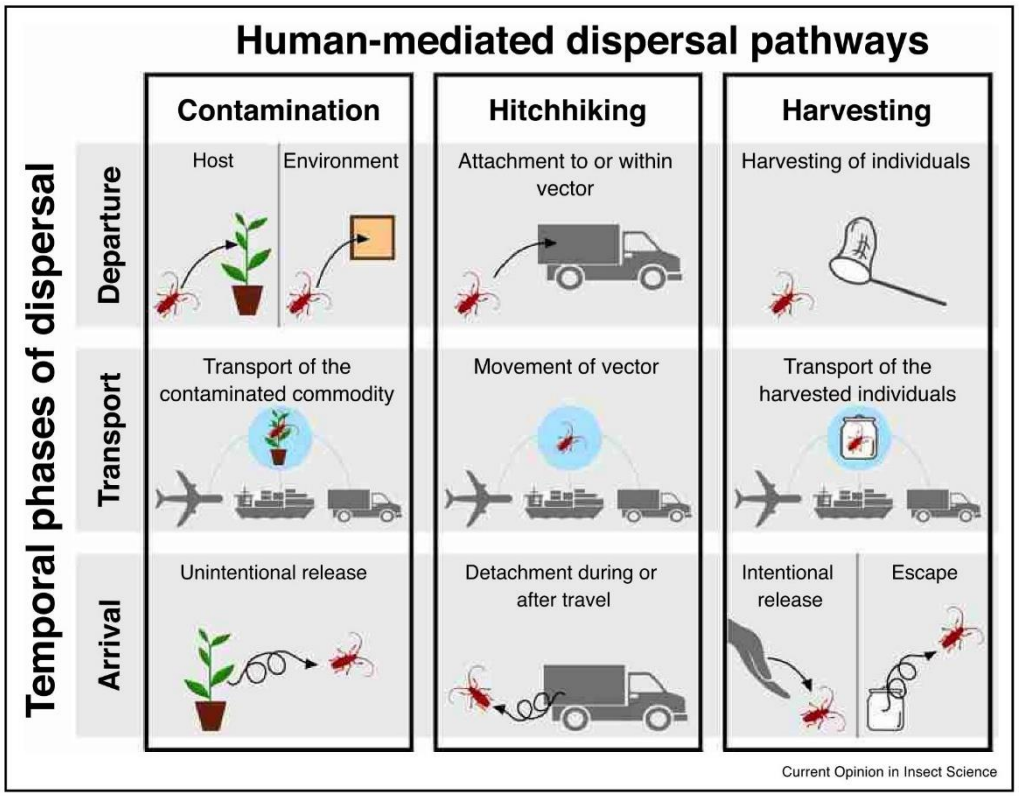


Figure 2. Phases of human-mediated dispersal (departure, transport and arrival) and invasion pathways (contamination, hitchhiking, harvesting) (figure by Gippet et al. 2019). **Accidential:** contamination and hitchhiking pathways, escape during harvesting pathway; **Intentional:** harvesting pathway (Gippet et al. 2019).

Phytosanitary measures antagonizing biological invasions

The major factor on a successful arrival for unintentionally introduced species is the effectiveness of phytosanitary treatments and inspections at ports of entry (e.g. fumigation, cold/heat treatments, irradiation) (Allen et al. 2017; ISPM 15, FAO 2019) and each incursion by alien species reveal occasional failed prevention measures (Pyšek et al. 2020). To secure healthy food supplies for a rising world population, resources and conservation of natural biodiversity, it is inevitable to national plant protection organizations (NPPO) to minimize the threat of IAS by state-of-the-art methods, preventing or at least impeding invasion processes. Therefore, substantial knowledge on biology, invasion history and genetic backgrounds of IAS, provides valuable information to develope effective control measures against all kinds of invasive species. (Roe et al. 2019) Within the process of a biological invasion, there are many points of actions along the way from transport, arrival, establishment and spread. These strategies range from prediction and prevention methods on the arrival, over methods on early detection, responses hindering an establishment, to treatments for mitigating further spreading of invaders. (Lampert and Liebhold 2021; Leung et al. 2002; Venette and Hutchison 2021) However, for all intervention strategies, the most important tool is a proper risk assessment

about invasive species that may develop a risk to plant production or the natural environment in invaded regions. Knowledge about their biology and the damage risk they pose supports prediction models, prevention measures, and at arrival or during early establishment, the knowledge where and how to look for to rapidly identify the species. (Mastin et al. 2020; Roe et al. 2019) For a globally successful biosurveillance, this information needs to be shared and communicated between NPPOs, to citizens and other research institutions (Roe et al. 2019). If strategies to predict and prevent a biological invasion have failed, the next step would be the localization of the new population as early as possible, and prompt eradication or containment (Venette and Morey 2019). One of the immediate reactions to biological invasions, are monitoring and intervention strategies. As soon as an invasive species is detected, attempts to eradicate initiating populations should be employed, which may differ according to regulations of an invaded region (e.g. in the EU when listed as quarantine pest; EU 2019). When the invasive species is irreversibly established, containment strategies need to be applied (EPPO 2013). Hence, research is needed to find the best methods to confirm an encountered invasive species. To support visual inspections, it is advantageous to deploy traps for monitoring if applicable (Venette and Hutchison 2021). The detection of already established populations in urban areas can also be supported by citizen science as assistance in monitoring, which was applied for the multicolored Asian lady beetle, *Harmonia axyridis,* in Europe (Roy et al. 2016) and for the Asian long-horned beetle, *Anoplophora glabripennis*, in the USA (Haack et al. 2010; Lingafelter and Hoebeke 2002).

Sequencing approaches in invasion science

An integrated approach including traditional taxonomy and genomic tools makes rapid determination of the species' identity more feasible. The advantage of genomic and molecular approaches is the possibility to distinguish differences within the species, enabling the assertion of the precise geographic origin of the invasive population, for instance as described with haplotype data for walnut twig beetles*, Pitophthorus juglandis* (Rugman-Jones et al. 2015) and brown marmorated stink bugs, *Halyomorpha halys* (Xu et al. 2014). The control agencies rely on fast and reliable tools for the precise identification of potential invasive species for an adequate management strategy (Boykin et al. 2012). The difficulties accompanied by morphological identification like a lack of expertises, specimen quality or morphological similarities (Thomas et al. 2016; Wu et al. 2017) can be resolved by supportive DNA sequencing approaches (Darling and Blum 2007). For best possible sustainable prevention and management strategies of invasive species, it is substantially important to address aspects like species identity, source populations, likelihood of colonization, invasiveness, introduction routes, dispersal patterns, ecological and evolutionary processes (e.g. adaption events) (Colautti and Lau 2015; Ibáñez et al. 2014; Roe et al. 2019). The identification of dispersal

pathways can be difficult when considering the extent and heterogeneity of transported commodities (Bilodeau et al. 2019), but also the complex evolutionary processes during biological invasions (Estoup et al. 2016) are challenging. **Box 2** describes some processes invasive populations pass through during invasion. Reconstructed dispersal patterns of invasive insects by genetic analyses exhibit that insect invasions encompass very complex invasion scenarios, including frequent jump dispersals, single or multiple introductions, back introductions into native range (Garnas et al. 2016) and also bridgehead effects (Bertelsmeier et al. 2018; Javal et al. 2019b; Lesieur et al. 2019; **Figure 3**).

Box 2. Main population processes during biological invasions (Javal 2017)

Bottleneck: During a biological invasion, the population size can be drastically reduced, hence only a small random sample of the source population is introduced into a new area. This can lead to a founder effect, where the frequency of originally rare and potentially deleterious alleles is increased, due to changes in allele frequencies among the sampled individuals (Dlugosch and Parker 2008).

Admixture: admixture is the genetic mixture of the same species by crossing of individuals from different source populations or lineages. Admixture events can happen either by human-mediated secondary displacement of introduced populations or if the individuals are capable to naturally disperse to another population (Garnas et al. 2016).

Bridgehead effect: This is a specific biological invasion scenario in which invasive populations are the source populations of other invasion events (Lombaert et al. 2010).

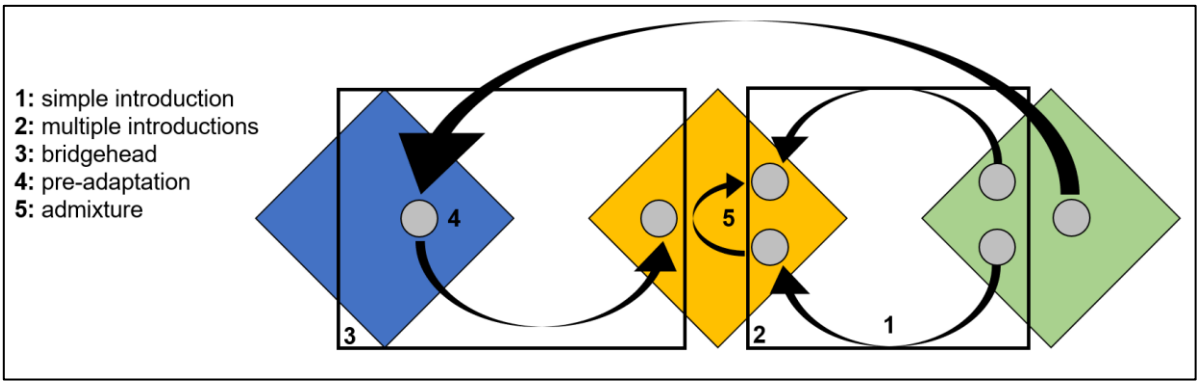


Figure 3. Summary of 5 different invasion scenarios (modified Javal 2017).
Two invasive (blue, yellow) and one native (green) geographic area are represented as diamonds. **1: simple introduction:** modified allele frequencies by genetic bottleneck; **2: multiple introductions**: several independent invasion events into a geographic area; **3: bridgehead:** an introduced population is the origin of the invasion in another geographic area; **4: pre-adaptation**: the invasive source population in a bridgehead event is adapted prior the secondary invasion; **5: admixture:** crossing of two invasive populations.
(Javal 2017)

Single introduction events can result in a drastic reduction of genetic diversity. However, this is strongly diminished by recurrent introductions which keep the propagule pressure high, enabling the introduced populations to establish and spread more likely. (Blackburn et al. 2015; Dlugosch and Parker 2008; Simberloff 2009) Multiple introduction events, which are more common than expected (e.g. Facon et al. 2003; Kelly et al. 2006), can ensure higher population sizes with an increased genetic diversity (Cristescu 2015), especially when the multiple introductions derived from different source populations (e.g. brown anole; Kolbe et al. 2004).

The genetic paradox of biological invasions

Invasive species are extraordinary in persisting and adapting to their new environment despite the probably depleted genetic variation by population bottlenecks (Cristescu 2015; Keller and Taylor 2010; Kolbe et al. 2004). The transient reduction of the population size (population bottleneck) during the process of introduction into a new environment and spatial expansion from there, can lead to reduced genetic variation (Allendorf and Lundquist 2003; Dlugosch and Parker 2008). The loss of genetic variation via inbreeding and genetic drift in the smaller founding populations (Founder-effect) can diminish the fitness of a population. However, many introduced populations can establish and even expand their range despite founder effects. (Allendorf and Lundquist 2003; Uller and Leimu 2011) The dilemma of invasion biology science, called genetic paradox of invasions, describes how bottlenecked invasive populations succeed despite their usually low genetic diversity (Allendorf and Lundquist 2003) by restoring evolutionary and reproductive potential, which is reviewed in Estoup et al. (2016). **Box 3** summarizes how non-native populations can be categorized as not paradox, genuine or spurious paradox.

Box 3. Genetic paradox of biological invasions (Estoup et al. 2016)

No genetic paradox: there is no loss of diversity by a shallow bottleneck which can be caused by multiple introductions or no adaptive challenge in the invaded area

Spurious genetic paradox: loss of genetic diversity is observable, but it is defined by a loss on neutral genetic markers rather than of ecological relevant traits, or the diversity loss is a result of successful selection

Genuine genetic paradox: there is a real loss of genetic diversity observable by genetic bottleneck, without negative consequences. This might happen when the bottleneck is beneficial due to purging of deleterious mutations, conversion of epistatic to additive variance, beneficial effects on particular traits, or when the adaptive potential is restored by *de novo* mutations, phenotypic plasticity or compensation by epigenetic processes

A genetic paradox population needs to have gone through a genetic bottleneck, which is determined by a lower genetic variation compared to the native source population. Despite the

bottleneck, introduced populations do not suffer from consequences of the founder effect and have successfully adapted to the conditions defined by the new habitat. Many invasive populations have a similar or greater genetic diversity rather than a lower diversity than native populations, or simply just do not encounter adaptive challenges in their new habitat. Hence, these populations are not genetic paradox. (Estoup et al. 2016) When a large number of individuals are introduced into a new habitat, the genetic diversity is higher than usually expected, since the introduced population can maintain the genetic diversity of the source population (Simberloff 2009; Uller and Leimu 2011). It is also possible to obtain a higher genetic diversity when there are multiple introduction events involved (Dlugosch and Parker 2008; Kolbe et al. 2004; Uller and Leimu 2011). When the multiple introductions derive from different native or invasive populations, the genetic diversity is considerably increased, as the genetic variation results from intraspecific admixture in the invasive range (Rius and Darling 2014). Admixture processes can result in completely new genotypes, which can be beneficial in local adaption and thereby for the colonization and expansion (Dlugosch and Parker 2008; Lavergne and Molofsky 2007; Rius and Darling 2014). Some species show a lower genetic diversity in the invaded area, but when the environment of the invaded and native range is similar, they do not face an adaptive challenge (Facon et al. 2006; Sax and Brown 2000). If the native range is characterized by fluctuating or heterogenous environments, the inhabitants usually sustain a high genetic diversity, hence these species may already be adapted in the new habitat they invade (Lee and Gelembiuk 2008). A special case of preadaptation is called anthropogenically induced adaptation to invade (AIAI) according to Hufbauer et al. (2012), which is induced by human-altered habitats in the native range. The concept of AIAI is parsimonious, because critical adaptions do not need to evolve multiple times (Estoup et al. 2016).

Some invasive populations just seem genetic paradox, because of inadequate measures for the genetic diversity, hence there is a spurious genetic paradox observable. This occurs, when neutral genetic markers show a diversity loss, but actual ecologically relevant traits do not. Another possibility of a spurious paradox is, when the low genetic variation reflects the results of successful adaptation. (Estoup et al. 2016) The usage of merely a few molecular markers with neutral genetic diversity, such as microsatellites, is not sufficient to predict the evolutionary potential of introduced populations, because they do not stand for the quantitative trait genetic variation (Pfrender et al. 2000; Reed and Frankham 2001). In general, quantitative traits are less affected by genetic bottlenecks (additive variance), than molecular markers during an invasion process e.g. by natural selection (Dlugosch and Parker 2008). Selective sweeps can also lead to a reduction of genetic variation in sequences near to a beneficial mutation (Smith and Haigh 1974). According to Robertson (1961), a low genetic variation can also be observed as a consequence of successful adaptation to new environments. When the effective

population size is reduced and homozygosity increased, some favorable individuals are selected for several subsequent generations, which is well known e.g. in animal breeding (Hill 2007). Hence, for multiple reasons the evolutionary potential to thrive in the new habitat can still be preserved by relevant traits, although the genetic diversity on neutral molecular markers is measured to be low (Bock et al. 2015; Dlugosch et al. 2015; Dlugosch and Parker 2008).

However, there are a lot of mechanisms of a genuine genetic paradox which actually help an invader to overcome deleterious consequences of a genetic bottleneck (Estoup et al. 2016). The reduction of genetic diversity due to bottleneck events or selection can in some cases improve the fitness of a population and thereby enhance the invasiveness of introduced species. This can be observed either by beneficial effects on ecologically relevant traits (e.g. red fire ant, *Solenopsis invicta* (Krieger and Ross 2002), increased invasiveness by purging of deleterious mutations, which enables high fitness even for inbred individuals (e.g. harlequin ladybird, *Harmonia axyridis* (Facon et al. 2011), or increased additive genetic variance in ecologically relevant traits during loss of total variation (e.g. moor frog, *Rana arvalis* (Knopp et al. 2007); pitcher-plant mosquito, *Wyeomyia smithii* (Armbruster et al. 1998)). Thereby, non-additive genetic effects (epistatic or dominance effects) are transformed into additive genetic variance, which allows selection to act on (van Heerwaarden et al. 2008; reviewed in Dlugosch and Parker 2008). New mutations usually hold a low frequency in the beginning, so if there are already existing beneficial mutations in an introduced population, it is asserted that these will have higher initial frequencies with increased probability of a fixation in the new population, which can be assumed during the timescale of an invasion (Barrett and Schluter 2008; Dlugosch et al. 2015; Hermisson and Pennings 2005). Also, deleterious mutations in the native range might be beneficial in the new environment and increase adaptive potential (Barrett and Schluter 2008). Other types of mutation can also contribute to restore adaptive potential in an invasion, like copy number variations (Dlugosch et al. 2015) and transposable elements (Stapley et al. 2015). Altered variation of gene expression and phenotypes due to epigenetic modifications as e.g. DNA methylation, histone modifications etc. allows organisms to promptly respond to environmental stressors (Dowen et al. 2012; Schrey et al. 2012). This might be an additional factor in new introduced populations to succeed with their invasion and could even help to overcome deleterious effects of the genetic bottleneck (Bock et al. 2015; Liebl et al. 2013). Also phenotypic plasticity may contribute to a better adaption despite a genetic bottleneck (Davidson et al. 2011).

To resolve this genetic paradox of invasions, more research is mandatory to optimize the genetic marker sets, and to include more factors in modeling as well as high throughput sequencing (HTS) and genotyping for further insights into the genetics behind invasions (Estoup et al. 2016). The usage of molecular markers covers taxonomic species identification,

source population identification, dispersal routes, secondary spread, observation of historical records and the ascertainment if single, multiple or successive introduction events and genetic bottlenecks were involved (Cristescu 2015; Roe et al. 2019). Anthropogenic driven invasions mostly have very complex introduction routes (Wilson et al. 2009), which makes them difficult to be reconstructed (Cristescu 2015), but by a reconstruction a better understanding of the evolutionary processes can be reached (Sax et al. 2007) and more information for reliable prediction models are gained (Kulhanek et al. 2011). In invasion science, often selectively neutral molecular markers, both nuclear (e.g. microsatellites) and mitochondrial markers, can be utilized for a detailed insight in the introduction patterns and source of introduction using population genetics and the genetic variation between the invading and the source populations (e.g. Cao et al. 2016; Carter et al. 2009b; Javal et al. 2019b), whereof multiple markers have a higher resolution (Cristescu 2015; Dupuis et al. 2012). Usually, dendrograms and networks are applied in a geographical context. For a detailed population genetic inference on population structure and demographic history by comparing the introduced populations with their corresponding native populations, source populations can be identified. This can be done e.g. with clustering of multilocus genotypes (Cristescu 2015) with methods provided by programs such as STRUCTURE (Pritchard et al. 2000), or ADMIXTURE (Alexander et al. 2009). All genetic and evolutionary changes need to be assessed against the original source populations to validate if the detected changes are only reflecting regional differences due to local adaption, drift and evolution, or if they are valid evolutionary changes between the source population and the sampled population (Dlugosch and Parker 2008; Keller and Taylor 2008). Methods based on approximate Bayesian computations (ABC) can test for the most likely invasion scenarios on few neutral genetic markers which were genotyped from native and invasive populations for the reconstruction of invasion routes (Beaumont et al. 2002; Cornuet et al. 2008), e.g. for *Anoplophora glabripennis* ABC methods were used by Javal et al. (2019b), Tsykun et al. (2019) and Lee et al. (2020). Nowadays, a range of genomic and transcriptomic approaches are adding information by the usage of genome wide DNA markers or whole genome sequencing with a much higher level of resolution and thereby a better understanding of the changes IAS populations undergoe during the invasion process. These approaches require high-quality data from multiple sources to trace invasion routes, source populations and adaption. (Chown et al. 2015; Roe et al. 2019) Genomics can be used to detect signatures of the stages of transport, colonization, establishment, spread and adaption during the invasion process by determining the genomic variation of invasive populations (Blackburn et al. 2011; Chown et al. 2015; Cristescu 2015; Roe et al. 2019). Thereby the present and past of populations can be documented in an unprecedented range (Chown et al. 2015). Early responses on a biological invasion event are the most cost-effective approaches to minimize risk and the inclusion of state-of-the-art genomic approaches improves the diagnostic process

enourmously (Chown et al. 2015; Roe et al. 2019). The advantages of genome-wide single nucleotide polymorphisms (SNPs) and the technology of Genotype-by-sequencing (GBS) in invasion science are decribed in more detail in ch. 3.5.

Invasive wood-boring beetles

This study focused on a cerambycid invasive wood-boring beetle which is introduced by WPM used in trade. This fact is of high importance in invasion science, since insects represent a large proportion of all IAS (Seebens et al. 2017). Insects are very easy transported accidentally by humans due to their small size (Meurisse et al. 2019). Juvenile stages of cerambycids can infest WPM such as pallets, crates, and dunnage in trading and only show little or no sign of infestation on the outside and therefore visual inspections alone are not sufficient to detect intercept infested materials (Eyre and Haack 2017). The international plant potection convention (IPPC) is regulating the global movement of WPM and requests the conduction of international phytosanitary measures (ISPM 15, FAO 2019) to reduce the introduction of wood-boring beetles (Wu et al. 2017). When WPM is not processed sufficiently, it becomes a pathway for cerambycid and other pests. The majority of the interceptions of long-horned beetles are related to WPM and are therefore of increased importance of national and regional plant protection agencies with expensive eradication programs. (Eyre and Haack 2017) In recent decades, repeated introductions of the cerambycid Asian long-horned beetle *Anoplophora glabripennis* (Motschulsky, 1853) have been of concern in many countries in Europe and North America. Most interceptions originated from WPM used in stone imports from China (Eyre et al. 2018; Haack et al. 2014).

3.2 The Asian long-horned beetle (Motschulsky, 1853)

The Asian long-horned beetle (ALB), *Anoplophora glabripennis* (Motschulsky, 1853), is a xylophagous wood-boring beetle from the family of the *Cerambycidae* (**Box 4**) with a wide range of host trees (Haack et al. 2010; Lingafelter and Hoebeke 2002). According to Lingafelter and Hoebeke (2002), the genus *Anoplophora* includes 36 species of wood-boring beetles occurring throughout Asia. Most of the species have remarkable colours on the elytra, pronotum, antennae, tarsi and venter. Likewise, the adult ALB beetles are also showing particular colours and maculations.

The imagines are typically cerambycid shaped, while males are smaller (19-32 mm long) than females (22-36 mm long) (Ric et al. 2006). The antennae of males are exceeding the body length a lot more than the female ones. The body of the beetle is shiny black with bluish colorings on antennomeres and tarsomeres. The elytral maculation is irregular-shaped in white or yellow (form *nobilis*), with a range of 0-60 dots (Lingafelter and Hoebeke 2002). All life stages (egg, larvae, pupa, imago) are depicted in **Figure 4**, even the morphotype *nobilis* (c).

Box 4. Taxonomic classification of the Asian long-horned beetle (CABI 2021)

Domain: *Eukaryota*
Kingdom: *Metazoa*
Phylum: *Arthropoda*
Subphylum: *Uniramia*
Class: *Insecta*
Order: *Coleoptera*
Family: *Cerambycidae*
Genus: *Anoplophora*
Species: *Anoplophora glabripennis*

A. glabripennis is native to East-Asia (China, Korean peninsula), and is considered as one of the most destructive invasive species worldwide (Lingafelter and Hoebeke 2002; Lowe et al. 2000). ALB was introduced and established outside of its native range in North America and Europe, which will be described in more details later in this chapter. In every infestation site this beetle has caused extensive economic and ecological damage (Haack et al. 2010). *A. glabripennis* has a wide host range of deciduous trees in urban environments, and in the native range also natural and planted forests. This polyphagous cerambycid can fulfill its complete life cycle on several species or genera, although *Acer* spp. appears to be the most appealing genus in the invasive range (Faccoli and Favaro 2016; Haack et al. 1997). However, the host range for *A. glabripennis* differs between native and invaded areas (Haack et al. 2010), which underlines the necessity to assess the actual host range in the invaded area. In the Asian native areas *A. glabripennis* infests healthy trees of the genera *Acer*, *Populus*, *Salix* and *Ulmus* (Haack 2006; Lingafelter and Hoebeke 2002; Wang et al. 2005; Williams et al. 2004), but *Populus* species and hybrids with *Aigeiros* are the major hosts (Hu et al. 2009). In the invaded areas in Europe and North America, the genera of broadleaf trees on which *A. glabripennis* can complete its life cycle comprise *Acer* spp., *Aesculus* spp., *Betula* spp., *Platanus* spp., *Populus* spp., *Salix* spp., *Sorbus* spp., and *Ulmus* spp. (CABI 2021; EFSA et al. 2019a; EFSA et al. 2019b; EPPO 2021a; Haack et al. 2006; Hérard et al. 2006). In Germany, *Acer* spp. and *Aesculus hippocastanum* have been the mainly infested hosts, though single damages of *Populus* spp., *Salix* spp. and very few of *Betula* spp. have been recorded as well (unpublished, Hoppe et al. in prep).

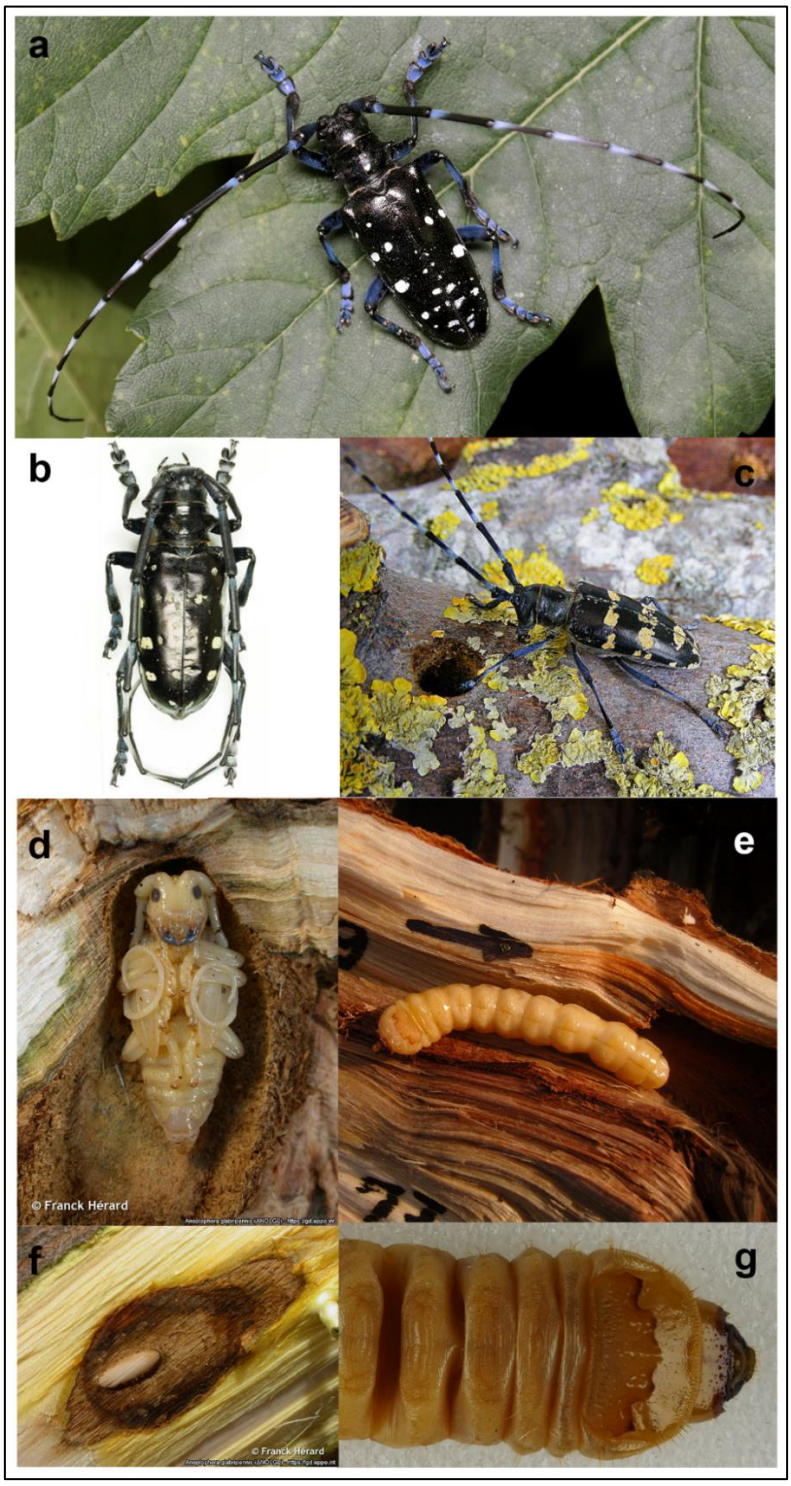


Figure 4. Life stages of the Asian long-horned beetle (ALB), *Anoplophora glabripennis*. Adult *A. glabripennis* beetles (a-c), pupa (d), egg (f) and larvae (e, g). Pictures provided by: **a, e)** Olaf Zimmermann (LTZ, Germany); **b, g)** Philipp Bauer (LTZ, Germany); **c)** Doris Hölling (WSL, Switzerland), *A. glabripennis* form *nobilis*; **d, f)** Franck Hérard (EPPO 2021b).

The duration of a generation of *A. glabripennis* depends on the climate conditions and nutrition. Generally, its life cycle takes one year, but sometimes it can take two to three years. In China,

for instance, the length of the lifecycle depends on the climate and latitude; the further north, the longer it takes to finish a lifecycle. For comparison, in Taiwan a generation takes only one year to develop. (reviewed in Lingafelter and Hoebeke 2002) For Europe, the same applies here, north of the Alps a generation takes two to three years, while in Italy it takes merely one year for a generation (EFSA et al. 2019b). Adult beetles of *A. glabripennis* can be found from April to December, but the main activities can be observed from May to July (Haack et al. 2010), whereas in Germany they mainly emerge in late July to August (Zimmermann, pers. comm.). On principle, the beetles stay for 10-15 days on the tree where they fulfilled their lifecycle, but some fly to neighbouring trees. The adults feed on twigs, petioles and veins of leaves. By contact or via short-range pheromones ALB beetles can find mating partners. (Lingafelter and Hoebeke 2002; Zhang et al. 2002; Zhang et al. 2003) The lifespan in the adult beetle stage is on average about one month (EPPO 2021a; Faccoli et al. 2015).

Approximately a week after mating, 30-32 eggs are deposited by females directly under the inner bark (phloem) via oviposition slits on the side of the trunk or in branches with >5 cm in diameter (EPPO 2021a). The females lay eggs typically in the upper trunk, along major branches or smaller branches in the upper canopy. The eggs hatch about two weeks later and the first two instar larva feed in the cambial layer and work their way into the sapwood as third instar larva. Near the oviposition site, the larva ejects frass (feces and sawdust) out of their tunnel. Frass from the second instar larvae that feed in phloem and xylem tissue is brown, while the late third and fourth instar larvae feed on xylem tissue and release whitish feces and sawdust. The maturing larva chews galleries moving up the trunk for 10-30 cm. The final instar larva builts a pupation chamber near the outer bark and transforms into a pupa, when the larva reaches a critical weight after overwintering. (reviewed in Lingafelter and Hoebeke 2002; Ric et al. 2006) The pupation process is induced in late spring or early summer inside the heartwood. After successful metamorphosis, the adult beetle exits from the tree via round holes above the oviposition with 10-15 mm in diameter. (EFSA et al. 2019b; Lingafelter and Hoebeke 2002) The observations in Germany were exactly 10 mm (Zimmermann, pers. comm.). After emerging, the adult beetles feed on tender bark of small twigs and branches, as well as on tender leaves and petioles which is called maturation feeding (Lingafelter and Hoebeke 2002). Differently to most of the other cerambycids, *A. glabripennis* attacks healthy, vigorous and stressed trees and even recently cut logs (reviewed in Lingafelter and Hoebeke (2002)). Often the adults return to the same tree from which they emerged to lay eggs again if the tree is still viable. The greatest damage is caused by the larval feeding inside branches and trunks. Even a small infestation may kill young host trees, heavier and successive infestations with several generations developing in one tree can weaken full-grown trees, so sometimes even strong winds or snow can cause branches to break or collapse. Eventually this process can become lethal to the host tree. (Lingafelter and Hoebeke 2002) Lethal infestations have been observed

in Germany as well (Zimmermann, pers. comm.; **Figure 5-a**). The visually detectable symptoms from the outside and the galleries within the trunk of an infestation tree are represented in **Figure 5**.

Phytosanitary measures to detect and control ALB infestations

The primary phytosanitary measures to prevent an infestation with ALB is applied on wood packing materials in compliance with the requirements of the ISPM 15 (FAO 2019). If ALB still could pass the surveillance on ports, the first step in the diagnostic process is the detection of the presence of *A. glabripennis* in any life stage by visual inspection of the trees that are at risk to be infested in the upper trunk, the main branches and the tree canopy (EFSA et al. 2019b; EPPO 2013; Haack et al. 2010). The inspections are conducted around crown level with binoculars, bucket trucks or tree climbers (**Figure 6-a**) (Haack et al. 2010). The symptoms of an *A. glabripennis* infestation can mostly be discovered about 1.5 m above the ground up to the canopy of the tree (EPPO 2013) and differ between the life stages (EFSA et al. 2019b; Ric et al. 2006). The oviposition slits are detected as transverse oval shaped pits on the bark with sap dripping off the fresh cut (Haack et al. 2010). The larval activities inside the tree can be observed although the larvae are hidden inside the wood (Haack et al. 1996; Haack et al. 2010) by the released frass (Rizzo et al. 2020; Taddei et al. 2021), **Figure 5-c** and **-d.** Wood shavings from the exit holes can be found at the tree base (EFSA et al. 2019b). The larval galleries under the bark or tunnels in the wood (**Figure 5-b**) cannot be observed in living trees, but can be detected on processed wood such as wood packing material, or can be traced by the sounds they make (Mankin et al. 2008). The symptoms of feeding from adult beetles on leaves and petioles (**Figure 5-f**) is only observable for a short time. The circular exithole of mature adults on the upper part of the trunk and main branches are also symptoms that can be detected by visual inspections for a longer time (**Figure 5-g**).

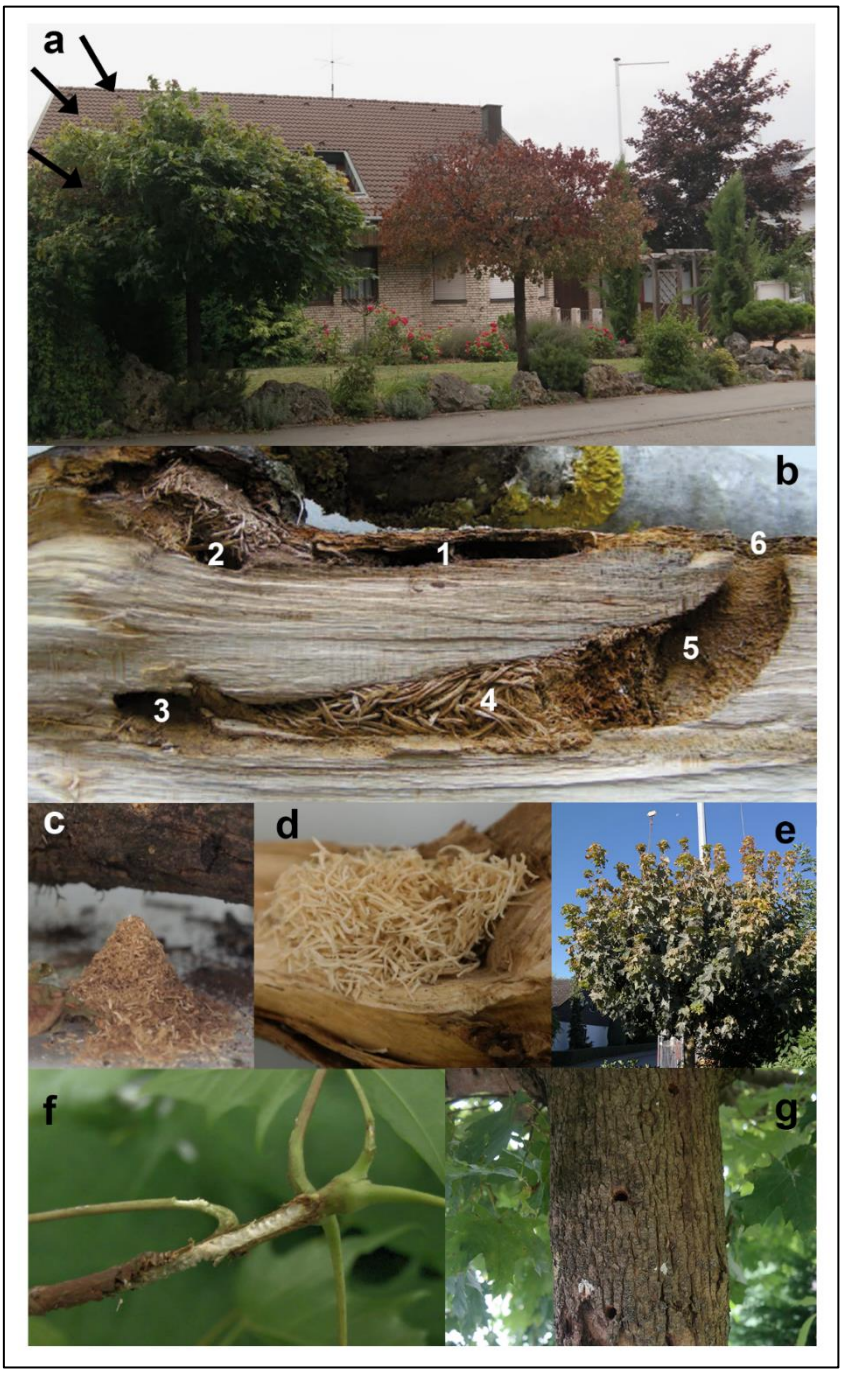


Figure 5. Symptoms and damage on host trees caused by ALB.
Different degrees of damage on host trees visible on the outside (a, c, d, e, f, g) and the inside (b). **a)** heavily infested trees (maple tree in the middle, already dead) can wither and die. Before an infestation becomes lethal, branches and the canopy of the tree start to wither (maple tree on the left, arrows). When infested trees are not robust anymore, neighboring vigorous trees are more attractive for new ovipositions (maple tree on the right). **b)** infestation of a maple tree; a young larva is feeding under the bark (**1**), works its way into the sapwood (**2**), is building larval galleries moving up the trunk (**3**) ending in a pupation chamber (**5**) that is plugged with shavings (**4**). After metamorphosis, the beetle exits the tree via a round hole (**6**). **c)** brown frass from second instar larvae; **d)** whitish frass from late third and fourth instar larvae; **e)** withering canopy of a maple tree; **f)** maturation feeding symptoms; **g)** round exit holes. Pictures provided by: **a, c-g)** Olaf Zimmermann (LTZ 2016, Germany), Hildrizhausen, Baden-Würrtemberg, Germany; **b)** Doris Hölling (Wermelinger et al. 2015, WSL Switzerland)

Heavily infested trees can show wilting foliage, crown discoloration, withering on branches and deformed bark (**Figure 5-a** and **-e**) (EFSA et al. 2019b). Other detection methods are also used to promote the monitoring process, like the application of sniffer dogs (Hoyer-Tomiczek et al. 2016, **Figure 6-e**) and pheromone traps (**Figure 6-d**) which are widely used to catch adult beetles in monitoring or infested areas (EFSA et al. 2019b; Nehme et al. 2014). Some traps are frequently used on sites with a higher risk, e.g. those with close proximity to stone importers for early detection (Regione Lombardia 2022). When symptoms are visible on the crown, the infestation has already proceeded at least 3-4 years (EPPO 2021a). The NPPOs must justify the expensive eradication programs, so they are dependent on reliable and rapid identification methods. But a reliable identification of a tree infested with *A. glabripennis* is not possible using only visual examinations. Thus, morphological, and molecular analyses are used for the identification. For juvenile specimens like eggs, larvae or pupae, obtaining specimens is usually only possible after chopping the infested tree (**Figure 6-b** and **-c**). Newer methods such as the real-time PCR method of Taddei et al. (2021) or the loop-mediated isothermal ampification (LAMP) approach of Rizzo et al. (2020) remain an exception as they can detect DNA of ALB in a non-destructive way in expelled frass. Late instar larvae and adults of *A. glabripennis* are best for the morphological identification, since there are several identification keys available (e.g. Lingafelter and Hoebeke 2002; Ric et al. 2006; PHID Coleo morphological key, unpublished). Pennacchio et al. (2012) provides a morphological identification key which is able to distinguish *A. glabripennis* larvae from the closely related *A. chinensis*, the Citrus Long-horned beetle (CLB). For standardized molecular identification according to EPPO Standard PM 7/129 (EPPO 2016), DNA barcoding with primers LCO1490 and HCO2198 of Folmer et al. (1994) and comparison with databases (BOLD, NCBI) are used, utilizing the partial coding sequence (CDS) of Cytochrome oxidase subunit 1 (COI). The early detection of Asian long-horned beetle is crucially facilitated by the commitment of residents, e.g. in Italy as described in Ciampitti and Cavagna (2014) for the Lombardia region. Many reports were from citizens, not from biosurveillance (Haack et al. 2010) like, for instance, the first identified infestation with ALB in New York city 1996 and additional infestations in Chicago 1998 (Lingafelter and Hoebeke 2002; Poland et al. 1998). In Germany among other information campaigns, signs as shown in **Figure 6-f** (Hildrizhausen) and **-g** (Weil am Rhein) were deployed to rise awareness of citizens in infested areas.

Control measures in the native areas from China differ from those applied in the invaded areas. There, the main approach lies in the application of insecticides, trap trees, or insect-pathogenic nematodes and preferred planting of resistant poplar hybrids. In the invaded areas of North America and Europe the main approach is eradication, which comprised the long process of long-term monitoring and destroying of infested trees. If an eradication is no longer possible, counter measures aim for containment, e.g. restrictions for wood transport outside certain

areas. In the EPPO region, the regularities for monitoring, eradication and containment of *A. glabripennis* is specified in EPPO Standard PM 9/15(1). (EPPO 2013, 2021a) According to the standards of the European and Mediterranean Plant Protection Organization (EPPO 2021c), *A. glabripennis* is registered in the A2 list (A2/296) for pests recommended for regulation as a quarantine pest, which is present in the EPPO region (52 member countries in Europe and the mediterranean area). Additionally, ALB is listed as quarantine pest in Annex II of the European Union (EU) commission implementing regulation 2019/2072 (EU 2019) as a quarantine pest with severe economic, environmental and social impact inside the EU territory. Respective phytosanitary measures for eradication, prevention of introduction and secondary spread are embedded in EU decision 2015/893 (EU 2015). Once an infestation is identified, regulations in this decision determine that all infested wood from host species, as well as infested wooden products, must be immediately destroyed. This includes immediate felling of infested trees and trees with symptoms, as well as chipping and burning of the wood. In addition, neighbouring, specified host plants must be removed in a radius of 100 m as a preventive measure. After this, demarcated areas that comprise the infested zone and a buffer zone with a radius of at least 2 km beyond the boundaries of the infested zone are defined. These demarcated areas must be at least annually monitored intensively and can only be lifted if there is no organism detected during four consecutive years. Additionally, the EU decision recommends raising public awareness about the threat of ALB and of the measures conducted.

After deep insight into the biology of *A. glabripennis*, the symptoms caused by an infestation, as well as the phytosanitary measures, the invasion of ALB in North America and Europe have only been broached. In the next section, the origin of the global spread, interceptions and especially the invasions in other areas of the world and where counter measures achieved eradication are addressed. This encompasses crucial information to link research efforts which already suceeded in practice or which will enhance the practice in the future. Naturally *A. glabripennis* is not moving further than 300 m (Favaro et al. 2015), as they can also finish their maturation feeding on the same tree where they emerged (Lingafelter and Hoebeke 2002). Rarely adults fly further than 2 km in natural conditions (Favaro et al. 2015) and they could fly up to 14 km in experimental conditions over their whole life span (Javal et al. 2018; Lopez et al. 2017). Thus, human mediated transportation is the only way how this organism could disperse (**Figure 7-a**) from their native origin in East-Asia (China, Korea) as evidently shown by its current extension to much more distant regions. It is known that mostly wood packing material and dunnage from host species used in trade containers (**Figure 7-b** and **-c**) are the common transport vehicles, as they can inhabit the life stages egg, larvae or pupae. Individual larvae or beetles (**Figure 7-d** and **-e**) have also been detected in wood packing material or living plants (e.g. bonsais). (EPPO 2013)

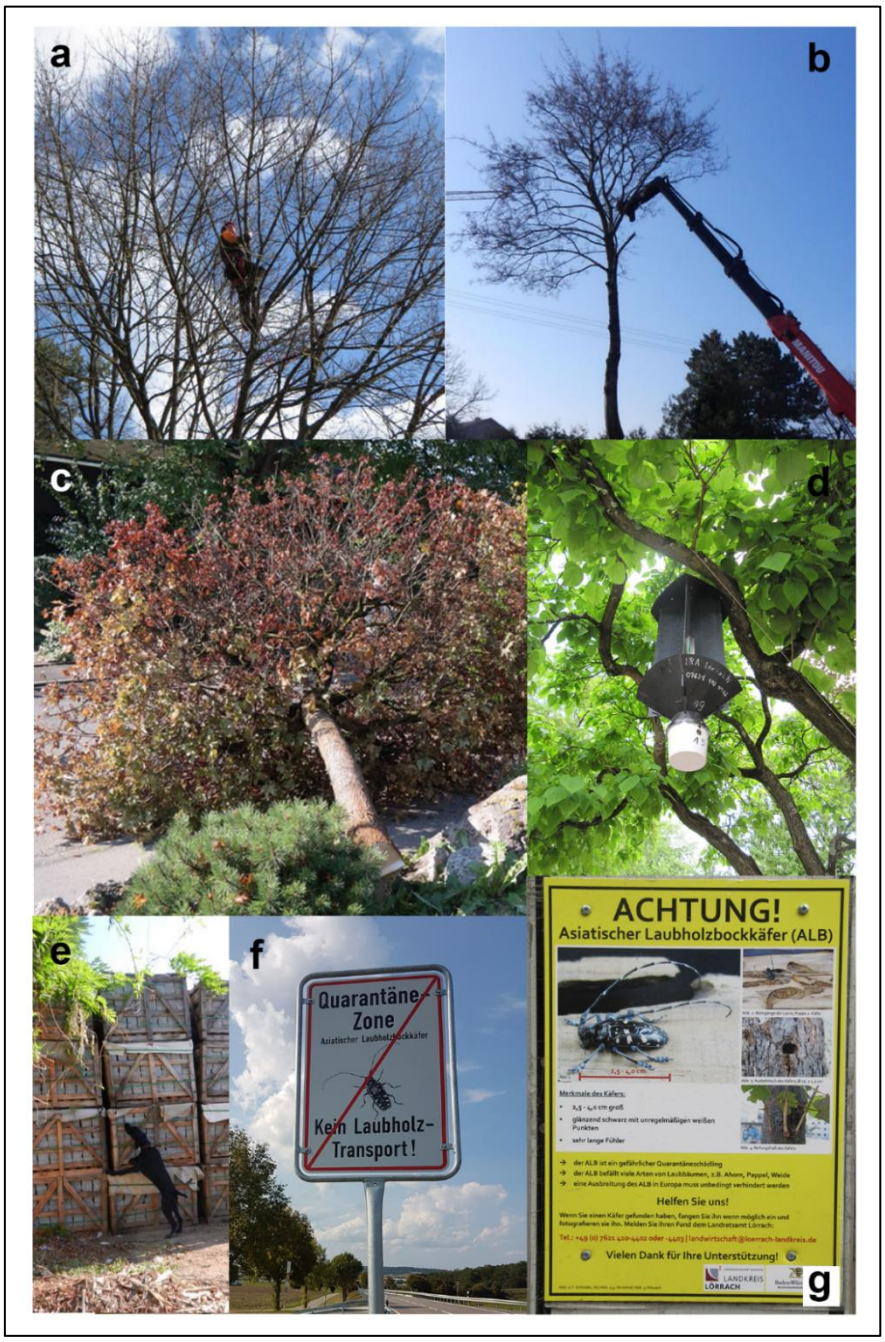


Figure 6. Management of invasive ALB populations in Germany.
a) visual inspection in the canopy by tree climbers; **b, c)** felling of dead infested trees; **d)** pheromone trap; **e)** sniffer dogs; **f, g)** information and warning signs for local residents; Pictures provided by: **a-c, e)** LTZ (2016) **a)** plant protection service of the German state Bavaria (LfL); **b)** Hermann Meier (RP Stuttgart) **c, e, f)** Olaf Zimmermann (LTZ) in Hildrizhausen (Germany, Baden-Württemberg); **d, g)** Birgit Gessler (University of Hohenheim) in Weil am Rhein (Germany, Baden-Württemberg).

Global distribution and invasion of *A. glabripennis*

A. glabripennis is indigenous in East Asia (China, Korea) and occurs in most Chinese provinces and thereby populates a vast geographic and climatic range. There have been some records from Japan, but it is considered as eradicated in this territory (2005). *A. glabripennis* has been introduced outside of its native range in North America and Europe (**Figure 7-a**; reviewed in Haack et al. 2010; Lingafelter and Hoebeke 2002), but even within the native range two of the colonized areas in South Korea were found to be introduced and invasive populations originating from China (Lee et al. 2020). In China, within the afforestation program 'Great green wall' campaign initiated by the Chinese government, vast numbers (millions) of hybrid poplars were planted over the last decades (starting in the 1960s) in many areas of the country along hedge rows and roads, in plantations, as agricultural shelter belts, in city parks and along Gobi desert to prevent desertification (Haack et al. 2010; Lingafelter and Hoebeke 2002; Wang et al. 2010). Some of the hybrids used for this purpose were very susceptible to *A. glabripennis* and were seriously damaged by this pest. The susceptibility of these artificial monocultures of poplar hybrids led to severe outbreaks and enormous proliferation of *A. glabripennis*, hence this beetle became a common pest in many regions of China. Some locals even named this beetle 'the forest fire without smoke'. (Lingafelter and Hoebeke 2002) Since the 1980s, new plantations of poplars were planted with only resistant hybrids and ALB turned into a less important pest in China again. *A. glabripennis* is not condsidered as a pest in the natural forests of China (EPPO 2021a) and Korea (Williams et al. 2004).

Through international trading, this beetle has been accidently introduced to North America (USA, Canada) and Europe on wood packing materials and is frequently intercepted worldwide (EPPO 2021b; Eyre et al. 2018; Haack et al. 2010), **Figure 7-b** to **-e**. In North America, *A. glabripennis* was first detected in 1996 in New York City, USA (Haack et al. 1996). To date (March 2022), ALB is currently present in Massachusetts (2008), New York (Long Island), Ohio (2011) and South Carolina (Coyle et al. 2021; EPPO 2021b). The latest established population was found in Hollywood in South Carolina (2020) and was estimated to be at least seven years old (Coyle et al. 2021). Due to extensive monitoring and phytosanitary measures, *A. glabripennis* was officially eradicated in Illinois (1998-2008) and New Jersey (2002-2008), as well as in New York City, parts of Massachusetts and parts of Ohio. In the USA, incursions in California and Washington states have been reported, but the pest did not establish there. (Coyle et al. 2021; EPPO 2021b) The first outbreak of *A. glabripennis* in Canada 2003 was in Ontario near Toronto (Turgeon et al. 2015). A second infestation occurred in the city of Mississauga (near Toronto) in 2013. The pest was officially declared as eradicated in Missisauga and Toronto and thereby from whole Canada after years of surveys without detection of this beetle. (EPPO 2021a, 2021b)

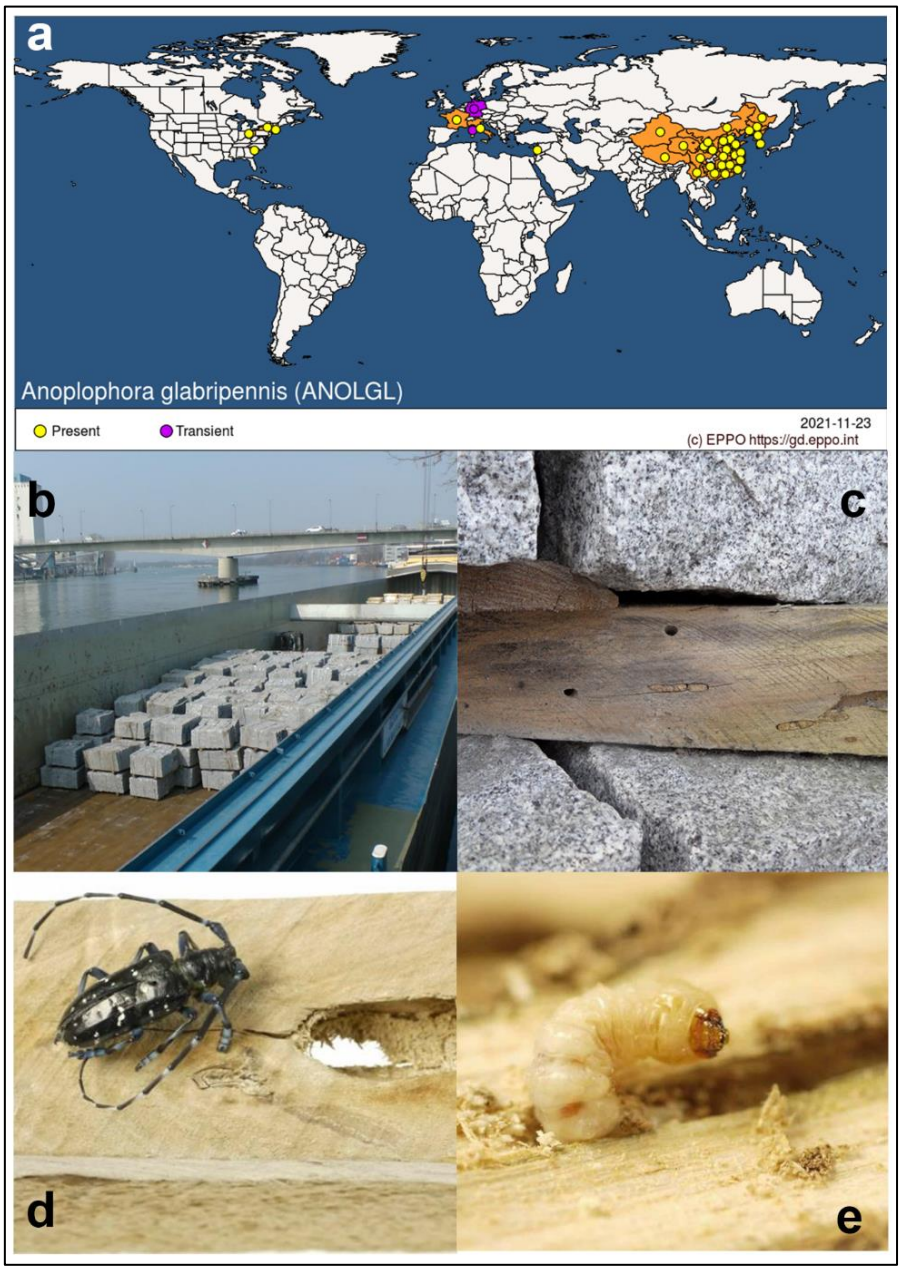


Figure 7. Global distribution of *A. glabripennis* and dispersal by trade associated WPM (March 2022).
a) global distribution of ALB showing countries or regions where this species is present (yellow) or transient (pink) at 2021-11-23 (EPPO 2021b): **EPPO Region:** France (mainland, Corse), Germany, Italy (mainland); **Asia:** China (Anhui, Fujian, Gansu, Guangdong, Guangxi, Guizhou, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Neimenggu, Ningxia, Qinghai, Shaanxi, Shandong, Shanxi, Sichuan, Xinjiang, Xizhang, Yunnan, Zhejiang), Korea Dem. People's Republic, Korea Republic, Lebanon; **North America:** United States of America (Massachusetts, New York, Ohio, South Carolina) (EPPO 2021a); Pictures provided by: **b-c)** Olaf Zimmermann (LTZ 2016; Germany); **d-e)** Matthias von Wuthenau (LTZ 2013)

The first found infestation with *A. glabripennis* in Europe was detected in Austria in 2001 (Hérard et al. 2006). New interceptions were frequently reported in other European countries since then (Belgium in 2008, Finland in 2015, France in 2003, Germany in 2004, Italy in 2007,

Montenegro in 2015 (Pajović et al. 2017), the Netherlands in 2010 (Loomans et al. 2013), Switzerland in 2011 (Forster and Wermelinger 2012) and the United Kingdom in 2011 (Eyre and Barbrook 2021) (EPPO 2021b). *A. glabripennis* has also been detected in Lebanon (EPPO region) in 2015 and 2016 (Moussa and Cocquempot 2017). The single locations of outbreaks of the three European countries examined in this study (Germany, Italy and Switzerland) will be listed in more detail here according to the reports from EPPO (2021b). In Germany the first outbreak was detected in Neukirchen am Inn (Bavaria, 2004), then in Bornheim (North Rhine-Westphalia, 2005), Alfter (North Rhine-Westphalia, 2009), Grenzach-Wylen (2011)/Weil am Rhein (Baden-Wuerrtemberg, 2012), Feldkirchen (Bavaria, 2012), Neubiberg (Bavaria, 2014), Schoenebach/Ziemetshausen (Bavaria, 2014), Magdeburg (Saxony-Anhalt, 2014), Kelheim (Bavaria, 2016), Murnau (Bavaria, 2016), Hildrizhausen (Baden-Würrtemberg, 2016), and Miesbach (Bavaria, 2019). In Italy, *A. glabripennis* was first found in Corbetta (Lombardia, 2007; Hérard et al. 2009), a few years later than in the neighbouring Vittuone (2010) and Sedriana (2013), Cornuda (Veneto, 2009), Maser (Veneto, 2009), Grottazzolina (Marche, 2013), in Ostra (Marche, 2016), in Trescore-Balneario (Bergamo, 2017), Vai (Torino, 2018), Cuneo (Cuneo, 2018), Fermo (Marche, 2019), Civitanova (Marche, 2019) and Sant'Elpidio a Mare (Marche, 2020). The first outbreak in Switzerland was found in Bruensried (Freiburg, 2011), after that it was detected in Winterthur (Zürich, 2012), Marly (Freiburg, 2014) and Berikon (Aargau, 2015). All underlined locations of these three countries were sampled for this study. In all outbreaks the beetle were under strict surveys and eradication measures. (EPPO 2021a)

In Europe, many infestations are already considered as successfully eradicated like sites in Austria (Braunau am Inn 2001-2012, St. Georgen 2012-2016, Gallspach 2013-2021), France (Strasbourg 2008-2019, Saint-Anne-sur-Brivet 2004-2013), Germany (Weil am Rhein 2012-2019, Hildrizhausen 2016-2020, Neukirchen am Inn 2004-2015, Feldkirchen 2012-2020, Neubiberg 2014-2019, Kelheim 2016-2020, Murnau 2016-2020), Italy (Cornuda 2009-2020, Maser 2009-2020), Montenegro (Budva 2015-2019), the Netherlands (Winterswijk 2012-2016, Almere 2010-2011), Belgium (2008-2011), Finland (Vantaa 2015-2021), Switzerland (Berikon 2015-2019, Marly 2014-2019, Bruensried 2011-2018, Winterthur 2012-2016) and the United Kingdom (Paddock Wood Kent 2011-2019; Eyre and Barbrook 2021) according to EPPO (2021b). The locations underlined in this list are from infestations which were investigated in this study and thereby only represent the history of European infestations.

However, the risk of new introductions into currently not infested areas, reintroduction in already eradicated areas, or even secondary spread events from currently infested areas is still present, since especially biological invasions which are facilitated via global trade in WPM will inevitably grow in the future (Hulme 2021; Seebens et al. 2017). Furthermore, ALB is not

only threatening urban landscapes, endemic biodiversity and poses a threat to pedestrians and vehicles by falling branches. Additionally, the enourmous economic impact caused by the damage, phytosanitary measures and eradication programs weigh down with e.g. total costs of 373 million USD in USA, 464 000 € in Austria, 55 000 € in France, 65 000 € in Germany and 23.5 million CAD in Canada in 2008 (Haack et al. 2010). Hence, this apparently demands more research adding knowledge to already existing research. Many studies addressed different parts on management strategies of ALB invasions, e.g. studies on the biology (e.g. laboratory rearing (Keena 2002, 2005, 2006; Torson et al. 2021a; Torson et al. 2021b), midgut community (Mason et al. 2019)), biological control (fungus *Metarhizium brunneum* (Clifton et al. 2020), parasitoid *Dastarcus helophoroides* (Coleoptera: Bothrideridae) (Gould et al. 2018), parasitoid *Ontsira mellipes* (Hymenoptera: Braconidae) (Wang et al. 2019; Wang et al. 2020c), RNAi (Dhandapani et al. 2020; Rodrigues et al. 2017) for improved management of invasive populations. In the following the focus will be on the previous molecular characterization approaches investigated on *A. glabripennis* to get an overview where the study presented here will build on the current state of knowledge.

Research status on the invasion of *A. glabripennis*

There are a lot of molecular studies available to answer key questions about the invasion of ALB. For example, Wu et al. (2017) examined taxon identification using the mitochondrial DNA marker COI on invasive wood-boring beetles of the Cerambycidae and Buprestidae families in combination with larval rearing of specimens found during controls of WPM arriving at ports. This should circumvent both, errors in taxonomic classifications due to the difficult morphology of immature insect life stages and database DNA barcoding gaps for commonly intercepted invasive species. However, other studies addressing the taxon identification of *A. glabripennis* used RAPD-SCAR markers (randomly amplified polymorphic DNA, sequence characterized amplified regions) (Kethidi et al. 2003), the mitochondrial DNA marker COI with quantitative real-time PCR (Taddei et al. 2021) and several other mitochondrial markers for a rapid LAMP approach (Rizzo et al. 2020). All were cross-checked on their identification capabilities, especially to distinguish from endemic species (e.g. *Monochamus galloprovincialis*) or other invasive species (e.g. *Aromia bungii*, *Anoplophora chinensis*) in Europe. Ohbayashi and Ogawa (2009) addressed the speciation and distribution of the lamiine *Anoplophora* genus on the basis of mitochondrial COI sequences for phylogenetic inferences and could classify *A. glabripennis* to its own lineage apart from other *Anoplophora* spp. with two sub-lineages. For ALB, also some transcriptomics studies are available studying either traits on olfactation to improve lures used in monitoring (Mitchell et al. 2017) or digestion to characterize metabolic plasticity (Mason et al. 2016; McKenna et al. 2016; Scully et al. 2014). These studies pointed out the extraordinary enzymatic repertoir to digest woody host plants and detoxify plant

defensive compounds, allowing ALB a high functional divergence which facilitates invasiveness. Furthermore, McKenna et al. (2016) sequenced (assisted by i5K consortium) and annotated the whole genome of a female *A. glabripennis* larva from a "mixed colony" from US infestation sites, creating a draft reference assembly of 710 Mb. Within the order Coleotera with a mean genome size of 974 Mb, the ALB genome (female 981.42 ± 3.52 Mb, male 970.64 ± 3.69 Mb) is average-sized (McKenna et al. 2016). The assembly was updated on 12th December 2017 and is available as representative genome GCF_000390285.2 (NCBI Agla 2.0) with a total length of 706.95 Mb with 9867 scaffolds (N50: 678 Kb) and 26749 contigs (N50: 80 Kb), with a prediction of 20619 proteins, 27 rRNAs, 447 tRNAs, 910 other RNAs, 16187 genes, 209 pseudogenes and a GC content of 33.4%.

With COI-II and microsatellite sequences from multiple sites in China and one site in Korea (natural forests, Kangwon Province, Seoraksan national park), Carter et al. (2009c) found genetic structure within and among populations in the native range. The genetic structure could hereby not be fully geographically discriminated, since a lot of haplotypes and microsatellite clusters were mixed within these collection sites. The genetic structure seems to be mostly shaped by multiple admixture events and secondary spread with the highest variation found within populations and might have been influenced by human-mediated movement of ALB in Asia. Also Javal et al. (2019a) found most diversity of COI-sequences within Asia, where some localities held up to 7 haplotypes, while some main haplotypes were observable very widespread over all sampled Asian sites. They suspected also a possible spread of the most widespread two haplotypes over common transportation routes of goods. By using the microsatellite markers of Carter et al. (2009a), the group of Javal et al. (2019b) found moderate genetic differentiation (most F_{ST} 0-0.15) with a blurred geographic structure in China showing a slight north-south gradient and high inbreeding (F_{IS}>0.2), which is quite uncommon in the native range. This indicates complex, and considering the natural dispersal capacities, most likely human-mediated migration events within the native range. For China this can be explained with the massive reforestation programm by the Chinese government to antagonize desertification (see p. 24), resulting in mixed populations which could proliferate enourmously in the Chinese native range. These results suggest that the source populations of further global invasions were invasive themselves in China, since natural forests could not be sampled. Hence, an insight in the historical genetic structure was impossible. (Carter et al. 2009c; Carter et al. 2010; Javal et al. 2019b) Additionally, Javal et al. (2019b) could date the merging of the ancestral native populations between 145 and 161 generations ago using microsatellites. But the putative ancestral structure is attenuated by the demographic events that were influenced by human activities. Different to the situation in China, in South Korea *A. glabripennis* was not ranked as a pest, since it just occurred in low densities in the natural forests with higher tree diversity and more regulating woodpecker predation (Williams et al. 2004). In contrast to China,

Javal et al. (2019b) could detect geographic differentiation in South Korea between different locations and lower effective population sizes. But recently, some disruptive ALB populations in the South Korean urban areas of Seoul, Incheon, Ulsan, Busan, Jeonju and Gimhae were observed, which were suspected to be invasive beside the indigenous populations in Pocheon, Suncheon, Goheung and Seoraksan national park (Kim et al. 2019; Lee et al. 2020). Kim et al. (2019) was able to find genetic divergence between the natural populations and the suspected invasive populations by COI-barcoding, but with the low sampling size the results were not evident enough. However, Lee et al. (2020) was able to find proof by ABC analysis based on COI sequences (primers of Carter et al. 2009c) that *A. glabripennis* in Incheon and Seoul are an invasive alien subgroup, which most likely originated from Northeastern China and ALB from Busan and Ulsan were also found to be an invasive alien subgroup, originating from Northwestern China. Hence, there was no range expansion from the natural forests of Kangwon Province detectable. This study is the first one which identified ALB populations as invasive ones within the native range. The invasive ALB populations in South Korea might have been established first in the respective cities and have been the source of bridgehead expansions subsequently. (Lee et al. 2020)

The invasive populations in North America and Europe were also investigated on their genetic variation and population structure using COI (Javal et al. 2019a), COI-II (Carter et al. 2009b; Carter et al. 2010) and microsatellite data (Carter et al. 2010; Javal et al. 2019b; Tsykun et al. 2019). The genetic patterns observed in North American and European invasive populations were broadly similar. For example, the three most dominant COI haplotypes found by data of Carter et al. (2010) and Javal et al. (2019a), occurred both in several native provinces as well as in invasive populations. Hence, all haplotypes found in invasive populations represented only samples from the native haplotype spectrum (Carter et al. 2010). Javal et al. (2019a) found European COI sequences were widedly dominated by one very common haplotype. The genetic diversity of ALB in Ontario (Vaughan, Toronto), based on COI-II and microsatellite data, was lower than from findings in the USA (New York City, Linden, Chicago, Long Island), which also hold limited genetic diversity in comparison with populations from China (Carter et al. 2009b; Carter et al. 2010). All Canadian populations were also assigned to the same genetic cluster by Javal et al. (2019b). The data suggests that the Toronto populations were derived from a small number of beetles and suffered more severe bottlenecks than the invasive populations in the USA (Carter et al. 2009b). In addition, the Toronto COI haplotype was only found in Neukirchen am Inn (Germany) so far and not in the sampled locations from China and Korea (Carter et al. 2009b; Carter et al. 2010). Another shared COI haplotype between North America and Europe was found in New York City and Milano (Javal et al. 2019a).

All results from previous studies on introduction patterns came to the same conclusion, in both North American and European invasive populations, most populations are homogeneous within and distant to other introduced populations. All studies suggested this pattern as a result of multiple separate introduction events and in some cases human-mediated or natural secondary spread within close vicinity of an outbreak or even over a far distance (bridgehead scenario). (Carter et al. 2010; Javal et al. 2019a; Javal et al. 2019b; Tsykun et al. 2019) In more detail, European secondary spreading events in close proximity were revealed by their related genetic structure based on microsatellite data, for instance in Marly and Bruensried and between all Corsican outbreaks (Javal et al. 2019b; Tsykun et al. 2019). However, multiple introductions of the same source population is also a possibility and difficult to exclude or proof (Javal et al. 2019b). In the US, the data also suggested an expansion from New York City to Jersey City and from Carteret to Linden (Carter et al. 2010; Javal et al. 2019a). ABC analysis based on microsatellites detected the Gien population in France to be a result of a bridgehead pathway from an invasive population in the USA (Javal et al. 2019b). Tsykun et al. (2019) was the first study for a fine-scale population genetic insight into invasive outbreaks in Switzerland. Out of these four outbreaks, three independent introductions were measured to be responsible for the Swiss invasions. The infestation in Bruensried was most likely a hitch-hiking translocation via firewood from Marly, which could be confirmed with ABC analsis. The source population of Marly is suggested to be from South Korea (Javal et al. 2019b; Tsykun et al. 2019). They also detected, that several other single findings could no be assigned to any of the Swiss established populations, hence these multiple independently introduced beetles have failed to establish (Tsykun et al. 2019).

The sole usage of mitochondrial DNA-markers was found to not be capable to identify source populations of ALB (Carter et al. 2010). The same can be said about microsatellite data, since no private alleles could be assigned to only one native population (Carter et al. 2009c), but the data were suffiently to precisesely demonstrate different source populations between findings in some cases, e.g. New York city districts differed from Carteret and Linden (Carter et al. 2010). Although all these studies facilitated insights in the population structure and dispersal patterns, they can only result in relatively low-resolution information on single outbreaks and are limited in the precise reconstruction of introduction histories. With low-resolution DNA markers, this may be hindered due to the suspected invasive characters of the Chinese source populations, which is assumed to be solved by the usage of genome wide DNA-markers. Given this research background, this study here abstained to raise more genotypes on microsatellites, since the previously findings did not confidentely assign invasive populations to native source populations. The decision was made to collect COI sequence data (3.3) to supplement the databases with new sequences from European infestation sites, which were not or rarely represented so far, to improve the identification reference basis. For a high-

resolution investigation of the introduction patterns, a new Genotype-by-sequencing approach (3.5) was chosen to collect genome-wide polymorphic data to characterize the genetic background of ALB invasions in Europe.

Very recently (after submission of this dissertation in March 2022), Cui et al. 2022 published their work on population structure of native ALB populations using high-resolution genome-wide SNP markers. This study aimed to disentangle the historical movements between regions in the native range to develop applicable biosurveillance tools. Using 6102 informative SNPs and 53 microsatellites they determined six distinct population clusters among the native ALB populations and a clear separation between South Korean and Chinese populations.

3.3 Mitochondrial marker and COI Barcoding

Morphological identification methods are sometimes very challenging when applied to different life stages of insects. Especially during the invasion process, any life stage can be introduced, which makes improved ability to identify even difficult to distinguish stages mandatory for good diagnostics and reliable specimen determination. The limitations of morphology-based identification like e.g. phenotypic plasticity, morphologically cryptic taxa and the shrinking amount of taxonomic expertise can often lead to misdiagnoses and therefore needs assistance by molecular specimen identification (Hebert et al. 2003). Mitochondrial based DNA-markers are often used for this purpose, as well as for phylogenetic studies (Mandal et al. 2014; Wu et al. 2017). The mitochondria, organelles of endosymbiontic origin, are involved in several key processes such as respiration (Fontanesi 2015), aging (Walter and Lee 2009) and apoptosis (Wang and Youle 2009). They possess their own genome, which is relatively conserved for most metazoa, where the mtDNA encodes 13 proteins involved in oxidative phosphorylation (OXPHOS), 2 rRNAs and 22-23 tRNAs for their own translational machinery (Ladoukakis and Zouros 2017; Mandal et al. 2014), see **Figure 8**.

Beside the important role in operational cells, the synergistic co-evolution of nuclear and mitochondrial DNA (mtDNA) is a very important part of the evolution of eukaryotes and has been found to be very valuable in population genetics and phylogenetics (Ladoukakis and Zouros 2017). MtDNA has already enabled important insights into invasive insects so far (Carter et al. 2010; Grapputo et al. 2005; Javal et al. 2019a; Lee et al. 2020; Scheffer and Grissell 2003). Despite the limitations (e.g. introgression, selection; Ballard and Whitlock 2004), mtDNA has been widely used to investigate phylogeographic structure. Evolutionary relationships can be inferred by examining the change of structure and function of molecular markers, which has to be chosen according to the objectives (Mandal et al. 2014). Herein, the application of mitochondrial DNA-markers is advantageous for several reasons. The small size (13-19 kb), lack of introns and untranslated regions apart from the D-loop, as well as the

conserved and simple gene organization among metazoa (Hebert et al. 2003; Mandal et al. 2014; Saccone et al. 1999) makes mitochondrial DNA-markers easy and cost-efficient to use, even without previous knowledge about the species. Different than nuclear DNA-markers, the multi-copy and homoplasmic character of mtDNA facilitates the more convenient direct PCR product sequencing. (Ladoukakis and Zouros 2017) Additionally, the high copy numbers enable the amplification of mtDNA markers of even small or degraded biological samples (Mandal et al. 2014). Furthermore, strict maternal inheritance, the rapid evolution (5-10 times elevated mutation rates compared to nuclear DNA), lack or limited recombination, infrequent rearrangements as well as less and highly conserved regions within the same mtDNA molecule are contributing to the advantages of mtDNA markers as valuable tools for population genetic and phylogenetic studies (Bernt et al. 2013; Brown et al. 1979; Ladoukakis and Zouros 2017; Mandal et al. 2014). By uniparental inheritance, individuals which possess only one type of mtDNA (homoplasmy) are derived, preventing the advance of selfish, fast replicating mutations (Greiner et al. 2015; Hastings 1992; Hurst 1995). Recombination in mitogenomes only occurs in some rare circumstances, such as the fusion of mitochondria with different mtDNA (Ladoukakis and Zouros 2017). The rapid evolution originates from the limited repair system in mitochondria (Brown et al. 1979). Moreover, all these properties together allow not only the phylogenetic reconstruction of each lineage as one evolutionary history at several taxonomic levels, but also the comparison among both, individuals of the same population and among distantly related species (Cox and Hebert 2001; Ladoukakis and Zouros 2017; Wares and Cunningham 2001). Additionaly, these unique properties eased the design and usage of universal primers for the amplification of robust mtDNA segments to discriminate taxonomic groups (Folmer et al. 1994; Hebert et al. 2003; Simmons and Weller 2001; Zhang and Hewitt 1996b). Of all mitochondrial DNA markers, some of the protein-coding sequenes were found to be most suitable to infer evolutionary relationships, either within or between phyla (ND4, ND2, cytb, and COI), among which cytochrome oxidase subunit I (COI) is the most widely used for species identification and phylogenies. Protein-coding genes are better DNA-marker, since indels (insertions and deletions) are rare, because they would inevitably lead to frame shifts. (Bourlat et al. 2008; Mandal et al. 2014; Zardoya and Meyer 1996) There are very robust universal primers available to amplify COI sequences for almost all animal phyla, such as a 710 bp fragment of COI which was proven to be reliable for 11 invertebrate phyla and is used as a common arthropod universal primer (Folmer et al. 1994; Hebert et al. 2003). The evolution of COI is fast enough to discriminate close species and phylogeographic groups within species, showing high levels of diversity (Bourlat et al. 2008; Carter et al. 2010; Hebert et al. 2003; Javal et al. 2019a; Kim et al. 2019).

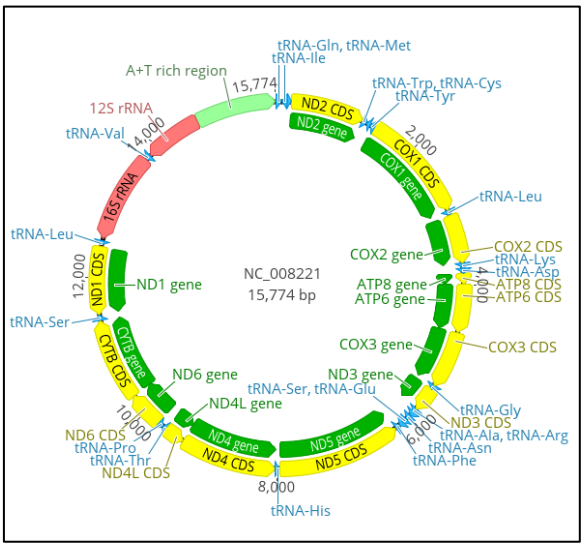


Figure 8. *Anoplophora glabripennis* mitochondrial genome (NC_008221; 15774 bp). ALB's genome contains 13 protein-coding genes (green; CDS yellow), 2 rRNA genes (red), 22 tRNA genes (blue) and an A-T-rich region (light green). The gene arrangement resembles other known beetle mitochondrial genomes and the nucleotide composition is strongly biased towards A/T (78.30%) (Fang et al. 2016), which is very common for insect mitochondrial genomes (e.g. *Apis mellifera*, Crozier and Crozier 1993) and other arthropodes (e.g. *Daphnia pulex*, Crease 1999).

DNA reference libraries can be used to assign unknown specimens to a taxonomic category, e.g. during assessment of biodiversity in conservation or at border biosecurity to identify IAS (Collins et al. 2012; Darling and Blum 2007). Widely used for this purpose are DNA barcoding approaches, which use standard universal genes to distinguish species according to species specific genetic variation, which can be compared with databases (Hebert et al. 2003). For IAS, DNA-barcoding is also used to characterize diversity and abundance of IAS even within an assemblage of indigenous and noningenious species in the sampled area (e.g. Smith and Fisher 2009). DNA barcoding is a commonly used, convenient, inexpensive and rapid diagnostic method which can help to identify the taxonomic level of specimens by estimating genetic distances within and between animal species. These distances are assessed by alignment with reference sequences deposited on public databases to find the closest reference match record. The first barcoding DNA-marker was COI which is used for metazoa kingdom. (Hebert et al. 2003) The species specific 'DNA barcodes' have to be assembled on databases in advance, such as Barcode of life database, BOLD (http://www.barcodinglife.org) (Ratnasingham and Hebert 2007). On this platform, barcode index numbers (BINs) are automatically assigned to specimen-clusters belonging to one species. Beside the BOLD database, lots of other phylogenetic studies deposited COI sequences in GenBank (https://www.ncbi.nlm.nih.gov), with whom BOLD has an exchange agreement, so GenBank barcodes can also be mined from BOLD (Kjærandsen 2022). However, the usage of a single barcoding region is not appropriate for all invasive species, since some lack a reference library

or the genetic variation is not informative enough and may limit the outcome (Collins and Cruickshank 2013; Dupuis et al. 2012; EPPO 2016). Several protocols for eukaryotes and prokaryotes are validated within the Quarantine Organism Barcoding of Life (QBOL) project. Regulated organisms have to be identified by finding the closest matching reference record using e.g. Basic Local Alignment search tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) hit identity. For *A. glabripennis* a COI-barcoding standard protocol is used with LCO1490 and HCO2198 primers of Folmer et al. (1994) according to EPPO (2016) and can also discriminate ALB from the closely related regulated pest *Anoplophora chinensis*. (EPPO 2016)

The outcome of DNA barcoding studies is strongly confounded by incomplete databases, incorrect species identities in databases, co-amplification of nuclear mitochondrial-like sequences (NUMTs), introgression or hybridization events (EPPO 2016), as well as recombination and heteroplamy. In the absence of recombination, deleterious mutations could accumulate much faster (Muller's ratchet), hence there are very litte cases of recombination in mtDNA sufficient to antagonize the mutation load (Gordo and Charlesworth 2000; Neiman and Taylor 2009). However, by using mitochondrial DNA-markers, the presence of recombination can have an undesirable effect on the branch lenghth in phylogenies, like longer total branch length and shorter times to the most recent common ancestor (MRCA) (Schierup and Hein 2000). As decribed before, the mitogenomes usually are homoplastic due to their maternal inheritance, which is maintained by the nucleus and reinforced by pre- and postfertilization bottlenecks which mitigate mitochondrial proliferation (reviewed in White et al. 2008). Additionally, paternal mtDNA is actively destroyed after fertilization (Al Rawi et al. 2011; Sato and Sato 2011). Therefore, heteroplasmy occurs when these mechanisms are disturbed or overcome. For instance, mother-inherited heteroplasmy may occur when the egg already contains two or more types of mitogenomes. Somatic mutations can also occur during mtDNA replication and can lead to the co-existence of mutated and normal mtDNA, and defected molecules may increase their frequency. Another source for heteroplasmy can be paternal leakage. Thereby the mitochondria of the sperm could avoid their destruction inside the egg during fertilization, e.g. in interspecific crosses. (Ladoukakis and Zouros 2017; Rokas et al. 2003) The point mutations or length variants in mtDNA genes are involved in some severe diseases in humans (Stewart and Chinnery 2015), while *Drosophila melanogaster* showes signs for an adaptive purpose of heteroplasmy, were both molecules in the same individual complemented deleterious effects (Ma et al. 2014). Another natural process confounding COI barcoding is caused by mitochondrial DNA sequences which enter the nucleus and generate nuclear mitochondrial DNA (NUMTs), which will be described in more detail in ch. 3.4.

Nevertheless, continued COI-barcoding efforts will help to fill gaps of public databases of some taxonomic groups and will aid to remove identification errors (Wu et al. 2017). Despite the

great potential, COI-barcoding will never substitute the important contribution of taxonomic experts to curate reference libraries and to describe new species (Kjærandsen 2022). These approaches should additionally be complemented by other molecular markers, for instance genome-wide SNP marker (3.5) for superior phylogenetic and population genetic inferences.

3.4 Nuclear-mitochondrial like sequences (NUMTs)

The more whole genome sequencing projects are available, the more transfers of mitochondrial DNA sequences are detected in nuclear DNA of many different species, which could have caused several unregognized biases in PCR-based studies, such as DNA barcoding approaches (Bernt et al. 2013). For eukaryotes, the exchange of DNA between mitochondria and nucleus has been quite common throughout evolution, which led to large mitogenomic copy fractions in the nuclear genome, especially during the early organelle evolution (Kleine et al. 2009). For this circumstance, the mtDNA must previously reach the nucleus to get integrated. Possible pathways of mtDNA into the nucleus contain the decay of abnormal mitochondria (Campbell and Thorsness 1998), lysis of mitochondrial compartments, encapsulated mtDNA in the nucleus and membrane fusions of mitochondria and the nucleus. Free mtDNA in the nucleus can be occationally integrated in nuclear chromosomes. (Hazkani-Covo et al. 2010) The insertion of mtDNA takes place during the repair process of double-strand break sites on nuclear DNA, whereby chromosomal DNA on one side and mtDNA on the other side are linked either with blunt-end or sticky-end repair mechanisms (Blanchard and Schmidt 1996; Hazkani-Covo et al. 2010; Kleine et al. 2009; Ricchetti et al. 1999). Some of the absorbed genes of mitochondrial origin from early evolution adopted new functions or have been re-imported into the mitochondria, while more recent transfers of mtDNA to the nucleus persist as NUMTs, non-coding and non-functional sequences (pseudogenes), in the nuclear genome or become new exons in preexisting nuclear genes (Kleine et al. 2009). NUMTs can occur as continuous stretches which are colinear with the mtDNA, but sometimes they occur in more complex patterns of non-continuous, rearranged and scattered NUMTs derived from different parts of the mitogenome and can even appear in different orientations (**Figure 9**; Leister 2005; Ricchetti et al. 1999; Tourmen et al. 2002).

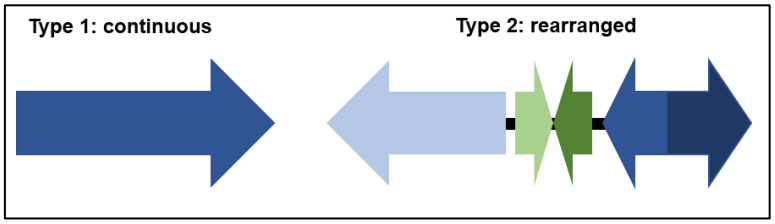


Figure 9: Types of NUMT organisation in the nuclear genome (modified Leister 2005). Two other types occur solely in plants and include nuclear plastide DNA copies (NUPTs).

With quite uniform insertion rates, subsequently the NUMTs can undergo post-insertion processes such as (multiple) duplications or deletions, just like the rest of the whole genome (Bensasson et al. 2003), whereas older NUMTs exist in more copies than recent ones (Hazkani-Covo et al. 2003). Almost all eukaryotic genomes have detectable ongoing processes of mitochondrion-to-nuclear DNA transfer and thereby always new supply of mitochondrial pseudogenes in the nucleus (Bensasson et al. 2001b; Leister 2005). NUMTs exist in many species in notable magnitude, for instance in humans over 400 kb (Hazkani-Covo and Graur 2007), in *Nasonia vitripennis* 43 kb and *Apis mellifera* over 230 kb, while *Drosophila melanogaster* almost lacks completely of NUMTs (Viljakainen et al. 2010). The frequencies, but also the size distribution of NUMTs differ in different genomes (Richly and Leister 2004). For instance, the genome of *Apis mellifera* has a very high NUMT sequence frequency, but those NUMTs are relatively short (Pamilo et al. 2007), while e.g. some of the human NUMTs comprise nearly the whole mitogenome as copy (Dayama et al. 2014; Mourier et al. 2001). NUMTs also very commonly occur in other insects, for instance in the stingless bee *Melipona capixaba* (Cristiano et al. 2012), several grasshoppers (Bensasson et al. 2000) and bark beetle species (Jordal and Kambestad 2014), as well as in the long-horned beetle *Monochamus galloprovincialis* (Koutroumpa et al. 2009), some of them in high frequencies.

The fraction of detectable NUMTs in genomes is commonly <0.1%, but they represent more or less the rather recently inserted NUMTs, as older NUMTs become undetectable over time by mutations or deletions (Hazkani-Covo et al. 2010). Recent insertions of mtDNA in the nuclear genome occur in unexpectedly high rates and correlate with genome size. Species-specific differences in NUMTs frequencies are influenced by the rate with which mitochondria are transfered, the frequency of chromosomal integrations and post-insertion processes (duplications, deletions) acting on the genome. The limitation of the integration events is derived by the rate of double-strand breaks that happen during growth or DNA damage. It is suggested but still controversial, that double-strand breaks are more frequent in larger genomes, hence it is possible that more NUMTs are integrated. (Bensasson et al. 2001a; Bensasson et al. 2001b; Hazkani-Covo et al. 2010; Wang et al. 2020b) Double-strand breaks can be increased during transcription events (Aguilera 2002). Consistent with this suggestion is the higher NUMT frequency in introns than in intergenic regions (Behura 2007; Ricchetti et al. 2004), indicating that open chromosomes facilitate NUMT insertions (Hazkani-Covo et al. 2010).

In phylogenetic studies, undetected divergent NUMTs can bring enourmous bias in form of overestimated diversity or in DNA metabarcoding approaches overestimated number of species. This can strongly compromise phylogenetic studies and specimen identification, due to the similarity of the NUMT sequences to the mtDNA. (Haran et al. 2015; Song et al. 2008;

Sorenson and Quinn 1998; Thalmann et al. 2004; van der Kuyl et al. 1995) The ambiguity introduced in DNA barcoding studies of insects led to enourmous confusions and misinterpretations (Haran et al. 2015; Song et al. 2008).

In the case of unwanted co-amplification of NUMTs, methods are required to identify contaminations in the collected sequences or in databases. Otherwise, when NUMT sequence incidences and troubles in sequencing approaches are already known for the studied species, averting strategies need to be implemented. To avoid NUMT coamplification prior to sequence analysis, several strategies have been investigated so far. One strategy aims at the enrichment of mtDNA by using tissue rich in mitochondria (e.g. muscle) for DNA extraction (Song et al. 2008; Sorenson and Quinn 1998), capture-enrichment (Wang et al. 2020a), or by selective sedimentation of mitochondria (Calvignac et al. 2011), where the chance of amplification of only mtDNA should be enhanced. Other PCR methods like RT-PCR, since mtDNA is transcribed and NUMTs are not, as well as long-range PCR or whole mitogenome sequencing can aid escaping NUMTs co-amplification (Bensasson et al. 2001b; Lopez et al. 2022; Sorenson and Quinn 1998; Wang et al. 2020a). Other studies use species specific mtDNA instead of universal primers (Malik et al. 2016; Moulton et al. 2010). All these avoidance strategies can enhance the chance of correct mtDNA sequencing, but are time-consuming and can not fully promise to not co-amplify NUMTs (Song et al. 2008). Therefore, strategies are required to also detect them when they are already in the dataset. NUMTs of species with both nuclear and mitochondrial genomes available can be inferred by standard data-mining tools, e.g. BLASTn of the whole mitogenome sequence as query against the nuclear genome. Among others, E-scores <0.0001 are used as a criterion of the BLAST hits to identify NUMTs. (Pamilo et al. 2007; Richly and Leister 2004) If NUMTs have disturbed sequencing approaches already, this can be identified by ambiguity of the sequences visible as double peaks in the chromatogrames of direct PCR product sequences (Haran et al. 2015), or by intra-individual variation detected by the sequencing of cloned PCR products, even though NUMTs are difficult to be discriminated from heteroplasmy in cloned sequences (Berthier et al. 2011). Additionally, mtDNA marker sequences should be checked for length variations and in-frame stop codons (Javal et al. 2019a; Kim et al. 2019; Song et al. 2008; Zhang and Hewitt 1996a).

Alternativey, some studies used NUMTs intentionally for phylogenetic inferences, e.g. in primates (Hazkani-Covo 2009) and for interspecific diversity (Richly and Leister 2004). Herein, the individual and population-wise polymorphisms in NUMTs, as well as prescence versus absence variations like those found for human populations can be informative features (Hazkani-Covo et al. 2010). A broader understanding of the post-insertion processes on population genetic level will benefit future studies using NUMTs as DNA-marker. The results of Bensasson et al. (2003) assume NUMTs do segregate in populations like neutral DNA-

markers. In this case, NUMTs could be used as 'molecular fossils' of ancestral mitochondrial sequences, since e.g. the oldest NUMTs found in humans were inserted 58 million years ago. Hazkani-Covo (2009) describes NUMTs as tools that might harbour more information than other segregating markers, as they resemble mitochondrial and nuclear polymorphisms within one marker. With NUMTs as DNA markers, insights into earlier evolution of both mitochondrial and nuclear genomes can be investigated and give hints of ancient admixture events. This approach makes it possible to reconstruct overlapping geographic distributions of different populations of a species. (Hazkani-Covo et al. 2010)

3.5 Genomic SNP marker identification via Genotype-by-sequencing

Routinely, genetic markers are utilized to identify source populations of species by using mtDNA markers in combination with microsatellites (3.1), but genomic markers provide a much higher resolution with thousands of SNP markers, with which processes in the present and past of populations can be documented in an unprecedented range (e.g. *Aedes aegypti*; Rašić et al. 2014 using ddRAD sequencing). New insights into the genetics of invasive species are on the rise, also for nonmodel species, by the advances in high-throughput sequencing (HTS) and genotyping methods (Estoup et al. 2016). Draft genomes and genomes of related species can also be used for variant determinations (Chown et al. 2015). Genomic approaches in the investigation of invasion processes can significantly enhance the knowledge about species identification, colonization, spread and adaptation. Thereby, genomics contribute to the reduction of impacts by IAS by improving early detection and intervention strategies. Genomic biosurveillance require high-quality data from multiple sources to trace invasion routes, source populations and adaption, but the resolution obtained by genomic approaches is highly increased (Bilodeau et al. 2019; Roe et al. 2019). Next-generation sequencing (NGS) technologies (Pochon et al. 2013) and genomic tools (Rašić et al. 2014) enable detection, sequencing and analyzing the genetic connections between samples from complex environmental samples which are either mixed, at low abundance or even partially degraded (Taberlet et al. 2012). The advent of genomic technologies enables a deep characterization of relationships between native and introduced populations quite fast and relatively inexpensive (e.g. *Mimulus guttatus*; Puzey and Vallejo-Marín 2014 using genome re-sequencing).

For many cases, whole genome sequencing approaches are not necessary, thus HTS methods are often concomitant with a reduction of the complexity and genome representation like Genotype-by-sequencing (GBS) (Elshire et al. 2011) and restriction-site-associated DNA sequencing (RAD-Seq) (Baird et al. 2008). The reduction of the genome is induced by specific restriction enzyme (RE) digestion and followed by ligation of adapters, PCR amplification and sequencing. Especially methods like GBS make HTS even more feasible with the opportunities to genotype a vast number of individuals at thousands of SNP markers at the same time

(Elshire et al. 2011). Thereby thousands of variants are genotyped for samples across the genome, which can be used as SNP-markers in studies of intraspecific genetic variation, population structure or GWAS approaches according to the research objective (Roe et al. 2019). The GBS procedure is reduced in laboratory and computationally steps in comparison to the RAD-seq method, while it is still very specific, reproducible and can be conducted also on large and very diverse genomes. Even the accessibility of complete reference genome sequences is not mandatory, because this can be replaced by the samples themselves around the restriction sites during the genotyping process. Thereby the consensus around the sequence tags becomes the reference. With this method, population structure can be investigated without prior knowledge about the genome of the species of interest. (Elshire et al. 2011) Baird et al. (2008) first described the advances of sequencing restriction-site associated genomic DNA (RAD tags) for the data collection of SNPs and genotyping. With the GBS method first described by Elshire et al. (2011), the multiplexing strategy is considerably enhanced by a barcoding system which is already included in an adapter (**Figure 10-a** and **-b**) and therefore saves steps and reagent costs in comparison to the original RAD-seq method. Another advantage by the GBS adapter system is the position of the barcode spares the indexing sequencing step usually necessary for Illumina sequencing (**Figure 10-c**) (Elshire et al. 2011). The terms RAD-seq and GBS discriminating the different methods are not quite straightforward and widely discussed (Campbell et al. 2018), since both original methods and later derivates of them use restriction-site associated sequencing and genotype while sequencing, so the terms are nowadays often used synonymously. Both methods RAD-seq (Baird et al. 2008) and GBS (Elshire et al. 2011) reduce the genome complexity by RE and use multiplexed NGS for high-density SNP discovery. For the first described RAD-seq method a barcode adapter is ligated to the cut genomic DNA, then the samples are pooled, randomly sheared and size selected (300-700 bp). Subsequently another Y-shaped adapter is ligated and the fragments are amplified. Differently, in the first described GBS the enzymatic and purification steps are reduced, no Y-shaped adapters are used, and the DNA is neither randomly sheared nor fragment size selected between the pooling and PCR steps. The two used barcode and common adapters are ligated simultaneously. After the sequences are aligned to the reference genome, with the RAD-seq approach downstream regions of all fragments >300 bp are sequenced, while with GBS short fragments of 150-350 bp are sequenced. (Davey et al. 2011; Elshire et al. 2011)

The GBS method was successfully applied to characterize the introduction patterns of the invasive gypsy moths *Lymantra dispar asiatica* and *L. d. japonica* to North America. With the large SNP dataset produced by GBS, geographic variants, determination of subspecies and introgression events could be detected. (Picq et al. 2018) Hence, this method is promising for the investigation of introduction patterns *of A. glabripennis* in Europe as well. In this study, the

normalised GBS (nGBS) protocol of LGC Genomics GmbH was used, which is applicable to any genome without the need of prior knowledge and is unaffected by genome size, methylation patterns or repetitive sequences. Commonly, for new species, several trials for the appropriate RE combination must be performed in advance, but nGBS uses for all genomes the blunt end cutter RE MsII. The fragments are reduced by a normalization step (RE treatment) to diminish high copy number fragments and paired-end 150 bp sequenced. (Arvidsson et al. 2016)

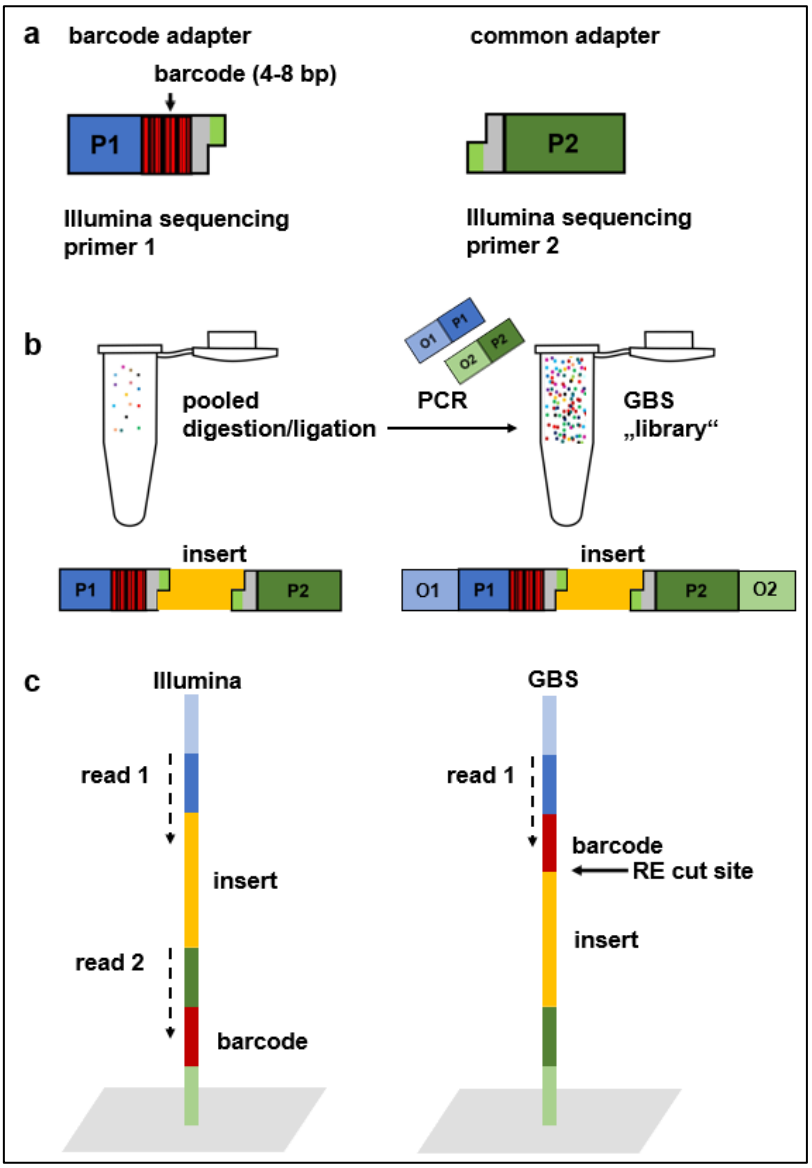


Figure 10. GBS adapters, PCR and sequencing primers (modified Elshire et al. 2014). **a)** double-stranded barcode and common adapter to which cut genomic DNA can be ligated at the “sticky ends”; **b)** two PCR primers which include the binding sites to the flowcell oligonucleotide 1 and 2 (O1 and O2), barcode and paired end sequencing primer 1 and 2 (P1 and P2); **c)** comparison of standard Illumina adapters and GBS adapters, showing that the barcode and RE cut site are physically attached to the DNA insert and can be sequenced at once, while standard adapters need at least two reads and more computational effort to combine the barcode and the sample information afterwards (Elshire et al. 2011; Elshire et al. 2014; workshop on youtube).

The resulting large datasets of SNP markers, which are randomly distributed across the genome, allow a lot of insight regarding demographic history and can estimate multiple origins and admixture, even when the source populations are not very differentiated. These estimates are quite important, since it is known that also admixture of poorly differentiated populations can affect the invasions success. (Dlugosch et al. 2015) Some native populations might be genetically more homogeneous if there are higher levels of gene flow or homogeneous environments due to human activities. Genomic approaches can solve this problem with their higher resolution, especially when large numbers of individuals are sampled. (Cristescu 2015) The informative value in the determination of admixture is comparatively much higher with high-resolution genome wide SNP markers in comparison with low resolution standard genetic markers. These can be stated with F-statistics which test for admixture events (F3-statistics) and admixture proportions (F4-statistics) (Patterson et al. 2012). The simulation tool ABC can also be applied on these large SNP datasets and can reveal the best fitting simulation of population history scenarios (Pudlo et al. 2016). Additionally, population and quantitative genomic approaches based on these HTS and genotyping methods can model the evolution of traits in natural populations (Wray 2013). These methods facilitate dense population genomic surveys which can resolve invasion history and bottleneck events (Estoup et al. 2010; Fitzpatrick et al. 2012). Major evolutionary processes (selection, gentic drift etc.), that shape demography (e.g. bottlenecks) and genomic structure during invasions can be investigated in large scales (Catchen et al. 2013; Fitzpatrick et al. 2012; Puzey and Vallejo-Marín 2014). With genome-wide data, the confidence and precision in the identification of species can be improved and additionally enable phylogenomic approaches to identify evolutionary relationships (e.g. Batista et al. 2016; Dowle et al. 2017; Picq et al. 2018; Storer et al. 2017).

3.6 Objectives

The Asian long-horned beetle (ALB), *Anoplophora glabripennis* (3.2), has been frequently found in Europe and North America in the last decades, clearly showing that more methods need to be applied to avert the global spread. Extensive knowledge on the introduction routes, as well as secondary dispersal is indispensable for effective control measures by local plant protection offices for prevention, eradication or containment of outbreaks. To achieve this improved understanding, this study aimed to investigate the intraspecific genetic differences of 196 samples from 22 European infestation sites in Germany, Italy and Switzerland, as well as three samples from China. The specimens have been provided from regional and federal plant protection institutes which are responsible for conducting the control measures. The samples originated from both already eradicated sites as well as from still active infestations. The German and Italian sites have not been studied on the intraspecific genetic differences in this dimensions so far, but the sites from Switzerland have already been investigated using

microsatellites (Tsykun et al. 2019). This study will give a first-time high-resolution contribution to the reconstruction of single outbreaks in Germany and Italy.

Following the population genetic processes of different invasion scenarios (e.g. genetic bottlenecks, preadaption, admixture), two assumptions were made on how to determine the introduction patterns into Europe.

First, with the detection of genetic differences between the different sampling sites in Europe caused by non-random mating, hence population structure, conclusions on their kinship can be drawn. **Second,** the introduction and dispersal patterns can be deviated by the kinship between the collection sites.

Hence, the discovery or lack of cross-regional and cross-country genetic differences can uncover spread promoting behaviour of humans like primary dispersal via wood-packing material and secondary dispersal to neighbouring locations, since the ALB is not a far distant flyer (Javal et al. 2018; Lopez et al. 2017). If the regions are strongly genetically differentiated from each other, they originate most likely from independent imports, either from native or other invasive areas. If regions are genetically related to each other, there are two possible explanations. Either it is the same source of introduction, or the low genetic differentiation is caused by recent secondary dispersal events.

For this study classical mitochondrial DNA-markers (3.3) were used for a first genetic characterization. The impact of nuclear mitochondrial-like sequences (NUMTs, 3.4) on DNA barcoding of ALB needed to be investigated due to disturbances that occurred during mitochondrial DNA-marker development (5.3.1). Additionally, genome wide single nucleotide polymorphisms (SNPs), which were obtained by Genotype-by-sequencing (GBS, 3.5) were used for the detailed population genomic insight. The DNA of the sampled regions were Sanger-sequenced, and Illumina sequenced for the described approaches. For the Sanger-sequencing, PCR products of the mitochondrial cytochrome oxidase subunits 1 and 2 were used and those of three NUMTs candidates, too. The determined haplotypes, thus the variants of the gene sections, were compared with other studies and unpublished data available in the databases BOLD and NCBI. The genome wide SNPs were gained by Nextseq500 Illumina sequencing and genotyped for each sample (GBS).

By combining two different approaches to study the population structure, aspects of basic research as well as applied methods for plant health controls could be covered. By improvements on DNA barcoding methods and databases, a low-resolution insight into the population structure is gained, which can give indications on single outbreak sources with the help of public databases, but no clear assignments yet. With the high throughput approach,

high-resolution insights in the population structure of the ALB are gained for the first time in European infestation areas. Integration of population genomic approaches in the considerations of federal plant health services is an inevitable state of the art strategy to effectively disrupt the routes and secondary dispersal, as well as to identify the characteristic introduction sources. Consequently, this study can supply recommendations for federal plant health services on how to use high-resolution insight into the introduction history in European sites, not only for future introductions of ALB, but also for other threats by invasive beetles with a similar mode of life (e.g. pests recommended by EPPO for regulation as a quarantine pest; A1 list (not present in EPPO region) *Apriona germari*, A2 list (present in EPPO region) *Aromia bungii*, *Anoplophora chinensis*; EPPO 2021c).

The results depicted in this dissertation will be grouped in three different chapters. **First** will be on the prospects and limits of COI-Barcoding practice to characterize the biological invasions of Asian long-horned beetle (ch. A). The **second** chapter will address the discovery of nuclear mitochondrial-like sequences in invasive Asian long-horned beetle and their impact on DNA barcoding (ch. B). The **third** one will cover the complex European invasion history of the Asian long-horned beetle which will be described based on new insights in its population genomic differentiation using Genotype-by-sequencing (ch. C).

4 Materials and methods

4.1 Chemicals and solutions

If not mentioned else, all chemicals were obtained in analytical purity level from the following companies and corporations:

Roth	Carl Roth GmbH + Co. KG, Karlsruhe BW, Germany
Thermo Fisher Scientific Inc.	Thermo Fisher Scientific Inc., Waltham MA, USA
VWR	VWR International GmbH, Darmstadt HE, Germany

Table 1. Nutritional media, stock solutions and buffers – molecular biological methods (1/2).

Solution identifier	Composition	Final concentration	Ch.
TFB1 solution	RbCl	100.0 mM	4.4.7
	$MnCl_4 \cdot H_2O$	50.0 mM	
	Potassium acetate (CH_3COOK)	30.0 mM	
	glycerol	15.0 % (v/v)	
	Set pH to 5.8 with acetic acid. Filter sterilize.		
TFB2 solution	MOPS	10.0 mM	4.4.7
	RbCl	10.0 mM	
	$CaCl_2 \cdot 2H_2O$	75.0 mM	
	glycerol	15.0 % (v/v)	
	Set pH to 6.8 with NaOH and autoclave (121°C, 1 bar, 20 min).		
LB medium (liquid)	tryptone	1.0 % (w/v)	4.4.7, 4.4.8
	yeast extract	0.5 % (w/v)	
	NaCl	0.5 % (w/v)	
	Autoclave (121°C, 1 bar, 20 min) the medium before usage. Store at room temperature.		
LB medium (solid)	agar-agar	1.5 % (w/v)	4.4.8
	Add the agar-agar to liquid LB medium. Autoclave (121°C, 1 bar, 20 min). Pour medium into petri dishes. Store at 4°C.		
LB_{Amp} medium (liquid)	ampicillin	100.0 µg/ml	4.4.8
	Allow the liquid LB medium to cool down to 50°C before adding 1:1000 volume of 100 mg/ml ampicillin sterile stock solution to a final concentration of 100µg/ml.		
$LB_{Amp/IPTG/X\text{-}\beta\text{-}Gal}$ medium (solid)	ampicillin	100.0 µg/ml	4.4.8
	IPTG	0.5 mM	
	X-β-Gal	80.0 µg/ml	
	Allow the LB medium (solid) to cool down to 50°C after autoclaving, before supplement while it is still liquid with 100µg/ml ampicillin, 0.5 mM IPTG and 80µg/ml X-β-Gal. Pour medium into petri dishes. Store at 4°C.		

Table 1. Nutritional media, stock solutions and buffers – molecular biological methods (2/2).

Solution identifier	Composition	Final concentration	Ch.
SOC medium (liquid)	tryptone	2.0 % (w/v)	4.4.8
	yeast extract	0.5 % (w/v)	
	NaCl	10.0 mM	
	KCl	2.5 mM	
	Mg^{2+}	20.0 mM	
	glucose	20.0 mM	
	Add tryptone, yeast extract, NaCl and KCl to distilled water. Autoclave (121°C, 1 bar, 20 min) and cool to room temperature Add filter sterilized 2M Mg^{2+} and 2M glucose stock solutions. Bring to final volume with sterile, distilled water. Aliquot and store at –20°C.		
Ampicillin stock solution	ampicillin	100.0 mg/ml	4.4.8
	Filter-sterilize the solution, aliquot and store at –20°C. The working solution is 100µg/ml (dilution factor 1:1000).		
IPTG stock solution	IPTG	0.1 M	4.4.8
	Filter-sterilize the solution, aliquot and store at –20°C. The working solution is 0.5 mM (dilution factor 1:200).		
X-β-Gal stock solution	X-β-Gal	25.0 mg/ml	4.4.8
	Dissolve in DMSO. Cover with aluminium foil and store at –20°C. The working solution is 80 µg/ml (dilution factor 1:625).		
Mg^{2+} stock solution	$MgSO_4 \cdot H_2O$	2.0 M	4.4.8
	Weigh 13.84g of $MgSO_4 \cdot H_2O$ and add distilled water to 100ml. Filter sterilize.		
Genomic DNA extraction buffer	tris	100.0 mM	4.4.1
	EDTA	20.0 mM	
	NaCl	1.4 M	
	CTAB	2.0 % (w/v)	
	β-mercaptoethanol	0.2 % (v/v)	
	Add β-mercaptoethanol freshly before every usage to an appropriate amount of buffer (e.g. 2 µl/ml buffer), whereas the rest is stored in 4°C.		
TBE electrophoresis buffer stock solution (5-times)	tris	450.0 mM	4.4.3
	boric acid	450.0 mM	
	EDTA	12.0 mM	
	The working solution is 0.5-times (dilution factor 1:10).		

4.2 Hardware and software

Table 2. Applied hardware for all molecular biological work.

Hardware	Model	Company/Inc.
Autoclave (L)	Systec VE-150	Systec GmbH, Linden, Germany
Autoclave (S)	Systec DB-23	Systec GmbH, Linden, Germany
Balance (fine)	Kern PCB 250-3	KERN & SOHN GmbH, Balingen, Germany
Balance	Kern PCB 6000-1	KERN & SOHN GmbH, Balingen, Germany
Centrifuge (L)	Eppendorf Centrifuge 5810	Eppendorf AG, Hamburg, Germany
Centrifuge (M)	Heraeus Fresco 17	Thermo Fisher Scientific Inc., Waltham MA, USA
Centrifuge (S)	VWR MiniStar Silverline	VWR International GmbH, Darmstadt HE, Germany
Clean bench	Clean Air DLF/BSS 4	CAE Innovative Engineering GmbH, Bielefeld, Germany
Fluorometer	Qubit 2.0 Fluorometer	Thermo Fisher Scientific Inc., Waltham MA, USA
Gel documentation	SynGene Ingenius LHR	SynGene UK, Cambridge, UK
Gel electrophoresis unit	Biozym EasyPhor System	Biozym Scientific GmbH, Hessisch Oldendorf, Germany
Gel electrophoresis power supply	Consort EV 245	Consort bvba, Turnhout, Belgium
Incubator (L)	Memmert Brutschrank	Memmert GmbH + Co. KG, Schwabach, Germany
Incubator (S, shaker)	Stuart Orbital Incubator SI500	Cole-Parmer, Staffordshire, UK
Magnetic stirrer	neoLab D-6010	neoLab Migge GmbH, Heidelberg, Germany
Microwave	SEVERIN 900&Grill	SEVERIN Elektrogeräte GmbH, Sundern, Germany
pH-Meter	Mettler-Toledo FiveEasy FE20	Mettler-Toledo GmbH, Gießen, Germany
Photometer	Eppendorf Bio Photometer	Eppendorf AG, Hamburg, Germany
Pipets	Eppendorf Research plus	Eppendorf AG, Hamburg, Germany
Spectrophotometer	NanoDrop 2000/2000c	Thermo Fisher Scientific Inc., Waltham MA, USA
Thermocycler	Mastercycler pro	Eppendorf AG, Hamburg, Germany
Thermomixer	Ditabis MKR 13	Ditabis AG, Pforzheim, Germany
Ultrapure Water Purification System	Purelab Option-R7	ELGA LabWater, High Wycombe, UK
UV table	UV Tisch	Bachofer GmbH & CO. KG Laboratoriumsgeräte, Reutlingen, Germany
Vortexer	Vortex-Genie 2	Scientific Industries Inc., New York, USA
Water bath	Thermostatic Waterbath	Fried Electric, Haifa, Israel

Table 3. Applied software for molecular biological work, all *in silico* data analysis, inference, visualization and publication of this study.

Software	Function	OS/Server	Reference/source
ADMIXTURE v. 1.3.0	Ancestry estimation software	ubuntu 18.04 LTS/bwUnicluster2	Alexander et al. (2009)
Adobe Photoshop CC	Image processing	Windows 10	Adobe Inc., San José, CA, USA
Arlequin v. 3.5	Population genetics software	Windows 10	Excoffier and Lischer (2010)
Citavi v. 6.8	Reference management software	Windows 10	Swiss Academic Software GmbH, Wädenswil, Switzerland
CLC Main Workbench v. 7.5.3	Advanced DNA, RNA, and protein analysis package	Windows 10	QIAGEN, Hilden NRW, Germany
DnaSP v. 6.12.01	DNA polymorphisms analysis software package	Windows 10	Rozas et al. (2017)
Geneious Prime® v. 2020.2.4. - v. 2021.0.1	Bioinformatics software for sequence data analysis	Windows 10	Biomatters, Auckland, New Zealand
GeneSnap	Gel documentation	Windows XP	SynGene UK, Cambridge, UK
JMP Pro v. 15	Statistic software	Windows 10	SAS Institute Inc., Cary, NC, USA
MEGA X	Molecular evolutionary genetics analysis	Windows 10	Kumar et al. (2018)
NanoDrop 2000	Spectrophotometer	Windows XP	Thermo Fisher Scientific Inc., Waltham MA, USA
PGDSpider v. 2.1.1.5	Data conversion tool	Windows 10	Lischer and Excoffier (2012)
PLINK v. 1.9	Whole genome data analysis toolset	ubuntu 18.04 LTS/ubuntu on WSL	Purcell and Chang (2021), Chang et al. (2015)
PopART v. 1.7	Population analysis with reticulate trees	Windows 10	Leigh and Bryant (2015)
R v. 3.6.1	Statistical computing and graphics	Windows 10	R Core Team (2021)
Splitstree v. 4.17.1	Computing unrooted phylogenetic networks	ubuntu 18.04 LTS	Huson and Bryant (2006)
VCFtools v. 0.1.15	Variation genome data analysis toolset	ubuntu 18.04 LTS/ubuntu on WSL	Danecek et al. (2011)

4.3 Sampling of Asian long-horned beetle (ALB)

All Asian long-horned beetle (ALB) samples analysed in this study were acquired from invasive European sites in Germany, Italy, and Switzerland, except three samples used for comparisons from the native range in China. The ALB samples were collected mostly from infestation sites in different years (2011–2019) and from members of federal plant health or research institutes. The samples were partly stored over several years before processed in this study. These ALB samples consist of different developmental stages (larva, pupa, or imago) and various conditions of conservation (dried, fixed in ethanol and/or frozen in –20°C). Some samples from Magdeburg (Germany) and the samples from Switzerland and China were directly obtained as DNA extracts. After arrival in Hohenheim, all samples, except the dried and pinned specimen, were preserved in 80% Ethanol and stored in –20°C further on. A list of all sample information provided by the federal plant health or research institutes about the collection years, developmental stages, the storing conditions before the arrival at the University of Hohenheim (UHO), as well as source and host information is available in the chapter Appendix I, **Table 20** on population identifier (Pop-ID) level.

A total of 199 ALB specimens were used in this study. These were obtained from 25 collection sites from Switzerland (CH), China (CN), Germany (D) and Italy (I), from which all European ones are considered as within the invasive range of ALB. Four populations from Switzerland are represented in this study, whereas the two locations CH-FR-MA and CH-FRS-BR are close to each other. Three single native specimens from three Chinese locations were used as representatives for common native infestation areas, but not as a robust dataset. The most collection sites were from Germany from 11 different sites, from which ten were from infestation sites and one specimen from D-BW-KA was obtained from an import control. Of the ten German infestation sites, only one was considered as eradicated to the time of the studies start (D-BY-NKI), whereas most sites in Baden-Württemberg, Bavaria and Saxony-Anhalt were still quite active and recent by the time of the studies start (2017). Finding sites in Bavaria from the locations D-BY-FK, D-BY-FKR and D-BY-NBB were in close proximity. From Italy, ALB specimens from seven collection sites were received, also from more recent and still active infestation sites to the time when the studies started. The locations I-MIL-CB and I-MIL-VI are also very close. These 25 sample collection sites were referred to as populations in the following. Further details on the Pop-IDs and positions for each of the sampling sites are compiled in **Table 4** and the locations are shown in **Figure 11**. The latitude and longitude (Lat Long) listed and shown are the geographical coordinates in decimal degrees from the middle of the locations, not of the precise finding spots. The Pop-IDs are composed of abbreviations of the country, the state or province and the location name (shown in brackets). The sample-IDs consist of the Pop-ID, the last two digits of the collection year and additionally a unique

sample-no. at the end. The location Feldkirchen Riem (D-BY-FKR) was added to D-BY-FK Pop-ID in all data analysis due to its proximity, but sample-IDs include the location specification D-BY-FKR. An additional Table with all individual sample-IDs and the corresponding information about sequenced PCR products and DNA-markers is also available in the chapter Appendix I, **Table 21** on specimen level.

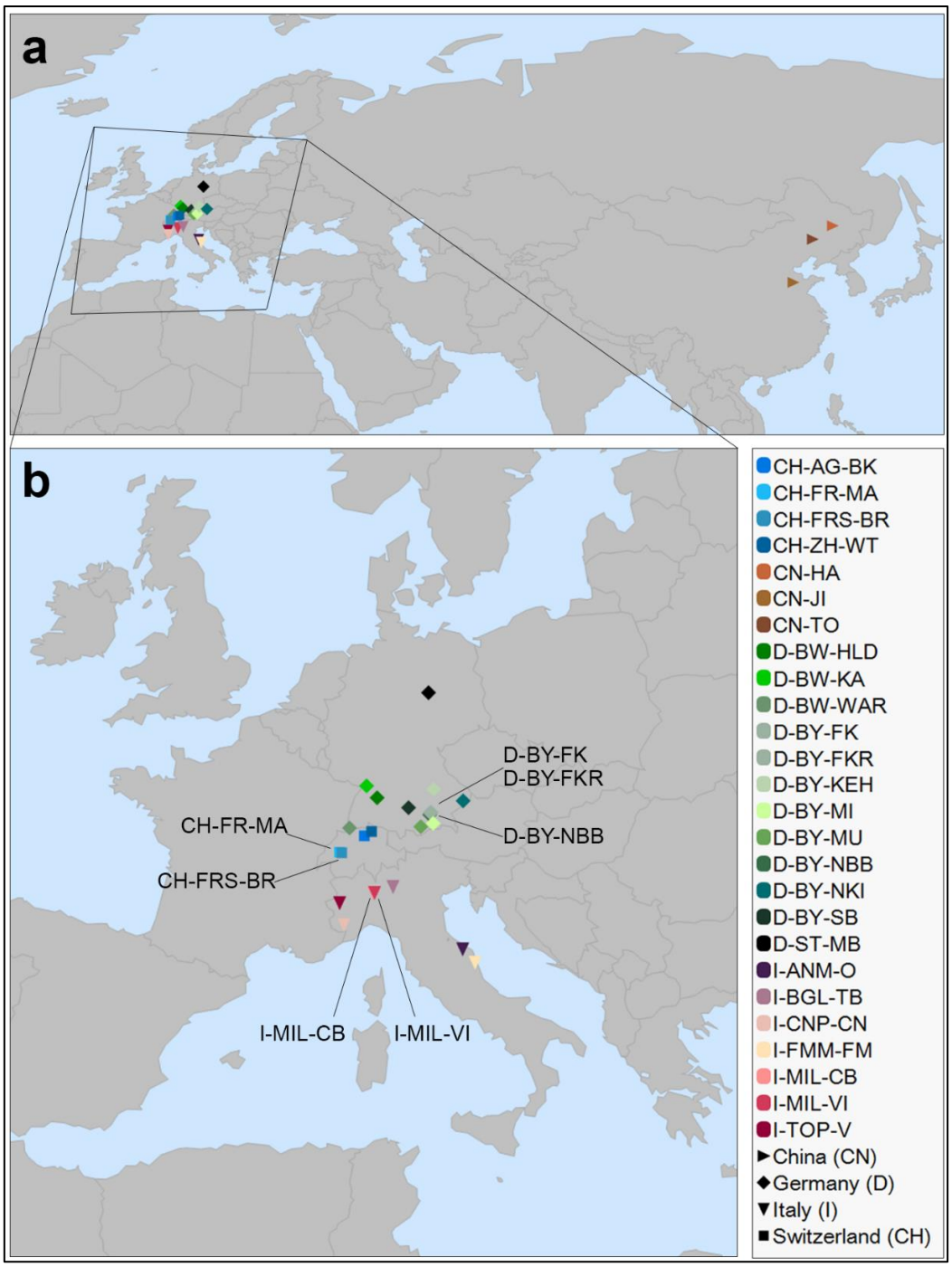


Figure 11. All collection sites of ALB from Europe and Asia used in this study.
a) European ALB collection sites from the invasive range and three Chinese reference specimens from the native range. **b)** 196 European ALB specimens that originated from 22 European ALB collection sites in Germany, Switzerland, and Italy.

Table 4. ALB collection sites (alphabetically) used in this study. – (1/2)

Country	State / province	Location	Lat Long	Year(s)	Pop-ID	n
Switzerland (CH)	Aargau (AG)	Berikon (BK)	47.349900, 8.373270	2015	CH-AG-BK	2
Switzerland (CH)	Freiburg (FR)	Marly (MA)	46.776089, 7.164610	2014	CH-FR-MA	12
Switzerland (CH)	Freiburg Sense (FRS)	Brünsried (BR)	46.759570, 7.278770	2013	CH-FRS-BR	5
Switzerland (CH)	Zürich (ZH)	Winterthur (WT)	47.499880, 8.726160	2012	CH-ZH-WT	2
China (CN)	Heilongjiang	Harbin (HA)	45.803776, 126.534966	2016	CN-HA	1
China (CN)	Shangdong	Jinan (JI)	36.651199, 117.120094	2016	CN-JI	1
China (CN)	Inner Mongolia	Tongliao (TO)	43.621910, 122.268051	2016	CN-TO	1
Germany (D)	Baden-Württemberg (BW)	Hildrizhausen (HLD)	48.625221, 8.966090	2016	D-BW-HLD	15
Germany (D)	Baden-Württemberg (BW)	Karlsruhe (KA)	49.006889, 8.403653	2016	D-BW-KA	1
Germany (D)	Baden-Württemberg (BW)	Weil am Rhein (WAR)	47.592419, 7.613150	2015	D-BW-WAR	9
Germany (D)	Bavaria (BY)	Feldkirchen (FK)	48.147380, 11.730140	2013, 2014	D-BY-FK	23
Germany (D)	Bavaria (BY)	Feldkirchen Riem (FKR)	48.140630, 11.683950	2016	D-BY-FK	4
Germany (D)	Bavaria (BY)	Kelheim (KEH)	48.918610, 11.872230	2016	D-BY-KEH	10
Germany (D)	Bavaria (BY)	Miesbach (MI)	47.789420, 11.833550	2019	D-BY-MI	7
Germany (D)	Bavaria (BY)	Murnau (MU)	47.677410, 11.200950	2017	D-BY-MU	10
Germany (D)	Bavaria (BY)	Neubiberg (NBB)	48.076910, 11.657100	2014	D-BY-NBB	15
Germany (D)	Bavaria (BY)	Neukirchen am Inn (NKI)	48.519482, 13.372620	2012	D-BY-NKI	2
Germany (D)	Bavaria (BY)	Schönebach (SB)	48.308160, 10.578670	2015, 2018	D-BY-SB	11

Table 4. ALB collection sites (alphabetically) used in this study. – (2/2)

Country	State / province	Location	Lat Long	Year(s)	Pop-ID	n
Germany (D)	Saxsony-Anhalt (ST)	Magdeburg (MB)	52.131672, 11.640320	2015, 2018, 2019	D-ST-MB	35
Italy (I)	Ancona-Marche (ANM)	Ostra (O)	43.614850, 13.159240	2017	I-ANM-O	2
Italy (I)	Bergamo-Lombardia (BGL)	Trescore-Balneario (TB)	45.694910, 9.846540	2017, 2018	I-BGL-TB	8
Italy (I)	Cuneo-Piemonte (CNP)	Cuneo (CN)	44.384476, 7.542671	2018	I-CNP-CN	3
Italy (I)	Fermo-Marche (FMM)	Fermo (FM)	43.158875, 13.720088	2015	I-FMM-FM	2
Italy (I)	Milano-Lombardia (MIL)	Corbetta (CB)	45.469566, 8.918932	2016, 2017	I-MIL-CB	8
Italy (I)	Milano-Lombardia (MIL)	Vittuone (VI)	45.485081, 8.958650	2014, 2016	I-MIL-VI	7
Italy (I)	Torino-Piemonte (TOP)	Vaie (V)	45.101662, 7.289800	2018	I-TOP-V	3

n is the number of specimens per collection site (Pop-ID).

4.4 Molecular biological methods

Table 5. Kits used for molecular biological methods.

Function	Kit	Content	Company/Inc.
Cloning	pGEM®-T and pGEM®-T Easy Vector Systems	2X Rapid Ligation Buffer (T4 DNA Ligase), T4 DNA Ligase (100 u), JM109 Competent Cells, High Efficiency (6 × 200µl), pGEM®-T Vector (50ng/µl), Control Insert DNA (4ng/µl)	Promega Inc., Fitchburg WI, USA
Fluorometer measurement	Qubit® dsDNA BR Assay Kit	Qubit® BR Buffer, BR Reagent, BR Standards	Thermo Fisher Scientific Inc., Waltham MA, USA
Gel extraction	MicroElute Gel extraction Kit	Binding Buffer, Wash Buffer, Elution Buffer, DNA columns	Omega Bio-Tek Inc., Norcross GA, USA
PCR	DreamTaq-Polymerase Kit	DreamTaq-Polymerase, DreamTaq-GreenBuffer	Thermo Scientific, Waltham MA, USA
PCR product purification	QIAquick PCR purification Kit	Binding Buffer, Wash Buffer, Elution Buffer, QIAquick columns	QIAGEN GmbH, Hilden NRW, Germany
Plasmid extraction	Plasmid Miniprep Kit I peqGOLD	Solution I, II, III, Plasmid Buffer, Wash Buffer, Elution Buffer, DNA columns	VWR International GmbH, Darmstadt HE, Germany

4.4.1 Genomic DNA preparation from ALB tissue

The different qualities and storing conditions of the tissue samples were the most challenging part in subsequent DNA extraction. After evaluating different DNA extraction protocols (not listed here in detail), the CTAB protocol according to the description in Rusterholz et al. (2015) was identified to be most suitable for the quality and amount of total DNA. A precooled metal rack (–80°C) was used to freeze down the tissue sections (approx. 5-10 mg) for mechanical cell disruption using a micro pestle. Furthermore, this protocol easily allowed certain modifications, depending on the conditions of the tissue sample. Thus, e.g. dried samples were incubated overnight instead of 90 min in the extraction buffer (4.1). For larval tissue, the extraction of the suspension with 500 µl chloroform/isoamyl alcohol (25:1) by inversions was repeated. In general, for adult beetles one leg was used for DNA extraction, if available. Otherwise, some tissue parts of the thorax were used. From larvae a small tissue section was used, avoiding the inclusion of intestine parts. The quality and quantity of the genomic DNA samples was assessed as described in chapter 4.4.2.

4.4.2 Quantification and quality control of DNA samples

Qubit® dsDNA BR Assay Kit (**Table 5**) was used to determine the quantity of the genomic DNA extracts from beetle tissue in ng/µl with a Qubit Fluorometer, following the manuals instructions. The quality was assessed by measuring the 260/280 and 260/230 ratios using a Nanodrop 2000/2000c Spectrophotometer. The molecular level of genomic DNA (high or low) was determined by visual check after agarose gel electrophoresis (4.4.3).

For plasmid DNA (4.4.9), Nanodrop 2000/2000c Spectrophotometer was used to quantify the concentration in ng/µl. For the evaluation of quality, 260/280 and 260/230 ratios were used. For some PCR products, the quantity was roughly estimated by visual check after agarose gel electrophoresis (4.4.3), whereas the quality was also determined by measuring the 260/280 and 260/230 ratios using Nanodrop 2000/2000c Spectrophotometer.

4.4.3 Agarose gel electrophoresis

Separation of DNA by size was performed via agarose gel electrophoresis for quality assessment of genomic DNA, to verify the fragment length after PCR reactions, as well as the estimation of the PCR product quantity after purification. A 0.5 times concentrated TBE buffer (4.1) served as running buffer, likewise for production of agarose gels. 1 % (w/v) TBE-agarose gels were applied for separation of genomic DNA and 1.5 % (w/v) TBE-agarose gels for PCR product separations. Therefore, the appropriate agarose amount was dissolved in 0.5-times TBE buffer in the microwave, and poured into a gel tray after cooling down to 50°C. Once solidified, the agarose gel was transferred into the electrophoresis unit and covered by a

running buffer. Genomic DNA samples (4.4.1) and purified PCR products (4.4.6) were mixed in a 1:6 proportion of DNA Gel Loading dye (6-times) (Thermo Fisher Scientific Inc., Waltham MA, USA) and loaded into the sample wells in the gel. Not purified PCR reactions were directly loaded to the wells, since the DreamTaq-GreenBuffer (**Table 5**) already contains green loading dye. GeneRuler 100 bp Plus DNA Ladder, or 100 bp+, (Thermo Scientific Inc., Waltham MA, USA) served as a molecular weight ladder into the first lane of the gel. Small gels were run at 60-90 V for 40-60 min and big gels at 160-190 V for 40-60 min, depending on the approach. If the agarose gel electrophoresis was performed for separating multiple bands from a PCR reaction, there was always a big gel applied for a subsequent gel extraction. For this purpose, the voltage was decreased to 150 V and the running time was extended to 3 h. For the visualization of the DNA bands in the agarose gel, post staining with a 3-times staining solution GelRed™ (VWR International GmbH, Darmstadt HE, Germany) in 0.1 M NaCl for 15-30 min was used. The gels were documented with SynGene Ingenius LHR gel imaging system. Gels for subsequent gel extraction (4.4.6) were put on a UV table instead to cut the single fragments.

4.4.4 Nuclear and mitochondrial single PCR products

All nuclear and mitochondrial single DNA sequences of interest were first amplified by PCR (4.4.5), subsequently purified (4.4.6) and then Sanger-sequenced (4.4.11). Primer design was performed with the help of CLC Main Workbench, Oligo Calculator (2021) and Multiple Primer Analyzer (2021). Subsequently the primers were checked via Primer-BLAST (2021) from Ye et al. (2012) against the reference scaffold genome of *Anoplophora glabripennis*. Oligonucleotides used for PCR amplification were obtained in lyophilized form from Eurofins Genomics (Ebersberg BY, Germany) and solved in sterile ddH_2O (Carl Roth GmbH + Co. KG, Karlsruhe BW, Germany) to receive a stock concentration of 100 µM. The final concentration of all primers in the reaction mix was 0.2 µM. The primer pairs, their corresponding sequences and respective PCR products are summarized in **Table 6**.

For each of the 199 specimens of ALB from 22 European and 3 Chinese collection sites (see Pop-IDs, **Table 4**), exact information about the respective sequenced PCR products/DNA-markers is given in the Appendix chapter I, **Table 21**. There were two single PCR products designed to amplify sections from mitochondrial origin and 9 single PCR products from nuclear origin, from which just 7 were Sanger-sequenced.

For COI-barcoding, a 709 bp fragment of the mitochondrial genome with the standard primers LCO-1490 and HCO-2198 according to Folmer et al. (1994) was used to amplify the position 1427 to 2135 on the NCBI Reference Sequence NC_008221.1 (NCBI 2021), corresponding to other ALB studies (Javal et al. 2019a; Kim et al. 2019; Tsykun et al. 2019). This PCR product was amplified for 150 specimens from four Swiss, three Chinese, nine German and seven

Italian populations. The only collection sites not represented by the mtCOI DNA-marker were D-BW-KA and D-BY-NKI.

ALB specific primers for the ALB-COI-II PCR product amplification were designed from position 1811 to 3081 on the mitochondrial reference genome NC_008221.1 to amplify a 1271 bp fragment of the partial cytochrome oxidase subunit I to the beginning of the second subunit of cytochrome oxidase, including the intermediate tRNA-Leu region between the genes for the subunits. For the amplification of ALB-COI-II, the primers mtALB/COX1_F and mtALB/COX2_R2 were applied on 55 specimens from seven German populations. The collection sites not represented by the ALB-COI-II PCR product were all sites from Switzerland, China, and Italy, as well as the German sites D-BY-KEH, D-BY-MI, D-BY-MU and D-BY-NKI.

The 9 ALB-specific nuclear PCR products from three different nuclear mitochondrial-like sequence (NUMT) regions, were amplified of 22 specimens from nine German and seven Italian populations with at least one specimen (013, 104, 194, 095, 085, 047, 170, 174, 124, 120, 180, 181, 122) per Pop-ID as representative. For sites with findings over several years from Germany like D-BW-WAR (189, 024), D-BY-FK (074, 058, 076), D-ST-MB (038, 129, 149, 164), more representatives were selected. The sample-no. shown here in brackets refer to the exact specimens per NUMT sequences in Appendix I, **Table 21** (blue). The collection sites not represented by the ALB NUMT PCR products were all sites from Switzerland and China, as well as the German sites D-BW-KA and D-BY-NKI.

For each NUMT region, primers were designed in the flanking regions up- (f = front end) and downstream (b = back end) of the NUMT, as well as internal (int = internal). The shorter NUMT ALBsc1709-Numt2 could be amplified and sequenced completely, whereas the other two NUMTs needed to be sequenced with three PCR products of the front, internal part and back end, respectively. For forward and reverse primers, the numbers in brackets (see **Table 6**) refer to their position upstream (-) or downstream (+) of the NUMT, whereas the numbers of internal, mid-forward or mid-reverse primers refer to the exact position on the NUMT itself.

For ALBsc158-Numt1 on ALBscaffold158, GenBank accession no. KZ486723.1, three PCR products were amplified and Sanger-sequenced. The amplification of the front end and the upstream flanking region ALsc158-Numt1(f), the primer pair ALBsc158_Numt1(-99)_F and ALBsc158_Numt1(745)_MR was used. The amplification of the internal part of ALsc158-Numt1(int) was done by the primer pair ALBsc158_Numt1(int)_F and ALBsc158_Numt1(int)_R. Furthermore, the primer pair ALBsc158_Numt1(726)_MF and ALBsc158_Numt1(+137)_R was utilized for the amplification of the back end and the downstream flanking region ALsc158-Numt1(b). On ALBscaffold1709, GenBank accession no. KZ488270.1, two NUMT regions 4081 bp apart from each other were amplified via PCR and

Sanger-sequenced. ALBsc1709-Numt2 PCR products of the front end and the back end, as well as the downstream and upstream flanking regions were checked via PCR, but not sequenced. The amplification of the front end and the upstream flanking region was done by using the primer pair ALBsc1709_Numt2(-70)_F and ALBsc1709_Numt2(156)_MR, whereas the back end and the downstream flanking region of ALBsc1709-Numt2 was amplified with the primer pair ALBsc1709_Numt2(136)_MF and ALBsc1709_Numt2(+89)_R. The whole ALBsc1709-Numt2 with its flanking regions was amplified by using the primer pair ALBsc1709_Numt2(-70)_F and ALBsc1709_Numt2(+89)_R. For the amplification of ALBsc1709-Numt3, three PCR products were employed. The front end ALBsc1709-Numt3(f) was amplified by using the primers ALBsc1709_Numt3(-160)_F2 and ALBsc1709_Numt3(793)_MR. For the amplification of the internal part of ALBsc1709-Numt3(int) the primer pair ALBsc1709_Numt3(int)_F and ALBsc1709_Numt3(int)_R was used. The back end and the downstream flanking region of ALBsc1709-Numt3(b) was amplified with the primer pair ALBsc1709_Numt3(774)_MF and ALBsc1709_Numt3(+79)_R.

Table 6. Primers for amplification of nuclear and mitochondrial single PCR products. – (1/2)

PCR product	Position (5')	Primer pair	Sequence (5'-3')	length (bp)
mtCOI	[1]1427	LCO-1490	GGTCAACAAATCATAAAGATATTG	709
	[1]2135	HCO-2198	TAAACTTCAGGGTGACCAAAAAATCA	
ALB-COI-II	[1]1811	mtALB/COX1_F	TTGCACATAGAGGTTCTTC	1271
	[1]3081	mtALB/COX2_R2	AGGAAAGTTGTTCTATTAGAGG	
ALBsc158-Numt1(f)	[2]95525	ALBsc158_Numt1(-99)_F	GTGCGTTCCTTGAAAGAGG	844
	[2]96368	ALBsc158_Numt1(745)_MR	GGAAATGCTATATCTGGAGC	
ALBsc158-Numt1(int)	[2]95765	ALBsc158_Numt1(int)_F	TTGTCTTCACACTCATCAC	747
	[2]96511	ALBsc158_Numt1(int)_R	CTGAAGAACCTCTATGTGC	
ALBsc158-Numt1(b)	[2]96349	ALBsc158_Numt1(726)_MF	GCTCCAGATATAGCATTTCC	1394
	[2]97742	ALBsc158_Numt1(+137)_R	TGGAGTAAATCGTCTGTTGG	
ALBsc1709-Numt2	[3]21449	ALBsc1709_Numt2(-70)_F	CTCACCAATTCCGGACAAC	469
	[3]21917	ALBsc1709_Numt2(+89)_R	GTCAAATGGTAGGTTACATGC	
ALBsc1709-Numt2(f)	[3]21449	ALBsc1709_Numt2(-70)_F	CTCACCAATTCCGGACAAC	226
	[3]21674	ALBsc1709_Numt2(156)_MR	CATTAGACGGCTGAAAGTAAG	

Table 6. Primers for amplification of nuclear and mitochondrial single PCR products. – (2/2)

PCR product	Position (5')	Primer pair	Sequence (5'-3')	length (bp
ALBsc1709-Numt2(b)	[3]21654	ALBsc1709_Numt2(136)_MF	CTTACTTTCAGCCGTCTAATG	264
	[3]21917	ALBsc1709_Numt2(+89)_R	GTCAAATGGTAGGTTACATGC	
ALBsc1709-Numt3(f)	[3]25750	ALBsc1709_Numt3(-160)_F2	CAGCAAGATCAATAACAGC	953
	[3]26702	ALBsc1709_Numt3(793)_MR	GCTGGAATAGTTAGAACATC	
ALBsc1709-Numt3(int)	[3]26511	ALBsc1709_Numt3(int)_F	GAGCACCCAGTATTAGTG	505
	[3]27015	ALBsc1709_Numt3(int)_R	TTTCGTTTCTCTGAGTGG	
ALBsc1709-Numt3(b)	[3]26683	ALBsc1709_Numt3(774)_MF	GATGTTCTAACTATTCCAGC	1206
	[3]27888	ALBsc1709_Numt3(+79)_R	GCCAAATCCTACTATTAACG	

Primer sequences are shown in pairs for the amplification of the corresponding PCR product. The length information refers to the length of the corresponding PCR products of mtCOI, ALB_COI-II, the NUMTs ALBsc158_Numt1, ALBsc1709_Numt2, ALBsc1709_Numt3 and their flanking regions. [1]*A. glabripennis* mitochondrion, complete genome, NCBI Reference Sequence NC_008221.1; [2]ALBscaffold158, GenBank accession no. KZ486723.1; [3]ALBscaffold1709, GenBank accession no. KZ488270.1.

4.4.5 PCR amplification

All PCR amplifications of 199 specimen of ALB from 22 European and three Chinese collection sites for two mitochondrial and 9 nuclear single DNA-segments were performed via DreamTaq-Polymerase Kit (**Table 5**) according to the manual's instructions. The standard Master mix composition is shown in **Table 7**. While the 15 µl reactions were applied on validation or test PCRs, the 25 µl reactions were used in double approaches per specimen before PCR product purifications (4.4.6). The programs for the PCR reaction condition during the amplifications are exhibited in **Table 8**. Almost all PCR reactions were performed with the standard program and their appropriate annealing temperature (T_A) in the Mastercycler pro (Eppendorf AG, Hamburg, Germany). For the two DNA-segments ALsc158-Numt1(b) and ALBsc1709-Numt3(int) the program ext60 was used. The annealing temperatures (T_A) were set to 54°C for mtCOI and ALB-COI-II, to 50°C for ALsc158-Numt1(f), ALsc158-Numt1(int), ALBsc1709-Numt2, ALBsc1709-Numt3(f) and ALBsc1709-Numt3(int) and set to 51°C for ALBsc1709-Numt2(b), ALBsc1709-Numt2(f).

Table 7. 1-times Master mix for PCR reactions using 1 µl genomic DNA as template.

Composition	Volume [µl]	Volume [µl]	Final concentration
10-times Dream Taq buffer	2.500	1.500	1-times
[1]dNTPs (2 mM)	2.500	1.500	0.2 mM
[2]Primer forward (10 µM)	0.500	0.300	0.2 mM
[2]Primer reverse (10 µM)	0.500	0.300	0.2 mM
Dream Taq DNA Polymerase (5 U/µl)	0.125	0.075	0.025 U/µl
H_2O (nuclease free)	17.875	10.325	
Total	24.000	14.000	

[1]2 mM dNTP Mix from dATP, dTTP, dGTP, dCTP stocks (100 mM) of Thermo Fisher Scientific Inc., Waltham MA, USA. [2]Primer solutions were 1:10 dilutions from 100 µM stock solutions.

Table 8. PCR conditions used for PCR reactions with Dream Taq (Thermo Fisher Scientific Inc., Waltham MA, USA).

Phase	Temperature [°C]	Time [s]	Cycles
Initial denaturation	94	120	1
Denaturation	94	20	
Annealing	[1]T_A	30	30
Elongation	72	[2]30 or 60	
Final elongation	72	120	1

[1]For the annealing the T_A was set according to the corresponding primer pair. [2]The elongation time was per default set to 30 s (standard), for some PCR reactions the time needed to be increased to 60s (ext60).

4.4.6 PCR product purification

Double approaches of 25 µl of ALB-COI-II PCR products were pooled and purified via ethanol purification. Therefore, each reaction was first charged with 0.1 volume of 3 M Sodium acetate (CH_3COONa) solution (pH 5.5) and then mixed with 2.5 volumes of pre-cooled (–20°C) absolute ethanol by inversions. The precipitation approaches were incubated at least for 15 min on ice and then centrifuged at 17,000 g for 20 min and 4°C. After pouring off the supernatant, the DNA pellets were washed with 300 µl 70 % ethanol by inversions and centrifugation at 17,000 g for 5 min, 4°C. After discarding the supernatant, the DNA pellets were air dried and solved in 15-20 µl of sterile ddH_2O (Carl Roth GmbH + Co. KG, Karlsruhe BW, Germany). The purified PCR products of ALB-COI-II were visually checked by agarose gel electrophoresis (4.4.3) and subsequently cloned (4.4.8). Double approaches of 25 µl of

mtCOI, ALBsc158-Numt1(f), ALBsc158-Numt1(int), ALBsc158-Numt1(b), ALBsc1709-Numt2, ALBsc1709-Numt3(int) and ALBsc1709-Numt3(b) PCR products were pooled and purified using QIAqick PCR purification Kit (**Table 5**). The purification was performed as described in the QuickStart protocol. The purified PCR products were directly sequenced (4.4.11) or subsequently cloned (4.4.8) when direct sequencing of the PCR product was not successful. Due to multiple bands when amplifying ALBsc1709-Numt3(f) PCR product, the bands were separated by slow gel electrophoresis as described in 4.4.3. The band with the expected size of 953 bp was extracted from gel by Micro Elute Gel Extraction Kit (**Table 5**), following the manuals instructions. Gels were put on a UV table to cut the appropriate fragment with a scalpel with the lowest UV exposure time as possible for a few seconds to avoid damage to the DNA. The PCR fragments which were purified via gel extraction were all subsequently cloned (4.4.8) before sequencing (4.4.11).

4.4.7 Production of chemical competent cells of *E. coli* strain JM109

For the transformation of plasmids into *E. coli* strain JM109, chemical competent cells were re-produced with rubidium chloride. The original chemical competent cells were provided by the pGEM®-T and pGEM®-T Easy Vector Systems (**Table 5**). The E. coli strain JM109 has the genotype *recA1, endA1, gyrA96, thi, hsdR17 (rK–, mK+), relA1*, *supE44*, Δ(*lac-pro*AB), [F´, *tra*D36, *pro*AB, *lac*I^{q}ZΔM15] according to the technical manual TM042.

For this purpose, 5 ml LB (4.1) was inoculated with a fresh colony of E. coli strain JM109 and cultivated overnight at 37°C, 200 rpm in a shaking incubator. This pre-culture was used to inoculate 100 ml of LB medium and incubated at 37°C, 200 rpm in a shaking incubator until an OD_{600} of 0.5, which was measures with a photometer. Every following step was performed on ice. Aliquots (1 ml) of the culture in smaller reaction tubes were cooled down on ice and centrifuged for 5 min at 5000 g and 2°C. The supernatant was poured off and the previous step was repeated until the whole culture was centrifuged to a pellet and removed supernatant of the liquid LB medium. Subsequently, 300 µl TFB1 (4.1) were added to each reaction tube and the pellets were carefully resuspended. 900 µl of the resuspended pellets were pooled together, so the whole amount of reaction tubes was reduced. These reaction tubes were incubated for 10 min on ice. All cell suspensions were centrifuged for 5 min at 5000 g and 2°C. The supernatants were discarded, and the reaction tubes dripped of on a paper towel. To each reaction tube 155 µl of TFB2 (4.1) was added and the cell pellets were resuspended carefully. The competent cells were aliquoted (50 µl each) and stored in –80°C.

The transformation efficiency of the chemical competent cells was assessed as described in the technical manual of the pGEM®-T and pGEM®-T Easy Vector Systems.

4.4.8 Cloning of PCR products in *E coli* strain JM109

In case of not successful direct PCR product sequencing (4.4.11), or in case of previous gel extraction of PCR fragments, these PCR products were cloned using the pGEM®-T and pGEM®-T Easy Vector Systems (**Table 5**) and chemical competent cells of JM109 E. coli strain (4.4.7) following roughly the manuals' instructions. The cloned samples are marked as such in Appendix chapter I, **Table 21**, including PCR products of mtCOI, some PCR products of ALsc158-Numt1 and ALBsc1709-Numt3, all PCR products of ALBsc1709-Numt2, all PCR products extracted from gel of ALBsc1709-Numt3(f) and all PCR products of mtALB-COI-II.

For cloning the amplified target PCR product, building constructs of pGEM®-T or pGEM®-T Easy vector and the PCR product as insert was the first step. Since the PCR product concentrations were not measured precisely and no optimization was necessary, there was no need to apply the molecular ratio of insert to vector in this case. To ligate the vector and PCR products respectively, 1 µl of purified PCR products (4.4.6) was gently mixed with 4 µl of ligation standard reaction (final concentrations 5 ng/µl pGEM®-T or pGEM®-T Easy vector, 1-times Rapid Ligation buffer, 0.3 Weiss units/µl T4 DNA Ligase) according to the instruction manual. To produce a maximum amount of transformants, the ligation reaction was incubated overnight at 4°C.

For the transformation of ligation reactions into *E. coli* strain JM109, rubidium chloride competent cells (4.4.7) were used. The competent cells (50 µl per reaction) from –80°C were thawed on ice for 5 min before they were gently mixed with 5 µl ligation reaction respectively. The suspensions were incubated on ice for 20 min and subsequently transformed via heat shock. For this purpose, the cells were set into a water bath at exactly 42°C for 50 s and immediately returned to ice for 2 min afterwards. To each transformation reaction 950 µl of SOC medium (4.1, RT) was added and the cells were incubated for 1.5 h, 150 rpm in a shaking incubator at 37°C for the phenotypic expression of the selection properties of the pGEM®-T vectors. From each transformation culture, two RT equilibrated LBAmp/IPTG/X-Gal plates (4.1) were prepared for cultivation. 100 µl of the transformation cultures were directly plated. After a short centrifugation at 5000 g for 2 min and pouring off most of the supernatant, the remaining pellet of transformed cells was resuspended in the remaining medium, and also plated on another $LB_{Amp/IPTG/X\text{-}Gal}$ plate. Both plates per transformation reaction were cultivated for blue and white screening and further applications (4.4.10).

Plated transformation reaction on $LB_{Amp/IPTG/X\text{-}Gal}$ plates were incubated overnight at 37°C. Afterwards, the plates were visually checked for white colonies, which generally should contain inserts, whereas blue colonies most likely don't (blue/white screening). The presence of insert in the transformed white colonies was checked via colony PCR (4.4.9).

4.4.9 Colony PCR to verify inserts in cloning vectors

Clones inhabiting the insert of interest, identified by blue and white screening as described in the manual, were further checked by colony PCR with the standard cycler conditions and PCR Kit as described above (4.4.5) with annealing temperature of 50°C, using T7 (5'-TAATACGACTCACTATAGGG-3') and SP6 (5'-CATTTAGGTGACACTATAG-3') primers. The PCR reaction composition is shown in **Table 9**.

Table 9. 1-times Master mix for colony PCR.

Composition	Volume [µl]	Final concentration
10-times Dream Taq buffer	1.500	1-times
dNTPs (2 mM)	1.500	0.2 mM
SP6 promotor primer (10 µM)	0.300	0.2 mM
T7 promotor primer (10 µM)	0.300	0.2 mM
Dream Taq DNA Polymerase (5 U/µl)	0.075	0.025 U/µl
H_2O (nuclease free)	11.325	
Total	15.000	

One dip of a colony with a sterile toothpick serves as template and is dipped into the 15 µl PCR reaction mix.

4.4.10 Plasmid-DNA preparation of cloning vectors from *E. coli*

Colonies of positive clones were previously identified via blue and white selection, as well as via colony PCR (4.4.9). Validated white colonies were cultivated for plasmid preparation to multiply the sequencing constructs. Therefore, cultures were prepared in 5 ml LB_{Amp} liquid medium (4.1) in culture tubes and inoculated with the whole remaining white colony by picking with a sterile toothpick or pipette tip. The cultures were incubated overnight at 37°C, 200 rpm in a shaking incubator and used for plasmid preparations. The Plasmid Miniprep Kit I peqGOLD (**Table 5**) was used to extract plasmid DNA of positive clones (4.4.8), following the instruction manual. The plasmids were eluted with 50 µl of the elution buffer. In this way, the sequencing constructs of pGEM®-T::insert were gained from the *E. coli* strain JM109 for subsequent Sanger sequencing (4.4.11).

4.4.11 Sanger sequencing of PCR products

All nuclear and mitochondrial PCR products (4.4.4), either plasmids or purified PCR products, were double strand Sanger-sequenced at Microsynth AG (Balgach, Switzerland) and GATC (Eurofins Genomics, Ebersberg BY, Germany). For this purpose, purified PCR products or plasmids were mixed with sterile ddH_2O (Carl Roth GmbH + Co. KG, Karlsruhe BW, Germany)

to gain a volume of 10 µl with 18 ng per 100 bases for PCR products and 40-100 ng/µl for plasmids, with 2.5 µl of the corresponding 10 µM primer solutions for sequencing at Microsynth AG. The plasmids containing the ALB-COI-II inserts were sequenced at GATC using 5 µl of 80-100 ng/µl plasmid and 5 µl of the respective primer (5 µM). For each of the PCR products or cloned inserts, these reactions were prepared for forward and reverse primers respectively.

4.4.12 Genotype-by-sequencing (GBS) of genomic DNA

Genotype-by-sequencing and SNP calling was conducted by LGC Genomics GmbH (Berlin, Germany) in three different runs using their specific normalised GBS protocol (nGBS) (Arvidsson et al. 2016). Therefore, the high molecular genomic DNA extracts were set to a volume of 30 µl with a concentration of 10-50 ng/µl. The total 183 DNA extracts sent to LGC Genomics GmbH were digested with the RE MslI before library preparation (insert size mean range: ~180-215 bp) and run on a llumina NextSeq 500 for the paired-end sequencing (two times 150 bp). The sample D-BY-SB-047 was sequenced on all three GBS runs and served as internal control.

4.5 *In silico* data analysis methods

4.5.1 Online databases and tools

To compare all new sequenced PCR products with already available and published datasets of *A. glabripennis*, the public databases Barcode of life database (2021) and NCBI (2021) were utilized. All primers were checked via Primer-BLAST (2021) of Ye et al. (2012) against the reference genomes of *A. glabripennis*, mostly prior to their application. Additionally, all Sanger sequenced PCR products were checked via BLAST (2021) against the representative reference genomes of ALB, as well as the whole nucleotide collection. Therefore, the setting megaBLAST was applied, looking for highly similar sequences.

For a global perspective on the current COI-barcoding situation, the 168 sequences received in this study were compared with the total available global dataset from Barcode of life database (2021). 809 published records of COI-5P for *A. glabripennis* were obtained from BOLD, with specimens from 9 countries (China, South-Korea, Canada, USA, Austria, France, Germany, Italy, Switzerland). According to the information stated from BOLD, 801 were mined from GenBank NCBI (2021) and 8 from Canadian Forest Service.

To have an evolutionary perspective on the sequence divergence in mitochondrial COI sequences within the genus *Anoplophora*, all COI sequences available on BOLD public database with the query *Anoplophora* were also mined (*A. beryllina, A. chinensis, A. davidis, A. elegans, A. flavomaculata, A. freyi, A. granata, A. horsfieldii, A. lurida, A. macularia, A. malasiaca, A. nobilis and A. stanleyana*). *Anoplophora* species names were processed and

used as obtained from BOLD for a better reproducibility, despite disagreements among taxonomists (Lingafelter and Hoebeke 2002).

By megaBLAST of ALB-COI-II sequences limited to whole genome sequence (WGS) contigs of *A. glabripennis* (April 2018), several contigs with high percentage identities and query coverages and additionally low E-values were discovered. Hence, the megaBLAST was repeated with partial NC_008221.1 mitochondrial reference sequence from position 1811-3081, covering the ALB-COI-II sequence limited to representative reference genomes of *A. glabripennis* and the whole nucleotide collection. This was performed as well with the sequence of the complete mitochondrial genome.

4.5.2 Quality assessment, data processing and filtering

4.5.2.1Sanger-sequenced single PCR products

From all Sanger-sequenced PCR products and plasmids (mtCOI, ALB-COI-II, ALBsc158_Numt1, ALBsc1709-Numt2 and ALBsc1709-Numt3), consensus sequences were built from forward and reverse reads after quality trimming using CLC Main Workbench v. 7.5.3 (Qiagen, Hilden, Germany) and Geneious prime v. 2020.2.4. (Biomatters, Auckland, New Zealand). For both, PCR-products and plasmids, the PCR product primer sequences were removed from consensus sequences. From cloned PCR products, meta-consensus sequences were formed out of 3-5 clone consensus sequences. Ambiguous nucleotides were called when the direct PCR product sequencing showed clear double peaks in the trace date, which was visually checked and counted as a real heterozygous locus. For cloned sequences, ambiguous nucleotides were called like they were called from Geneious alignment or assembly algorithms for consensus sequence. In addition, the protein coding DNA regions mtCOI and ALB-COI-II were translated into amino acid sequences and checked for frame shifts or sudden aborts.

For the NUMT-sequences, the ambiguous sequence variants were just used in TCS networks and Maximum likelihood trees, if no polymorphic sites are lost due to masking of ambiguous sites. Thus, sequences with too many ambiguous sites were also removed from the dataset and excluded from the analyses. For ALBsc1709-Numt3, consensus sequences out of ALBsc1709-Numt3(f), ALBsc1709-Numt3(int) and ALBsc1709-Numt3(b) consensus sequences were formed for each individual respectively, to cover the whole Numt3 on ALBsc1709 and its flanking regions. For ALBsc1709-Numt2 this was not necessary and for ALBsc158-Numt1 it was not possible.

All consensus sequences from the Sanger sequenced single PCR products were aligned with their appropriate reference sequence from ALBs genomes (mitochondrial and nuclear). The

default settings for the multiple alignments were used from CLC Main Workbench v. 7.5.3 (ALB-COI-II) and Geneious prime v. 2020.2.4-2021.0.1 (mtCOI, ALBsc158_Numt1 (f, int, b), ALBsc1709_Numt2 and ALBsc1709_Numt3 (f, int, b)).

For the COI-5P dataset on a global perspective of *Anoplophora glabripennis*, the 168 sequences from this study and the COI-5P BOLD dataset were aligned in Geneious prime v2020.2.4. The length of the alignment was adjusted to 360 bp from position 1693-2052 on the mitochondrial reference genome NC_008221.1 covered by all sequences.

For the evolutionary perspective on the genus *Anoplophora* of the COI-5P dataset, all 110 sequences of the *Anoplophora* spp. COI-5P-datasets were aligned with the haplotype sequences COI-Type1 to COI-Type8 of the European *A. glabripennis* specimens emerged from this study (Geneious prime v2021.0.1). After adjusting the length to maximal overlap, there were 74 sequences with a total length of 618 bp covering the position 1471-2088 on ALB's mitochondrial reference genome. Thus, a total of 74 *Anoplophora* spp. sequences were used for further analyses, including the 8 COI-haplotypes of ALB samples sequenced for this study and the mitochondrial reference (NC_008221.1).

4.5.2.2Genomic SNP-markers from GBS approach

To obtain genomic SNP-markers via Genotype-by-sequencing approach, after llumina NextSeq 500 paired-end sequencing the reads were pre-processed and subsequently aligned with the reference genome of ALB for calling the variants (SNPs). These steps were conducted by LGC Genomics GmbH (Berlin, Germany). Conducted by LGC Genomics GmbH (Berlin, Germany), first all library groups were demutliplexed using the Illumina bcl2fastq 2.17.1.14 software. When the barcode distances between all libraries on the lane allowed for it, one or two mismatches or Ns were allowed in the barcode read. The library groups were demultiplexed according to their inline barcodes and verification of restriction site. In the inline barcodes, no mismatches or Ns were allowed, but Ns were allowed in the restriction site. Afterwards, sequencing adapter remnants were clipped from all reads. All reads with a final length <20 bases were discarded. With the restriction site filtering, reads with 5' ends not matching the RE site were discarded. The adapter clipped Illumina reads were subsequently quality trimmed by removing the reads containing Ns and trimming of reads at 3'-end to get a minimum average Phred quality score of 20 over a window of ten bases. Again, reads with final length <20 bases were discarded. Single reads were written into a separate file if one read in a pair has been discarded. For all FASTQ files, a FastQC report was created. The subsampled quality trimmed reads were aligned against the reference (*A. glabripennis*: NCBI Agla 2.0; https://www.ncbi.nlm.nih.gov/assembly/GCF_000390285.2) using BWA-MEM version 0.7.12 (http://bio-bwa.sourceforge.net/), outputting a combined alignment for all

samples in coordinate-sorted BAM format for each of the three respective GBS-runs. The variant discovery and genotyping of samples was conducted with Freebayes v1.0.2-16 (https://github.com/ekg/freebayes#readme) by using the following specific parameters:

```
--min-base-quality 10 --min-supporting-allele-qsum 10 \
--read-mismatch-limit 3 --min-coverage 5 --no-indels \
--min-alternate-count 4 --exclude-unobserved-genotypes \
--genotype-qualities --ploidy 2 --no-mnps --no-complex \
--mismatch-base-quality-threshold 10
```

The VCF-files received from LGC Genomics GmbH (Berlin, Germany) from three different runs of nGBS were merged as a first step. Merging of the VCF-files from three sequencing runs of 183 specimens with two internal controls ("185Indv") was performed in VCFtools v. 0.1.15 (Danecek et al. 2011) according to the manual by using `vcf-merge` by first zipping each with `bgzip`, tabulate with `tabix -p`, then merge and unzip with `gunzip`.

To sort all the individuals alphabetically according to the Sample-IDs from the previously merged VCF-files, single VCF-files were generated for each specimen, then `vcf-merge` as described before was repeated with an automated bash-script for all samples at once, that regulates the order in the VCF-file columns. For later analysis steps it was necessary to replace the Accession no. of the 9866 scaffolds and contigs in the VCF-file with e.g. scaffold_1 for KZ487478.1 or contig_10422 for AQHT02010422.1 by using an automated bash-script with `paste | sed` for all sites.

To evaluate the raw data before filtering, outputs were generated with VCFtools v. 0.1.15 for allele frequency (`--freq2 --max-alleles 2`), mean depth per individual (`--depth`), mean depth per site (`--site-mean-depth`), site quality (`--site-quality`), proportion of missing data per site (`--missing-site`) and individual (`--missing-indv`), as well as heterozygosity and inbreeding coefficient (`--het`) according to the VCFtools Manual (2018), (https://vcftools.github.io/man_latest.html). All these parameters were subsequently plotted in R by using the R script 11_filtersPT3.R from José Cerca (2021b), (https://github.com/jcerca/population_genomics/tree/master/SNPcalling_GATK).

According to the results (5.4.1) of the output plots in R, the raw VCF data set was filtered (filter 7, f7) by quality (`--minQ 30`), filter flag (`--remove-filtered-all`) which removes all sites with a FILTER flag other than PASS, minor allele frequency (`--maf 0.01`), minor allele count (`--mac 3`), number of alleles (`--max-alleles 2 --min-alleles 2`), read depth (`--min-meanDP 3 --minDP 3 --max-meanDP 55 --maxDP 55`) and filters on the individual missingness (`--remove <filename>`). The file for removing individuals from the

dataset contained a list of individuals with F_miss values (relative count of missingness) above 0.9. Seven specimens had this criterion fulfilled and were thereby sorted out (D-BW-WAR-11-191, D-BY-MU-17-094, D-ST-MB-19-200, D-ST-MB-19-201, D-ST-MB-15-036, I-MIL-VI-14-185, D-BW-WAR-11-192). In the end, there were 178 specimen (178Indv) left from which two were internal controls. Next step was position filtering (filter 8, f8) to remove SNP clusters (`--thin 10`). This was done to prevent very dependent SNPs with high linkage disequilibrium within a distance of 10 bases. The filtered VCF-file was checked and plotted for allele frequency, depth, site quality and missingness again as described above. Since there were still some specimens with quite huge missingness, again some individuals were removed (`--remove <filename>`), this time by using a cut-off for F_miss of 0.7. Eight specimens had this criterion fulfilled and were thereby sorted out (D-ST-MB-19-154, I-FMM-FM-15-169, D-BY-FK-13-069, CH-FR-MA-14-219, D-BY-SB-18-110, D-BY-NKI-12-021, D-BW-WAR-11-188, D-BW-WAR-15-025). In this subset 170 specimen (170Indv) were left, from which two were internal controls.

Both subsets 170Indv-f8 and 178Indv-f8 were additionally filtered according to the site missingness (`--max-missing <float>`), where sites were excluded based on the proportion of missing data. Four different float values were used 0.5, 0.75, 0.9 and 0.99. Here 0.5 means that 50% missing data (m50) on SNPs is tolerated, so the genotype call rate would be 50% (g50), 0.75 means 25% missing data (m25) is tolerated but genotype call rate would be 75% (g75). A float value of 0.9 and 0.99 have genotype call rate of 90% (g90) and 99% (g99), as well as tolerate 10% (m10) and 1% (m1) missing data on SNPs.

Additionally, to test some less rare SNPs, a third subset was filtered (filter 9, f9) using the same filtering conditions as f7, except for the minor allele frequency (`--maf 0.1`). The position filtering to remove SNP clusters was also performed on this subset (filter 10, f10) and again the same eight specimens were removed according to the F_miss value, remaining with 170 individuals. For the subset with higher minor allele frequencies of 0.1 just one site missingness filter was set (`--max-missing 0.5`), resulting in the dataset 170Indvf10g50m50.

Since there were three different GBS runs conducted, it was mandatory to test for potential batch effects. These are effects on the data caused by non-biological factors due to the experimental design or from technical circumstances like different libraries for sequencing. Therefore, the raw dataset with 185 specimens, as well as the filtered VCF-files (f8g50m50, f8g75m25, f8g90m10) with the 178 specimens and 170 specimens datasets were tested with several output statistics and plotted emphasizing location and library with different colours in R. This was done by using an adjusted guideline from José Cerca (Instructor Physalia courses). The outputs generated with VCFtools v. 0.1.15 for this purpose were mean depth per individual (`--depth`) and proportion of missing data per individual (`--missing-indv`).

As second step, a csv/tab separated file was needed listing the individuals assigned to their corresponding location and library (GBS runs 1-3). Next step was a principal component analysis (`--pca 22`) in PLINK v. 1.9 with the extra parameters `--allow-extra-chr --vcf-half-call m --double-id --set-missing-var-ids @:# --allow-no-sex`. This was done to also plot the principal components (PC) 1 and 2, to check for batch effects on the population clusters. Next, the measures of depth, missingness and PCA were plotted in R and coloured according to the library (GBS-run) and/or the location.

4.5.3 Population genomic and evolutionary analyses

4.5.3.1Sanger-sequenced single PCR products

Subsequently to the alignments of all Sanger-sequenced PCR products (4.5.2.1), they were analysed via DnaSP v. 6 (Rozas et al. 2017), Arlequin v. 3.5 (Excoffier and Lischer 2010), PopART v. 1.7 (Leigh and Bryant 2015), MEGA X (Kumar et al. 2018) and Splitstree4 (Huson and Bryant 2006) for the population genetic and evolutionary inferences. All haplotype sequences of protein-coding sequences were translated to the corresponding amino acid sequences. The assignment of the specimens to haplotype sequences of the respective PCR products was performed in DnaSP v. 6 as well as the transformation into Arlequin project- and haplotype-file formats. All individuals from populations with less than 5 specimens (CH-AG-BK, CH-ZH-WT, CN-HA, CN-JI, CN-TO, D-BW-WAR, I-ANM-O, I-CNP-CN, I-FMM-FM, I-MIL-VI, I-TOP-V, mtALB-COI-Reference) were assigned to a population named "out". Afterwards the haplotype sequences were used for the analysis in PopART and the Arlequin-files for the calculation of the F_{ST}-values and conducting the Analysis of molecular variance (AMOVA).

PopART v. 1.7 was used to create haplotype networks using the TCS method according to the manual by creating trait matrices and combining them with Phylip-format alignments of the haplotype sequences (Geneious prime v2021.0.1) on 658 bp-COI-5P, 360 bp-COI-5P on a global perspective, Anoplophora-COI-5P, ALBsc158_Numt1 (f, int, b), ALBsc1709_Numt2 and ALBsc1709_Numt3 (f, int, b)**.** Haplotype projections on the global map were also conducted in PopART v. 1.7 by expanding the previous in PopART created Nexus-files with the geographical coordinates of the collection sites according to the manual's instructions for the sequences of 658 bp-COI-5P, 360 bp-COI-5P, ALBsc1709_Numt2 and ALBsc1709_Numt3.

With Arlequin v. 3.5 software, the genetic structure of the datasets of mtCOI with the 151 sequences of European 658 bp-COI-5P and 959 sequences of global 360 bp-COI-5P was tested. For the AMOVA, the standard setting for haplotypic formats was used with 1000 permutations. For population comparisons the pairwise F_{ST} was calculated with 100 permutations and a significance level of 0.05. Non-hierarchical AMOVA was used for testing a panmictic scenario for both datasets. For the 658 bp-COI-5P dataset the hierarchical AMOVA

was grouped by country, whereas for the 360 bp-COI-5P dataset the hierarchical AMOVA was just grouped by continent.

Evolutionary analyses were conducted in MEGA X. Previous model testing resulted in Tamura 3-parameter (T92) model (Tamura 1992) for the mtCOI DNA-markers 658 bp-COI-5P and 360 bp-COI-5P, ALBsc158_Numt1, ALBsc1709_Numt2, ALBsc1709_Numt3(f) and ALBsc1709_Numt3 (int) sequences, γ-distributed T92 model (T92+G) for ALB-COI-II, γ-distributed General time reversible model (Nei and Kumar 2000) with invariant sites (GTR+G+I) for the Anoplophora-COI-5P approach and Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al. 1985) for the ALBsc1709_Numt3(b) sequences. The evolutionary history was inferred by using the Maximum Likelihood method and the appropriate substitution model for each DNA-segment. Bootstrap method was used to test the inferred tree with 500 iterations. Estimates of evolutionary divergence and diversity between sequences were inferred by using Tamura 3-parameter model for all DNA-sequence alignments and validated with Bootstrap method (500 iterations).

For a better understanding of the evolutionary relationships and cross-links, also phylogenetic networks generated with Splitstree4 (Huson and Bryant 2006) were used on the alignments in the Nexus format from the DNA-segments mtCOI and ALB-COI-II. The distances in Splitstree4 were estimated with the models with the lowest BIC which were previously tested in MEGA X, NeighborNet (computes Neighbor-Net splits to produce a network) and EqualAngle (computes a planar split network). Since T92 and GTR models were not available in Splitstree4, for 658 bp-COI-5P the model HKY was used and for 360 bp-COI-5P HKY+G+I (α parameter for γ distribution 0.37; proportion of invariable sites 0.76). For the Anoplophora-COI alignment, as well as the ALB-COI-II alignment, the HKY+G model (α parameter for γ distribution 0.23 Anoplophora-COI and 0.70 ALB-COI-II) was applied.

4.5.3.2Genomic SNP-markers from GBS approach

For analyses such as Principal component analysis (PCA) and Admixture, the results are much more representative for the genetic structure when the used SNPs are independent from each other, or in other words when they are in linkage equilibrium (LE). Thus, it is important to measure the linkage disequilibrium (LD) decay with physical distance between SNPs using the r^2 value as a measure, whereas r^2 of 1 stands for full linkage disequilibrium (associated SNP pairs) and 0 are in full linkage equilibrium (fully independent SNP pairs). When LD decay is plotted, a value for r^2 can be determined for pruning the dataset for all highly correlated SNPs and use the remaining independent and non-correlated markers as most representative markers of population stratification.

All steps from r^2-value exploring, plotting in R and the actual pruning steps were adjusted for the dataset of this study using the tutorial of José Cerca (2021a) as guideline (https://github.com/jcerca/population_genomics/tree/master/LD_prunning_HowTo).
To explore the measures of linkage, r^2 was calculated (`--r2`) in PLINK v. 1.9 (Chang et al. 2015; Purcell and Chang 2021) using a sliding window approach (`--ld-window 100 --ld-window-kb 100 --ld-window-r2 0`) with the following extra parameters `--allow-extra-chr --const-fid 0 --allow-no-sex --vcf-half-call m.` According to the results of the LD decay plot (**Figure 68**), the threshold for LD with a r^2-value of 0.15 was used, assuming that all SNPs with an r^2-value below 0.15 are in LE and independent. To prune the SNPs exceeding the set r^2-threshold in PLINK v. 1.9 was used with a sliding window approach (`--indep-pairwise 25 5 0.15`) and the extra settings `--double-id --allow-extra-chr --set-missing-var-ids @:# --allow-no-sex --vcf-half-call m`. The `--indep-pairwise` command produces beside the pruned subset of markers in LE a *.prune.in file which contains the IDs. To extract the actual data `--extract <*.prune.in file> --make-bed --recode vcf` and the extra settings `--double-id --allow-extra-chr --set-missing-var-ids @:# --vcf-half-call m` were used in PLINK v. 1.9. As an estimate for the nucleotide diversity (π) on the filtered 170Indvf8g50m50 dataset, a sliding window approach (`--window-pi 10000 --window-pi-step 10000`) in VCFtools v. 0.1.15 was used on the variants. The π-bins were plotted in R on the 8 biggest scaffolds (scaffold size >3 mb). The R script was adjusted from a script of José Cerca (Instructor Physalia courses). As measures of heterozygosity on a per individual basis (`--het`), the inbreeding coefficient F was estimated using a method of moments, as well as the expected and observed homozygous sites with VCFtools v. 0.1.15 on 170Indvf8g50m50 dataset. The dataset was previouly filtered again on depth (`--min-meanDP 9 --minDP 9`), since the precise estimation of the homozygous alleles is not possible on lower DP values. Plotting of the Heterozygosity was done in R as described before (4.5.2.2) and in JMP Pro v. 15.

To measure the genetic structure of the filtered datasets 170Indvf8g50m50 and 178Indvf8g50m50, the fixation indices F_{ST} and the molecular variance were calculated with Arlequin v. 3.5. First for the input files a list of individuals and populations was created and merged into a population map by first using e.g. `bcftools query -l ./Input-files/170Indv_f8_g50_m50.vcf > 170ind_f8_g50_m50.list.` Afterwards a Pop.list file was created with the same number of rows. Both lists were then merged with `paste` command into a tab separated tsv-file and a header "Individual<tab>Population" was added to the population map (popmap). All individuals from populations with less than 5 specimens were assigned to the population named "out". The translation from vcf-file to

Arlequin format was done with PGDspider v. 2.1.1.5 (Lischer and Excoffier 2012). The popmap files were used in the spid-file to define the populations. In Arlequin, the project files created with PGDspider were selected. For the AMOVA, the standard setting for locus-by-locus AMOVA was used with 1000 permutations. For population comparisons, the pairwise F_{ST} was calculated with 100 permutations and a significance level of 0.05. Non-hierarchical AMOVA was used for testing a panmictic scenario and the hierarchical AMOVA was grouped by countries, excluding China. Additional to the pairwise F_{ST} estimates between the populations, the pairwise F_{ST} estimates on a per site basis (e.g. `--weir-fst-pop Pop1 --weir-fst-pop Pop2`) was also calculated in VCFtools v. 0.1.15 with a sliding window approach (`--fst-window-size 50000 --fst-window-step 10000`) by using an automated Python-script for all populations with defined individuals in one folder. The pairwise F_{ST} was calculated between Italy, Germany and Switzerland as well as between the German states Baden-Wuerttemberg, Bavaria, and Saxony-Anhalt. The F_{ST}-bins were plotted in R on the eight biggest scaffolds as described for the π-bins with adjustments of the R script of José Cerca (Instructor Physalia courses). As another control measure of population stratification, a principal component analysis (`--pca 22`) was performed in PLINK v. 1.9 on the LD-pruned VCF-files as described before with an automated bash-script with the extra parameters `--allow-extra-chr --double-id --set-missing-var-ids @:# --allow-no-sex`. The Eigenvalues were plotted in JMP Pro v. 15 and the PCA plots were made in R by using the same script as described before (4.5.2.2).

As a method of inferring the individual genetic ancestry to deviate the kinship of the analysed specimens to each other, an admixture analysis using ADMIXTURE v. 1.3.0 (Alexander et al. 2009) was conducted with the b-files created by PLINK v. 1.9 on the LD-pruned VCF-file. Prior to the analysis, the copied bim-files needed to be adjusted, since ADMIXTURE v. 1.3.0 just allows integers as "chromosome" information. Thus, `sed 's/contig_//g' $k | sed 's/scaffold_//g'` was run on all bim-files from the LD-pruned datasets 170f8g50m50.LDpruned0.15 and 178f8g50m50.LDpruned0.15. Sequential admixture analysis was performed for K2-K22 (`for K in `seq 2 22``) with additional cross validation (`--cv`) according to the manual. The CV-errors were extracted from the log files using `grep -h '^CV error' log*` and subsequently plotted in JMP Pro v. 15. The Admixture Plots were created with R using two scripts from Lewis Spurgin (Instructor Physalia courses).

For the phylogenetic networks done with Splitstree4 (Huson and Bryant 2006) first the VCF-files needed to be converted into a nexus file, by using a Python script from Ortiz et al. (2021), https://github.com/edgardomortiz/vcf2phylip/blob/master/vcf2phylip.py. The Nexus files of 170Indvf8g50m50 and 178Indvf8g50m50 datasets were loaded into Splitstree4 and calculations were run using NeighborNet and EqualAngle setting. The distance method was

chosen accordingly to the results obtained from model testing in MEGA X software. For both, 170Indvf8g50m50 and 178Indvf8g50m50 datasets, GTR+G+I model resulted in the lowest BIC. But, since GTR model is not available in Splitstree4, γ-distributed Kimura-2-parameter with invariant sites (K2+G+I) was used (α parameter for γ distribution 3.73 and 3.70; proportion of invariable sites 0.02).

5 Results

5.1 Asian long-horned beetle sampling and DNA sequencing

Depending on the quality and integrity of the tissue, contaminations with 230 nm absorbing substances in some samples were higher than optimal with 260/230 ratios below 2.0. But, for most samples 280 nm contaminants were completely removed with the CTAB DNA extraction method (4.4.1), measured by 260/280 between 1.8 and 2.0. In some samples there was a higher proportion of low molecular, degraded DNA beside the high molecular DNA, which was visually checked after gel electrophoresis of gDNA. Particularly some dried samples seemed to have a strongly reduced DNA quality with higher low molecular weight proportion. Nevertheless, these samples were all used for sequencing to reflect the reality of samples occurring in plant health controls also under not optimal conditions.

For the mtALB-COI DNA-marker and subsequent sequencing, mostly dried specimen (D-BW-WAR-15-023, D-BY-MI-19-193, D-BY-SB-18-110) or larva needed to be cloned due to bad trace data quality in direct PCR product sequencing. The samples for which the gDNA was detected to be partly or completely low molecular weight (D-BW-WAR-11-188, D-BW-WAR-11-190, D-BW-WAR-11-191, D-BW-WAR-11-192, D-BY-SB-110, I-MIL-VI-14-183, I-MIL-VI-14-184, I-MIL-VI-14-185) were sequenced with the GBS approach anyway (4.4.12), but most of them where filtered out from the final SNP dataset. The 22 specimens from nine German and seven Italian populations used for the PCR amplification of the NUMT DNA-sections were chosen by the best quality GBS-results (lowest missingness and highest SNP call).

Most of the total 199 individuals were sequenced for mtCOI and with the GBS approach, while just a few individuals of each population were used for the NUMT PCR product sequencing. However, not all samples could be successfully sequenced for the three different approaches as pictured in **Figure 12** and marked in **Table 21**, Appendix, I Asian long-horned beetle sample, PCR product and DNA-marker information. Most of the samples from Europe and the three Chinese samples were collected in the years 2014, 2016 and 2019, as shown in **Figure 13**. The samples from the populations D-BW-WAR (2011), CH-ZH-WT (2012) and D-BY-NKI (2012) were the oldest ones. Some of them had very bad storing conditions and high levels of degraded DNA (**Table 20**, Appendix, I Asian long-horned beetle sample, PCR product and DNA-marker information).

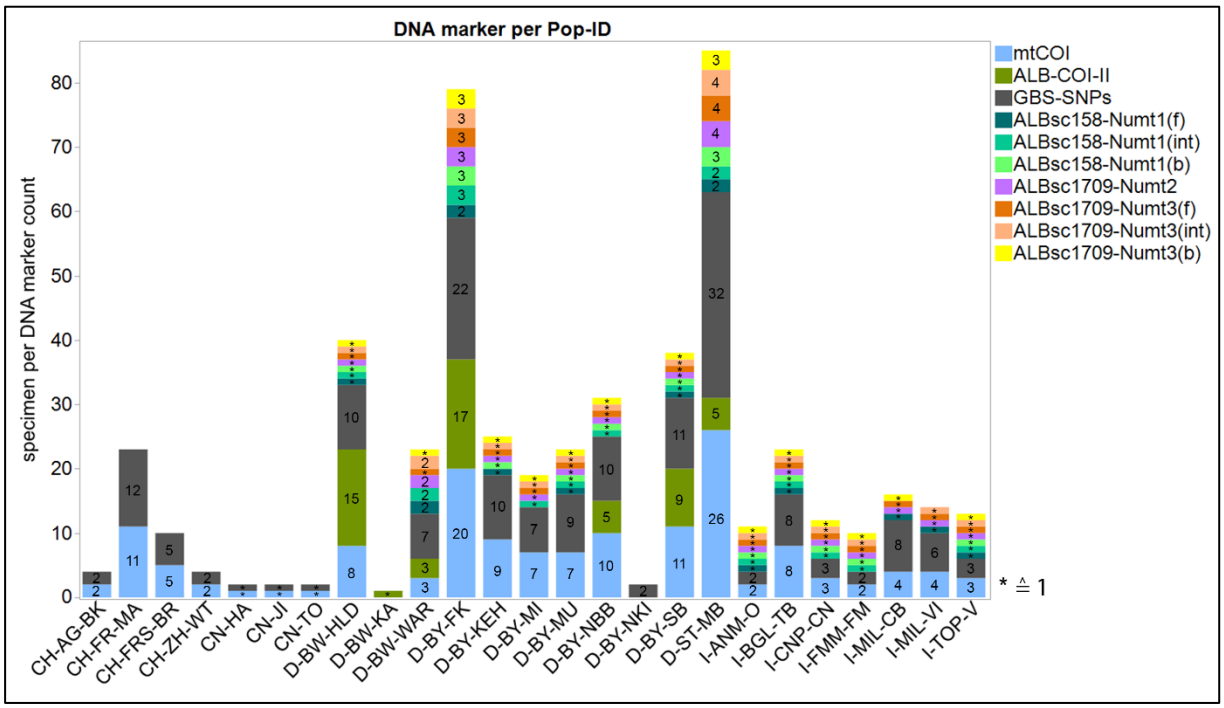


Figure 12. DNA marker and PCR products per Pop-ID.
The numbers in the bar sections represent the exact specimen count for each Pop-ID and the corresponding sequenced PCR products and DNA-markers.

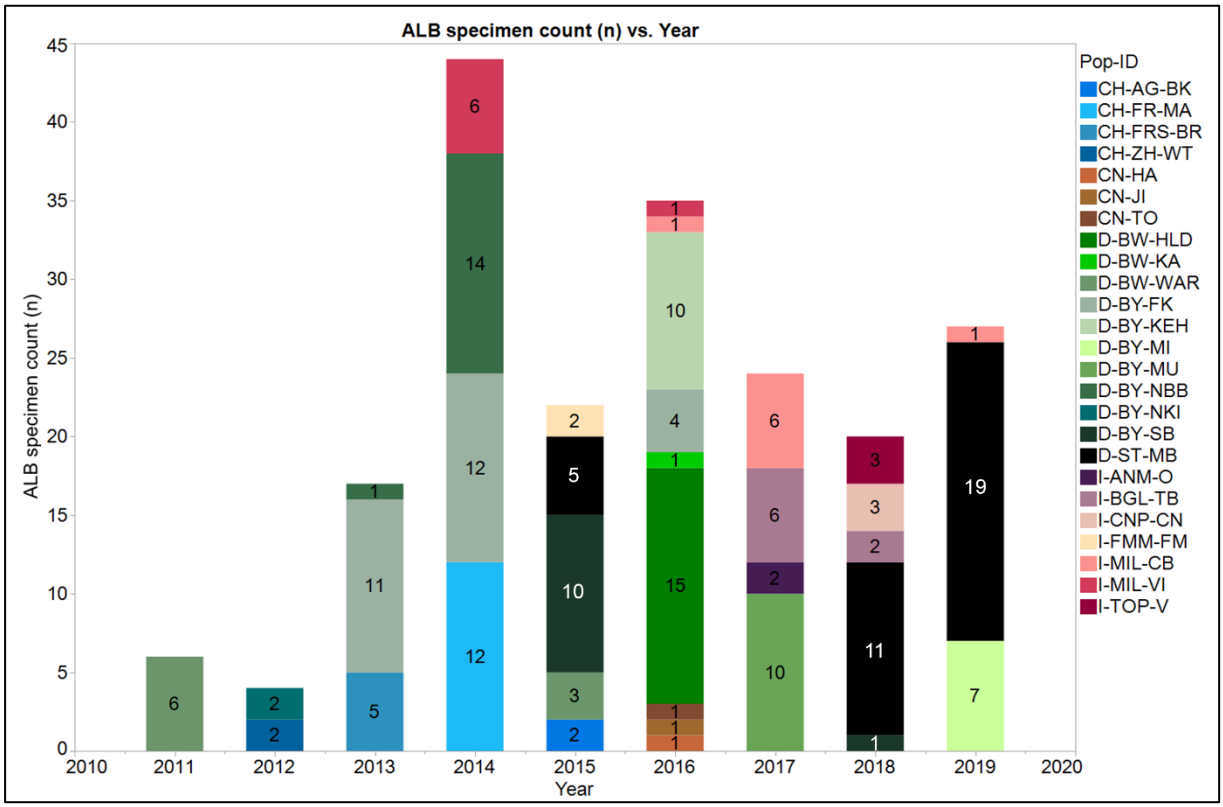


Figure 13. Pop-IDs represented per collection year.
The numbers in the bar sections represent the exact specimen count for each Pop-ID and the corresponding collection year.

For the NUMT PCR-products, except the internal ones, Primer-BLAST (2021) was conducted before using the primers in the PCR amplification. Some of the primers for the NUMT amplifications showed conflicts with other regions of the ALB genome as visible in **Table 10** with red marks. The conflict for the DNA-segment ALBsc158_Numt1(int) was unintentionally, but the conflict shown for ALBsc1709_Numt2 DNA-segment was the reason why all PCR products were sequenced directly from the PCR-products and from 5 clones each, to obtain additional information about the two different scaffold-regions. The multiple banding during the PCR-amplification of ALBsc1709_Numt3(f) and the necessity of the gel extraction of the PCR product with the right length can be explained with the results of the Primer-BLAST. The primers for mtCOI barcoding and ALB-COI-II PCR products were tested with Primer-BLAST retrospectively after the results depicted in "Chapter B: results". The universal primers from Folmer et al. (1994) showed a conflict with scaffold1356 (NW_019417445.1) of ALB's reference genome with a similar length than the desired PCR product (**Table 10**).

Table 10. Primer-BLAST of all primers used for PCR amplification – (1/2).

Primerpair	Target template	product length	mismatches
mtCOI	scaffold1356 (NW_019417445.1)	702 bp	5 2
	mitochondrion, complete genome (NC_008221.1)	709 bp	4 3
ALB-COI-II	mitochondrion, complete genome (NC_008221.1)	1271 bp	0 0
	scaffold109 (NW_019416246.1)	72 bp	4 4
Numt1(f)	scaffold158 (NW_019416265.1)	844 bp	0 0
	scaffold40 (NW_019416268.1)	100 bp	2 3
	scaffold107 (NW_019416230.1)	2002 bp	5 1
Numt1(int)	scaffold158 (NW_019416265.1)	747 bp	0 0
	scaffold1709 (NW_019417922.1)	783 bp	1 0
	scaffold1709 (NW_019417922.1)	788 bp	2 0
Numt1(b)	scaffold158 (NW_019416265.1)	1394 bp	0 0
	scaffold275 (NW_019416374.1)	1760 bp	5 3
	scaffold584 (NW_019416664.1)	1553 bp	4 5

Table 10. Primer-BLAST of all primers used for PCR amplification – (2/2).

Primerpair	Target template	product length	mismatches
Numt2	scaffold164 (NW_019416480.1)	469 bp	0 0
	scaffold1709 (NW_019417922.1)	469 bp	0 0
	scaffold8 (NW_019416350.1)	3137 bp	5 1
Numt3(f)	scaffold1709 (NW_019417922.1)	953 bp	0 0
	scaffold1709 (NW_019417922.1)	1353 bp	1 0
	scaffold1709 (NW_019417922.1)	1135 bp	0 4
	scaffold1709 (NW_019417922.1)	720 bp	1 4
	scaffold1709 (NW_019417922.1)	275 bp	1 5
	scaffold1709 (NW_019417922.1)	301 bp	4 4
Numt3(int)	scaffold1709 (NW_019417922.1)	505 bp	0 0
	scaffold289 (NW_019416639.1)	2613 bp	3 2
	scaffold180 (NW_019416335.1)	3533 bp	2 3
Numt3(b)	scaffold1709 (NW_019417922.1)	1206 bp	0 0
	scaffold159 (NW_019416271.1)	447 bp	4 5
	scaffold314 (NW_019416415.1)	1004 bp	5 4

green marked boxes are the desired PCR-products, red marked are conflicts with other regions in ALB's genome.

5.2 Chapter A: results

Chapter A

"Prospects and limits of COI-barcoding practice to characterize the biological invasions of Asian long-horned beetle, *Anoplophora glabripennis*"

5.2.1 Genetic diversity of mtCOI in European dataset of ALB

The conserved gene for Cytochrome oxidase 1 (COI) is a well-established marker for species identification in arthropods. The data availability is very good, since it was used in several ALB studies already, also in an intraspecific manner (Carter et al. 2009c; Carter et al. 2010; Javal et al. 2019a; Kim et al. 2019; Lee et al. 2020; Tsykun et al. 2019). Therefore, first the 658 bp COI barcoding region (excluding primers) of 150 individuals from 20 sites in Germany, Switzerland and Italy were Sanger-sequenced for a better insight in the European genetic diversity and three Chinese samples were included too. The proportion of the specimens per population sequenced for this purpose is pictured in **Figure 14**. The populations D-BW-KA and D-BY-NKI were not represented by the mtCOI marker. Most specimens sequenced for this marker were from D-ST-MB and D-BY-FK, while the Italian, Chinese and CH-AG-BK and CH-ZH-WT samples were just represented by a low sample number.

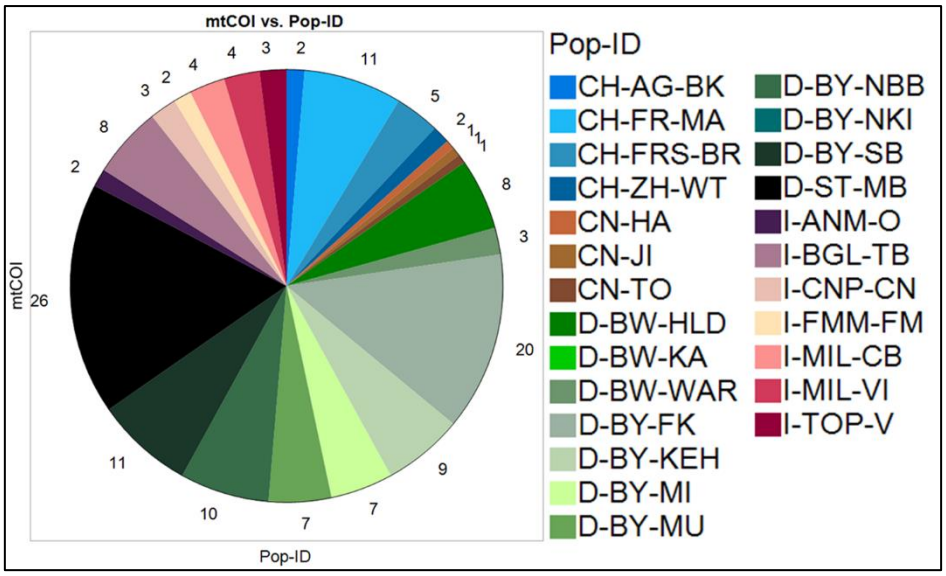


Figure 14. Represented ALB-populations (Pop-ID) from barcoding mtCOI DNA-marker.

The dataset resulted in 168 sequences for the 153 specimens, that could be assigned to eight different haplotypes which were different from the reference mitochondrial genome available on NCBI (NC_008221). Most specimens of D-BY-FK (18 out of 20) showed one ambiguous nucleotide at the same position with clear double peaks in the trace date, which was visually checked. The haplotypes were named from COI_Type1-8 according to their abundance in the

dataset. All haplotypes were blasted with megaBLAST against the nucleotide collection. The first three hits of each of the haplotypes are represented in **Table 22** in the Appendix, II Chapter A. All eight haplotypes had >99% similarities with other *A. glabripennis* sequences in the NCBI nucleotide collection database in the first three hits.

The haplotypes (including the reference) showed 17 polymorphic sites on the 658 bp gene section of COI, hence there were 641 monomorphic sites. Of the total 17 mutations, four were singleton variable sites and 13 parsimony informative. None of the 658 sites was missing at any of the sequences. Some of the haplotypes shared mutations with other haplotypes. As shown in the alignment of all haplotypes (**Figure 15**), there are some unique polymorphic sites (4), like among others e.g. position 23 with an A instead of an G in COI_Type4. Additionally, there are also some polymorphic sites shared between a minority of haplotypes. The COI_Types 3, 7 and 8 shared a T at position 284, as well as a C on position 310. COI_Type 6 and 8 shared an A at position 473. Half of the haplotypes (COI_Type3, 5, 7, 8) shared a G on position 550, whereas the other four haplotypes and the reference have an A on that position.

To check for synonymous or non-synonymous substitutions due to the 17 polymorphic sites, the sequences were translated to amino acid sequences on the alignment (frame +2) reflecting the open reading frame of the whole cytochrome oxidase I (COI) coding sequence (CDS). The alignment of the translated amino acid sequences had a total length of 219 amino acids which is pictured in **Figure 15**. Of the 17 polymorphic sites, there were eight leading to non-synonymous substitutions in comparison to the consensus of all nine haplotypes. Two different mutations in COI_Type4 and COI_Type6 both lead to a substitution in the same amino acid position. The mutation in COI_Type4 at position 23 in the DNA sequence in comparison to COI_Type1 led to an amino acid change from Alanin (Ala, A) to Threonine (Thr, T) at position 8 in the amino acid sequence. The ambiguous site in the D-BY-FK population was counted as a real heterozygous locus since the translation to amino acid sequences resulted in the correct open reading frame (ORF) for both haplotypes. On the same position 23, the translation of COI-Type6 showed a non-synonymous substitution resulting in Valine (Val, V) instead of Alanin. Similar, two different polymorphisms in COI_Type6, COI_Type 7 and COI_Type8 led to two different amino acid changes on the position 158 on the amino acid sequence. The other three non-synonymous substitutions just occurred in one haplotype each. Of the 17 polymorphic sites, nine mutations were synonymous in comparison to the consensus.

Figure 15. Alignment of the eight haplotypes of mtCOI 658 bp DNA-marker from 153 ALB specimens and the haplotype of the reference NC_008221 in Geneious Prime® v. 2021.0.1 with amino acid translation (frame +2).
Disagreements to the DNA consensus sequence were marked here coloured according to their base (G = yellow, C = blue, T = green, A = red). Genetic code for amino acid translation: invertebrate mitochondrial (translation table 5). Disagreements to the amino acid consensus sequence were marked here colored according to their amino acid.

The abundance and connections of the described haplotypes are shown in more detail in a TCS network (**Figure 16**), geographic haplotype projections (**Figure 17**) and in a Venn diagram (**Figure 18**).

The most abundant haplotype COI_Type1 was represented by 78 specimens in the dataset from all four sampled countries, separated by only one mutation from the reference haplotype COI_Ref. This haplotype COI_Type1 was also determined in Harbin (CN-HA). Some scattered haplotypes were clearly distinguished from the others e.g. in Magdeburg, Feldkirchen, Marly and Bruensried, as well as in Corbetta and Vittuone. The haplotype COI_Type2 is unique to the 26 specimens from the German location Magdeburg (D-ST-MB) and is also one mutation apart from COI_Type1. Some collection sites inhabited two different haplotypes like Feldkirchen (COI_Type1 and COI_Type4), Kehlheim and Fermo (COI_Type1 and 3), as well as Cuneo (COI_Type1 and 8). COI_Type3 was observed in Germany (Kelheim, Schönebach) and Italy (Ostra, Trescore-Balneario, Fermo) with a total of three mutations from the main haplotype COI_Type1. The two haplotypes COI_Type1 and COI_Type4 observable in 18 out of 20 of the sequenced specimens in D-BY-FK population of all three collection years were separated by one mutation and COI_Type4 was only abundant in Feldkirchen. Another exclusive haplotype COI_Type5 was observed in the two Swiss neighbouring locations Marly and Bruensried, which were already described in Tsykun et al. (2019) as COXC. The most distant haplotypes COI_Type6 (eight mutations), COI_Type7 (four mutations) and COI_Type8 (six mutations) in comparison to the main haplotype COI_Type1 were all observed in Italy and China, whereas COI_Type7 (Jinan, China) and COI_Type 8 (Cuneo, Italy) were unique in the whole dataset. COI-Type7 from Jinan is differentiated by only one mutation from some samples of Bavaria and Italy. The most diverged haplotype COI_Type6 occurred in the two neighbouring Italian locations Corbetta and Vittuone, as well as in Tongliao, China. COI-Type6 is most clearly differentiated with eight and more mutations from all the other haplotypes. In **Figure 16** the exact number of mutations described above setting the different haplotypes apart from each other can be seen as hatch marks. Most strikingly was, that most haplotypes were only a few mutations apart from each other. By taking the haplotype projections (**Figure 17**) into account, there was no specific pattern of haplotypes linked to a specific location, except of the distinct ones described above. With a look on the Venn diagram (**Figure 18**), the only haplotype shared between all three European countries Switzerland, Germany and Italy was Type1. Type5 occurred in Switzerland only, Type2 and Type4 in Germany and Type6 and 8 in Italy only. Type 3 was shared between Germany and Italy.

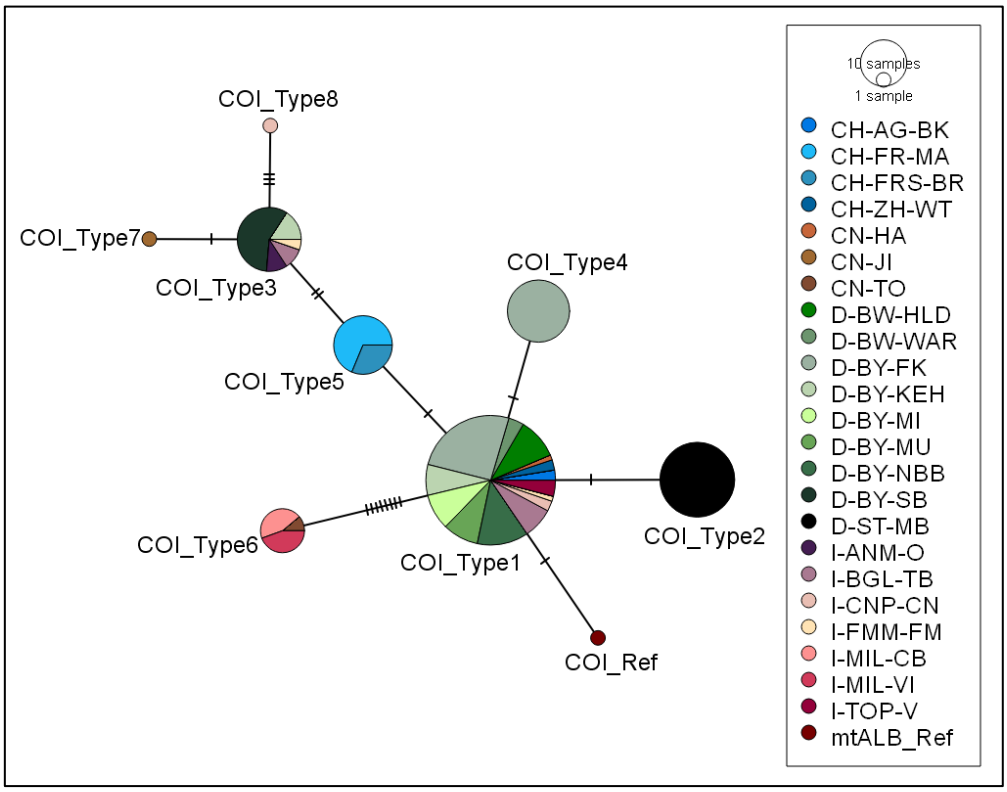


Figure 16. TCS haplotype network of nine haplotypes of mtCOI-658 bp from 150 European ALB specimens from Switzerland, Germany, Italy, three Chinese specimens and the reference sequence (PopART v. 1.7).
Mutations are represented by hatch marks; the size of the circles is proportional to the number of specimens contributing to the haplotype.

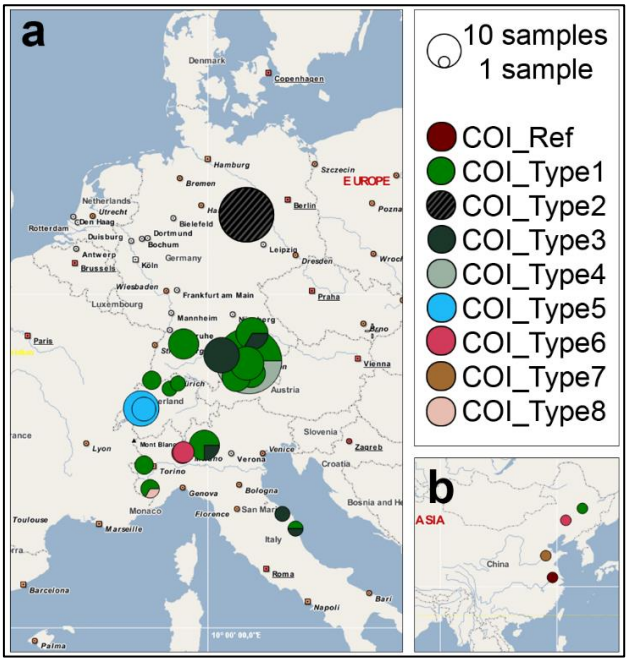


Figure 17. Haplotype projections of nine COI haplotypes in Europe (a) and China (b) are shown (PopART v. 1.7).
For the mitochondrial reference haplotype, the location of the laboratory was used, since there was no further information on NCBI (Laboratory of Plant Quarantine, Jiangsu, China).

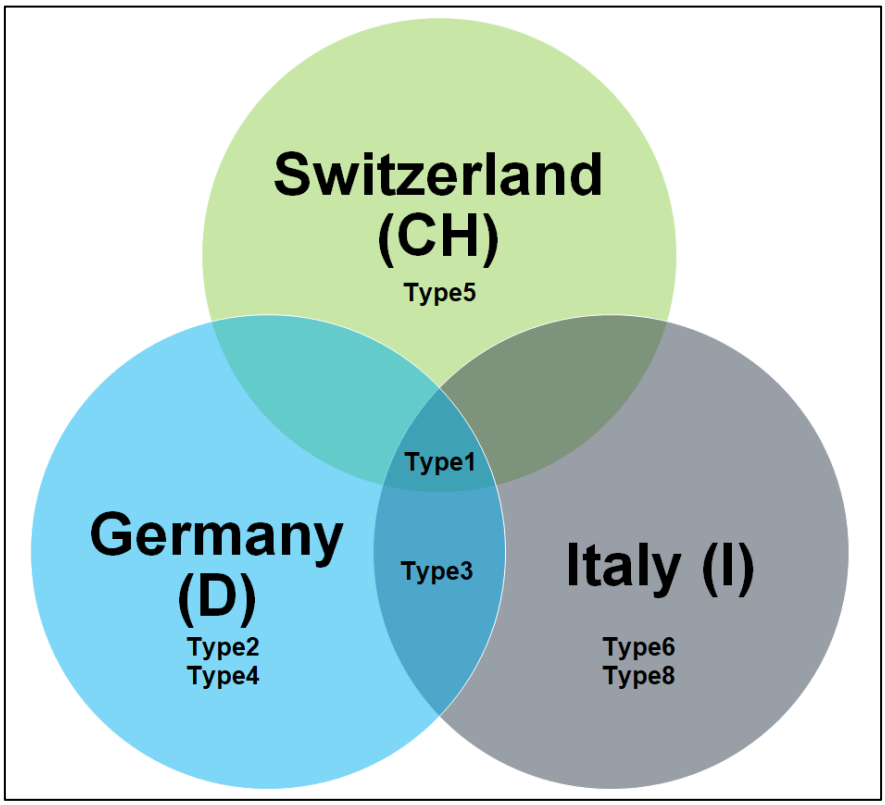


Figure 18. Venn diagram on the distribution of the mtCOI-658 bp DNA-marker haplotypes on the three European countries Switzerland, Germany and Italy.

In the following analyses on genetic diversity, population structure and molecular variance, the dataset was split in two for the two different haplotype sequences to only include one haplotype on those 18 specimens affected in D-BY-FK to avoid mistakes in calculation of the measures. Hence, there was a dataset mtCOI-658 bp-a including all nine haplotypes with 151 sequences, whereas the dataset mtCOI-658 bp-b lacks the haplotype COI_Type4 (polymorphic sites: 16, monomorphic sites: 642, singletons: 4, parsimony informative sites: 12). In both analyses the both populations I-MIL-CB (four specimens) and I-MIL-VI (four specimens) were grouped together as population Milano (I-MIL) to not lose the entire COI_Type6 haplotype due to the set threshold for populations to have at least five specimens. The other populations not fulfilling the minimum of specimens were set as population "out".

First, common parameters on genetic diversity were obtained with DnaSP v. 6 (Rozas et al. 2017) and integrated in **Table 11** for mtCOI-658 bp-a (number of haplotypes h: 9, haplotype diversity, Hd: 0.7726) and mtCOI-658 bp-b (number of haplotypes, h: 8, haplotype diversity, Hd: 0.6773).

Table 11. Summary of common parameters of genetic diversity of mtCOI-658 bp-a and -b.

population \ parameters	n	S	h	Hd	π
CH_FR_MA	11	0	1	0.00000	0.00000
CH_FRS_BR	5	0	1	0.00000	0.00000
D_BW_HLD	8	0	1	0.00000	0.00000
D_BY_FK	20	1 (0)	2 (1)	0.18947 (0.00000)	0.00029 (0.00000)
D_BY_KEH	9	3	2	0.50000	0.00228
D_BY_MI	7	0	1	0.00000	0.00000
D_BY_MU	7	0	1	0.00000	0.00000
D_BY_NBB	10	0	1	0.00000	0.00000
D_BY_SB	11	0	1	0.00000	0.00000
D_ST_MB	26	0	1	0.00000	0.00000
I_BGL_TB	8	3	2	0.42857	0.00195
I_MIL	8	0	1	0.00000	0.00000
out	21	15	6	0.55238	0.00360
Total Data Estimates	151	17 (16)	9 (8)	0.77263 (0.67726)	0.00353 (0.00321)

13 populations were included in this analysis of gene flow and genetic differentiation
n = number of sequences, S = segregating sites, h = number of haplotypes, Hd = haplotype diversity, π = nucleotide diversity, values in brackets belong to mtCOI-658 bp-b.

The genetic structure of populations was investigated by AMOVA, where the null hypothesis assumed no differentiation within and between groups. By defining groups of populations, a particular genetic structure was tested. Here the two scenarios panmixia and a restricted mating between the different countries were simulated. The AMOVA showed very high molecular variance among the populations in the datasets, as described in **Table 12**. This could be observed in both the non-hierarchical (91.78%, 92.17%) and the hierarchical approaches. The approach to test the structure according to the countries depicted the highest molecular variance among the populations within the countries (63.87%, 59.68%). For the dataset without COI_Type4 (mtCOI-658 bp-b) showed less variation in both hierarchical levels. The molecular variance within the populations for mtCOI-658 bp is estimated higher (8.22%, 7.83%) in the non-hierarchical approach, than with a country-wise population hierarchy (6.90%, 6.37%). For the dataset without COI_Type4 the values were similar, but in the non-hierarchical approach the variation within the populations was slightly higher.

The measures to determine the degree of differentiation, the fixation indices F_{ST}, F_{SC} and F_{CT} are depicted in **Table 12**. The highest degree of differentiation was estimated in the hierarchical approach with F_{ST}-values to show the differentiation among the demes with >0.9. The differentiation among the populations within a country (F_{SC}) is very high as well with >0.9. Otherwise, the degree of differentiation among the countries (F_{CT}) was less but still high with values about 0.3.

Table 12. AMOVA design and results on dataset of 153 ALB specimens with the DNA marker mtCOI-658bp conducted in Arlequin v. 3.5.

Source of variation	d. f.	Sum of squares	Variance components	Percentage variation	Fixation indices	p values
(a) AMOVA - No. of groups 1 (mtCOI-658 bp-a)						
Among populations	11	136.254	1.16425 (Va)	91.78***		0.00000
Within populations	118	12.300	0.10424 (Vb)	8.22	[1]0.91783***	0.00000
Total	129	148.554	1.26849			
(a) AMOVA - No. of groups 1 (mtCOI-658 bp-b)						
Among populations	11	122.546	1.04758 (Va)	92.17***		0.00000
Within populations	118	10.500	0.08898 (Vb)	7.83	[1]0.92171***	0.00000
Total	129	133.046	1.13656			
(b) AMOVA - No. of groups 3 CH, D, I (mtCOI-658 bp-a)						
Among countries	2	41.508	0.44200 (Va)	29.24*	[3]0.29238*	0.06158+ -0.00571
Among populations within country	9	94.746	0.96549 (Vb)	63.87***	[2]0.90256***	0.00000
Within populations	118	12.300	0.10424 (Vc)	6.90***	[1]0.93105***	0.00000
Total	129	148.554	1.51173			
(b) AMOVA - No. of groups 3 CH, D, I (mtCOI-658 bp-b)						
Among countries	2	40.694	0.47451 (Va)	33.95***	[3]0.33949*	0.04594+ -0.00687
Among populations within country	9	81.852	0.83420 (Vb)	59.68***	[2]0.90361***	0.00000
Within populations	118	10.500	0.08898 (Vc)	6.37***	[1]0.93634***	0.00000
Total	129	133.046	1.39769			

Significance tests (1023 permutations): *** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.1$; [1]F_{ST}, [2]F_{SC}, [3]F_{CT} a) no hierarchical structure, "out" population excluded b) structure to test: populations of Switzerland, Germany and Italy grouped to the respective country, "out" population excluded.

The **Figure 19** represents the pairwise differentiation between the subpopulations (F_{ST}) on nine haplotypes (mtCOI-658 bp-a) using pairwise difference as distance method. The values for the degree of differentiation are pointed out with grades of coloring from low to high. The pairwise differences on eight haplotypes (mtCOI-658 bp-b) (**Figure 91**), the exact F_{ST}-values (**Table 23**, **Table 25**) and p-values (**Table 24**, **Table 26**) are given in Appendix II, Chapter A. Most of the significant pairwise F_{ST}-values show very high differentiation with values >0.8-1. The population Milano (I-MIL) is without exception extremely differentiated from all other populations. The same can be said about Marly (CH-FR-MA), Bruensried (CH-FRS-BR), Schoenebach (D-BY-SB) and Magdeburg (D-ST-MB). Marly and Bruensried showed a bit less differentiation to Kelheim (D-BY-KEH) and Trescore-Balneario (I-BGL-TB), but still very strong differentiation (~0.3-0.4). Feldkirchen (D-BY-FK), Schoenebach and Magdeburg showed lower F_{ST}-values in comparison to Kelheim and Trescore-Balneario but are also still highly differentiated (>>0.5). The lowest distance was detected between the populations of Miesbach, Murnau, Neubiberg and Hildrizhausen, as well as between Kelheim and Trescore-Balneario, but all this lower F_{ST}-values (except the comparison to themselves) were not significant. The pairwise differences of the mtCOI-658 bp-b data (**Figure 91**) showed the same results.

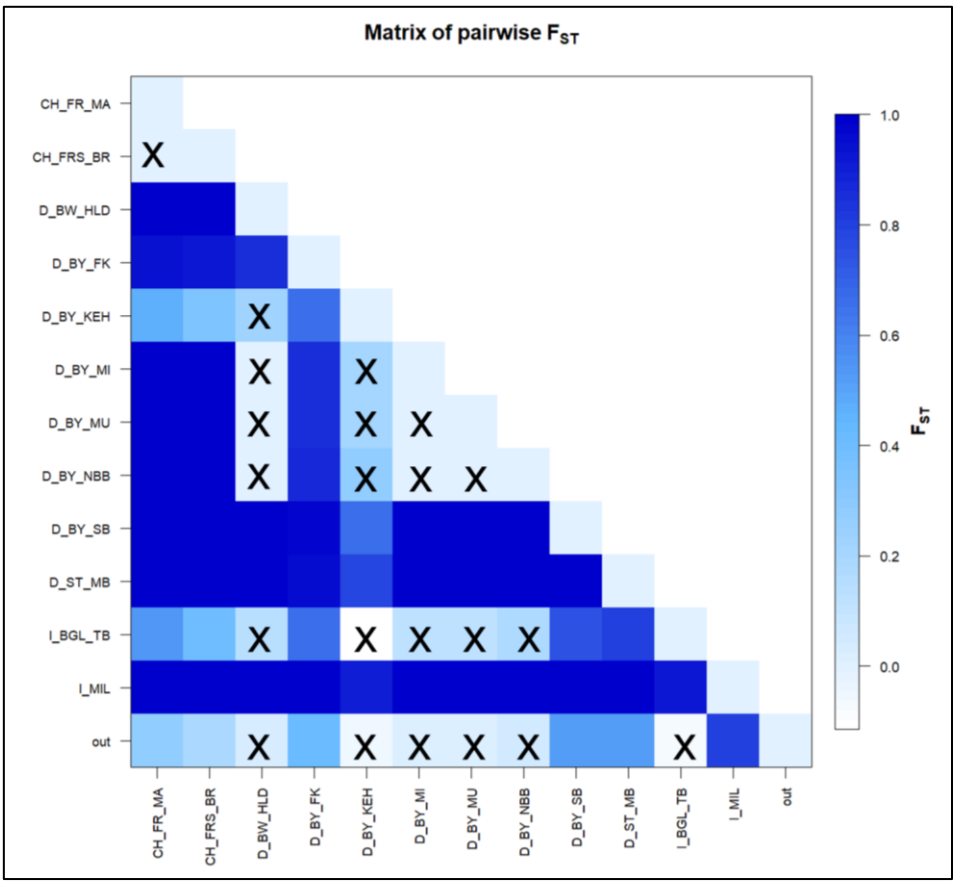


Figure 19. Population differentiation on nine haplotypes (mtCOI-658 bp-a).
Pairwise F_{ST}-values among populations mtCOI-658 bp-a.; x = not significant (significance level=0.05). Population labels were used as set in **Table 4**.

Very similar results like in the TCS network (**Figure 16**) and the pairwise distances (**Figure 19**) were obtained by the estimates of evolutionary divergence from Mega X (Kumar et al. 2018) as depicted in **Table 13**. The haplotypes COI_Type6 and COI_Type8 are most distinct from the other haplotypes, with mostly over 1% different nucleotides. With divergences of 1.229-1.699% COI_Type 6 is the most distinct haplotype from all the others. COI_Type8 has most similarity with divergence of 0.458% to COI_Type3 and 0.611% to COI_Type7 but is still very divergent. The least differences were observable in the main cluster of COI-Ref, COI_Type1, 2, 4 and 5. The most abundant haplotype in this dataset, COI_Type1, has lower differences to COI_Ref, COI_Type2, 4 and 5 with divergences of 0.152%, but it shows a higher difference to all the other haplotypes.

Table 13. Estimates of evolutionary divergence between nine haplotype sequences of mtCOI-658 bp conducted in MEGA X converted into percentage.

haplotypes	Ref	1	2	3	4	5	6	7	8
COI_Ref									
COI_Type1	0.152%								
COI_Type2	0.305%	0.152%							
COI_Type3	0.612%	0.458%	0.612%						
COI_Type4	0.305%	0.152%	0.305%	0.612%					
COI_Type5	0.305%	0.152%	0.305%	0.305%	0.305%				
COI_Type6	1.385%	1.229%	1.385%	1.699%	1.385%	1.385%			
COI_Type7	0.765%	0.611%	0.765%	0.152%	0.765%	0.458%	1.854%		
COI_Type8	1.075%	0.920%	1.075%	0.458%	1.075%	0.765%	1.860%	0.611%	

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992). This analysis involved 9 nucleotide sequences with a total of 658 positions in the final dataset.

The evolutionary tree (**Figure 20**) drawn by the maximum likelihood method reflects the observations from TCS network and the different estimates on the differences. Prior model testing of the alignment (169 sequences of mtCOI-658 bp) measured the lowest BIC score (Bayesian information criterion) on the nucleotide substitution model T92 (Tamura-3-parameter), so this model was chosen to describe the substitution pattern best. The five highly similar haplotypes COI_Type1, 2, 4, 5 and COI_Ref group together. The bootstrap values (57-65) to separate these haplotypes measured lower confidence. Another Group is built by haplotypes COI_Type3, 7 and 8 that is supported by a bootstrap value of 73. Differently, COI_Type6 is building an outgroup, that is divergent from the rest and its branch shows the highest bootstrap value of 99.

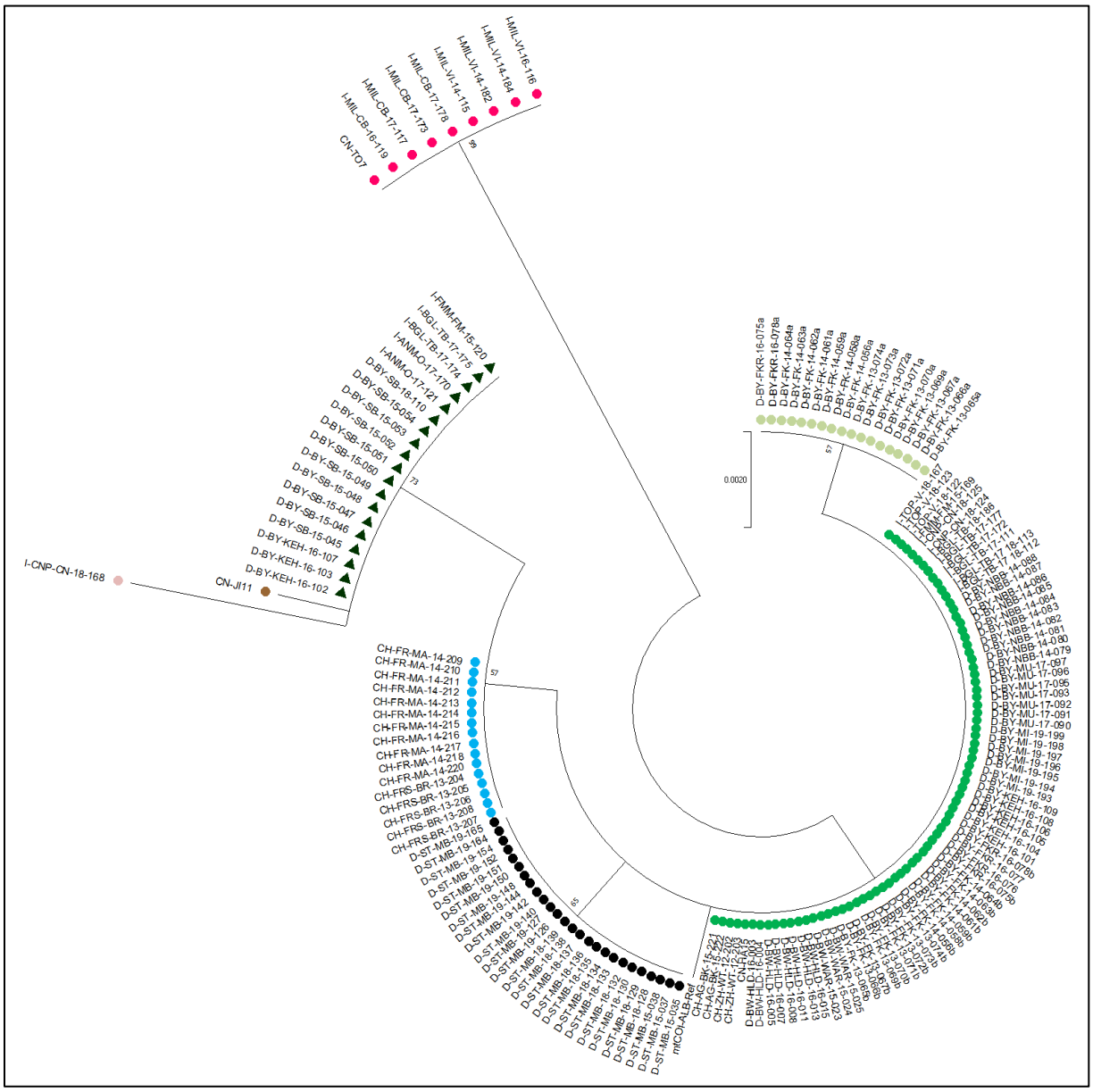


Figure 20. Evolutionary history inferred by using the Maximum Likelihood method on 169 sequences of mtCOI-658 bp conducted in MEGA X (Kumar et al. 2018).
The Tamura 3-parameter model (Tamura 1992) was used for the pairwise distance estimation to draw the tree and bootstrap method to test the results. The branch lengths measure in the number of substitutions per site. This analysis involved 169 nucleotide sequences. There was a total of 658 positions in the final dataset.

The NeighborNet phylogenetic network (**Figure 21**) drawn with Splitstree v. 4.17.1 supports all previous findings and gives more detailed insight in the phylogenetic connections of the haplotypes. When having a look on the crosslinks, it seems like all haplotypes from COI_Type1-5, COI_Type7 and 8 show a parallel branch with connections to each other long time ago (measured in substitutions per site). In more detail, COI_Type2 and COI_Type4 are slightly split from COI_Type1 (~0.001). One branch is including the haplotypes COI_Type1, 5, 3 and a bit more distant COI_Type7, whereas COI_Type8 has its own branch, which is more distant to the others, but shows crosslinks to the haplotypes COI_Type3, 5 and 1 at distances from ~0.003-0.008 substitutions per site. The most strikingly distant haplotype is again COI_Type6, which does not show crosslinks to other haplotypes for a long distance but has the strongest support with a bootstrap-value of 99.9.

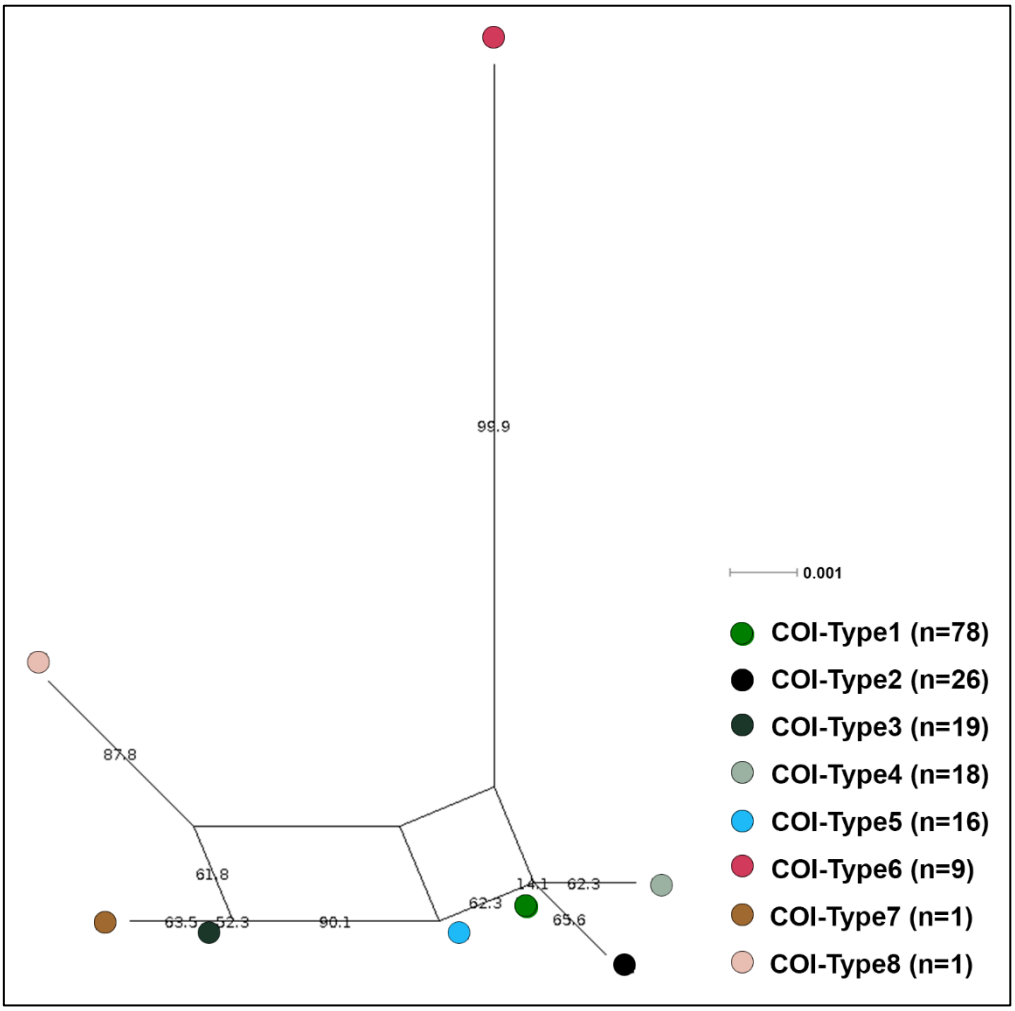


Figure 21. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on 168 sequences representing 8 haplotypes of mtCOI-658 bp.
HKY85 distances and Equal Angle were used to draw the network with 10 splits; Fit = 99.83; 1000 runs of Bootstrap. 8 out of 168 taxa are shown on 658 characters.

5.2.2 Genetic diversity of mtCOI in global dataset of ALB

Since COI was used in several ALB studies before, the data availability was very good (Carter et al. 2009c; Carter et al. 2010; Javal et al. 2019a; Kim et al. 2019; Lee et al. 2020; Tsykun et al. 2019). This was enabling the comparison with public databases (BOLD, NCBI) with 809 global ALB sequences of nine countries from North America (Canada, USA), Asia (China, South-Korea) and Europe (Austria, France, Germany, Italy, Switzerland) to have an insight into the Global genetic diversity within the conserved COI-gene. All these mined sequences were aligned to the 168 sequences of the 150 European specimens from this study. For including all the available sequences, the section on the COI-barcoding region was determined which was represented by all 977 sequences and was used for further analyses. Since the alignment for the two haplotypes in 18 specimens from D-BY-FK showed no sequence differences, just one sequence of each was used in the following. Hence, from a total of 959 sequences from position 1693-2052 (= 360 bp) on the mitochondrial reference genome NC_008221.1, the haplotypes were determined. In total there were 33 haplotypes (COI_h01-h33) observable for the 959 individuals from nine countries. The original haplotypes from the mtCOI-658 bp DNA-marker were incorporated into the haplotypes from the smaller proportion of the COI gene with 360 bp. The most abundant haplotype COI_h01, represented by 359 sequences, included COI_Type3 (19 specimens) and 5 (16 specimens). The second most abundant haplotype COI_h02, represented by 311 sequences, included COI_Type1 (78 specimens), 4 (18 specimens) and COI_Ref. COI_Type6, which was described as the most distant one, was incorporated in COI_h03 and occurred 136 times. The exclusively in Magdeburg (Germany) occurring haplotype from the European set COI_Type2 was represented 36 times in the global dataset as COI_h04. COI_Type8, which was unique in the European dataset, was represented four times in the global dataset with COI_h11, whereas COI_Type7 remained unique as COI_h33. The other haplotypes COI_h05 (30 specimens), h06 (20 specimens), h07 (13 specimens), h08 (11 specimens), h09 (7 specimens), h10 (6 specimens), h12-h15 (2 specimens each) and COI_h16-h32 (1 specimen each) did not represent any of the COI_Types of the mtCOI-658 bp DNA-marker found in the dataset from this study.

The 33 haplotypes (including the reference) showed 22 polymorphic sites within the 360 bp gene fragment of COI. There were 14 sites with missing information, hence there were 324 monomorphic sites. Of the total 22 mutations, eight were singleton variable sites and 14 parsimony informative. As visible in the alignment of all haplotypes with the COI-barcoding reference (**Figure 22**), there were two haplotypes with ambiguous nucleotides, COI_h02 and COI-h10. For the COI_h02 haplotype, there were 11 sequences assigned to this haplotype by DnaSP v. 6 (Rozas et al. 2017) with N at 12 positions in the sequence (positions 14, 30, 43,

69, 176, 215, 227, 240, 247, 285, 318, 333), of which five are on a parsimony informative site (30, 43, 69, 240, 333) and are therefore masked in further analyses. COI_h10 has an ambiguous nucleotide R on pos. 227 because DnaSP v. 6 assigned them to the same haplotype although they are actually two very similar ones. Within COI_h10 ambiguous sites exist on the positions 30, 186 and 227 (encoded as N), of which 30 and 227 are parsimony informative sites. Therefore, the positions 30 and 227 were masked in further analysis as well. Of 14 parsimony informative sites, six were masked in further analyses due to the ambiguous nucleotides in COI_h02 and COI_h10. Also, the singleton variable sites with Ns were masked in the further analyses.

To check for non-synonymous or synonymous substitutions due to the 22 polymorphic sites, the sequences were translated to amino acid sequences (frame +1) reflecting the open reading frame of the whole cytochrome oxidase I (COI) coding sequence (CDS). The alignment of the translated amino acid sequences from the 33 haplotypes (frame +1) in **Figure 23**, show that almost half of the polymorphic sites in the DNA-sequence resulted in a non-synonymous substitution in comparison to the consensus of all haplotypes. From the 120 amino acids in total, there were three singleton substitutions and seven substitution sites that occurred in more haplotype translations. The positions 76, 77 and 80 were polymorphic sites with at least one X (Xaa, any amino acid) due to COI_02 and/or COI_h10 haplotype translations. On position 78, no consensus in the amino acid sequence could be found, since there were three different amino acids present Q (Gln, Glutamine), K (Lys, Lysine) and P (Pro, Proline), from which P was just found in COI_h33.

With DnaSP v. 6 (Rozas et al. 2017) common parameters on genetic diversity were obtained and are integrated in **Table 14**. Of the total 360 sites, 346 were considered for mtCOI-360 bp DNA-marker (number of haplotypes: h: 33, haplotype diversity: Hd: 0.7321)

Table 14. Summary of common parameters of genetic diversity of mtCOI-360 bp.

populations \ parameters	n	S	h	Hd	π
Austria	8	0	1	0.00000	0.00000
Canada	80	1	2	0.02500	0.00007
China	205	19	24	0.81000	0.00658
France	57	3	3	0.38534	0.00217
Germany	129	3	5	0.56892	0.00195
Italy	37	4	5	0.65766	0.00453
South-Korea	57	5	4	0.68233	0.00550
Switzerland	67	1	2	0.50475	0.00146
unknown	57	6	11	0.75063	0.00618
USA	262	4	5	0.60952	0.00486
Total Data Estimates	959	22	33	0.73206	0.00516

10 populations were included in this analysis of gene flow and genetic differentiation. n = number of sequences, S = segregating sites, h = number of haplotypes, Hd = haplotype diversity, π = nucleotide diversity.

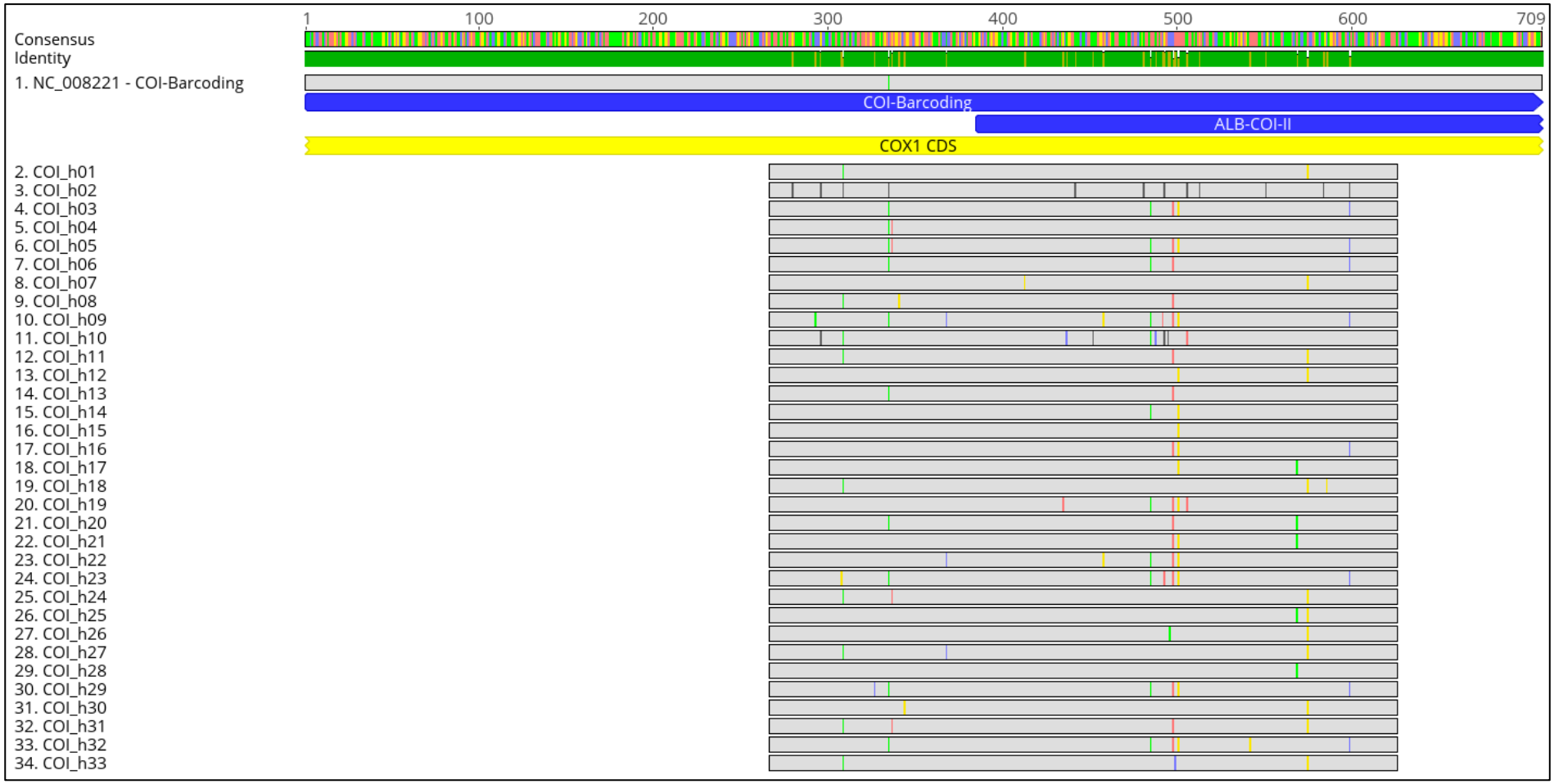


Figure 22. Alignment of 33 haplotypes of the 360 bp mtCOI DNA-marker with the reference sequence of the whole COI-barcoding region gained with the primers of Folmer et al. (1994) in Geneious Prime® v. 2021.0.1.
809 sequences were received from the global BOLD dataset and were combined with the 150 sequences of the European set (959 ALB sequences in total). Disagreements to the consensus sequence were marked here coloured according to their base (G = yellow, C = blue, T = green, A = red).

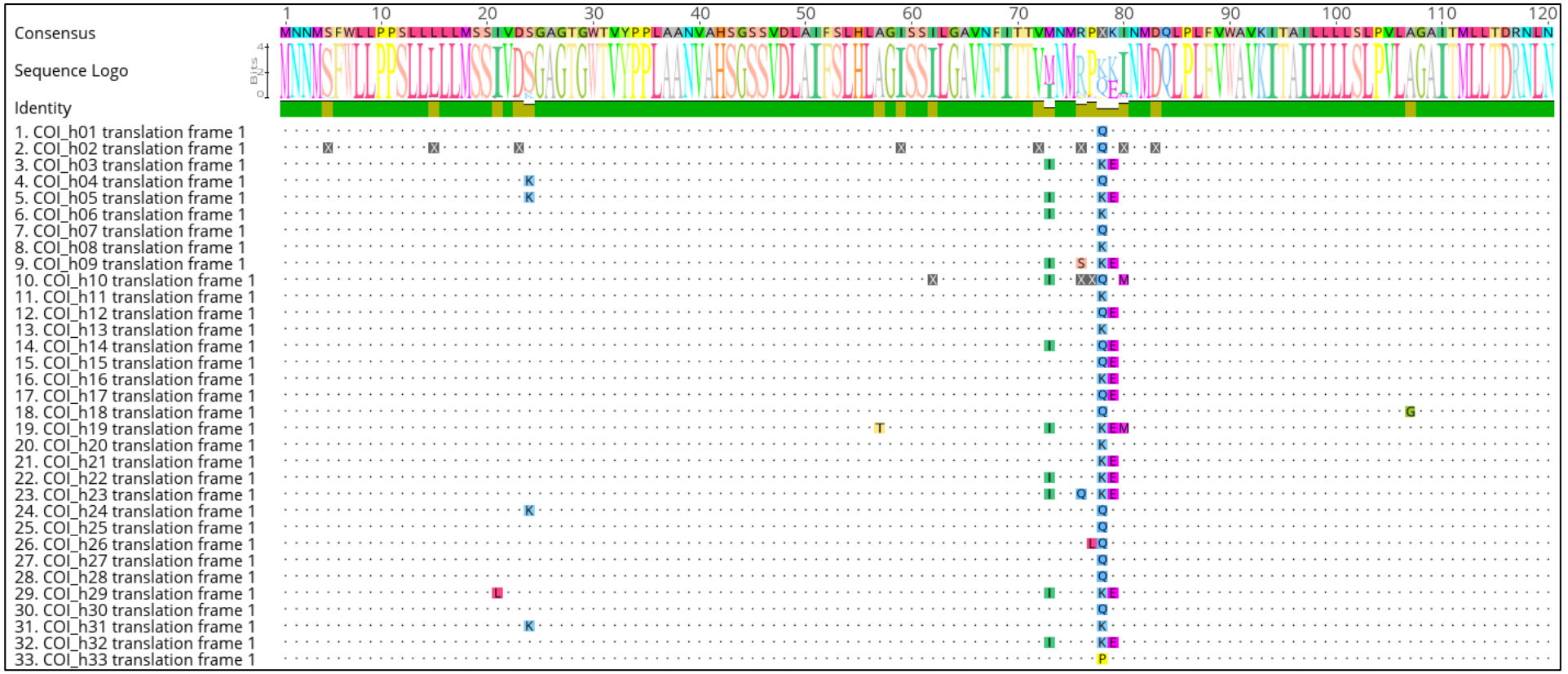


Figure 23. Alignment of the 33 haplotypes of mtCOI 360 bp DNA-marker from 959 global sequences from BOLD combined with the dataset of this study in Geneious Prime® v. 2021.0.1.
Genetic code for translation: invertebrate mitochondrial (translation table 5). Disagreements to the consensus sequence were marked here coloured according to their amino acid.

The abundance and connections of the described 33 global haplotypes are shown in more detail in TCS networks (**Figure 24**, **Figure 92**), geographic haplotype projections (**Figure 25**, **Figure 26**, **Figure 27**, **Figure 29**) and in a Venn diagram (**Figure 28**).

All information about location affiliation was used as mined from BOLD/NCBI and publications (Carter et al. 2009c; Carter et al. 2010; Javal et al. 2019a) and sometimes readjusted by google maps information. From the 959 sequences represented in the alignments, 57 of them were without country information. All sequences without country information were unpublished. From the remaining 902 sequences with knowledge on the country of origin, from 27 of them no further information on the collection site was available and some of the 33 haplotypes were not represented anymore (h15, h16, h17, h20, h21, h25, h28). For these 902 sequences, a TCS network was calculated (**Figure 24**), showing the country affiliations per haplotype and the exact number of mutations between them. For 875 sequences, all information including the collection site were available. With the 875 sequences with knowledge on city level also a TCS network was drawn (Appendix, II Chapter A, **Figure 92**) as well as haplotype projections on the maps from Asia, North-America and Europe (**Figure 25**, **Figure 26**, **Figure 27**, **Figure 29**). From the 875 sequences the haplotypes h19, h27 and h32 were not represented anymore.

When having a look on the TCS network **Figure 24**, which is based on 902 sequences, it is clearly visible that all haplotypes are highest three mutations different from the most abundant haplotypes COI_h01, h02 or h03. COI_h01 and COI_h02 are just one mutation apart, whereas COI_h03 is three mutations from h02 and four mutations from h01 apart. The most abundant haplotype COI_h01 is represented in seven countries, h02 in eight countries and h03 in just four countries. COI_h01 was not found in Austria and France, but h07 from France is just one mutation apart. COI_h02 was not found in Canada, but in all other countries. Austria inhabited just the h02 haplotype. COI_h03 was not found in Canada, Austria, France, Germany or Switzerland. All the unique haplotypes were from China. One haplotype COI_h07 just occurred in France and COI_h05 was found mainly in South-Korea. The haplotype COI_h04 is mostly represented by the Germany population D-ST-MB, but also occurred in China and France. In general, it is also observable that all haplotypes are just a few mutations apart from each other.

The haplotype projections (**Figure 25**, **Figure 26**, **Figure 27**, **Figure 29**) show in more detail the global distribution of the 33 described haplotypes in four figures.

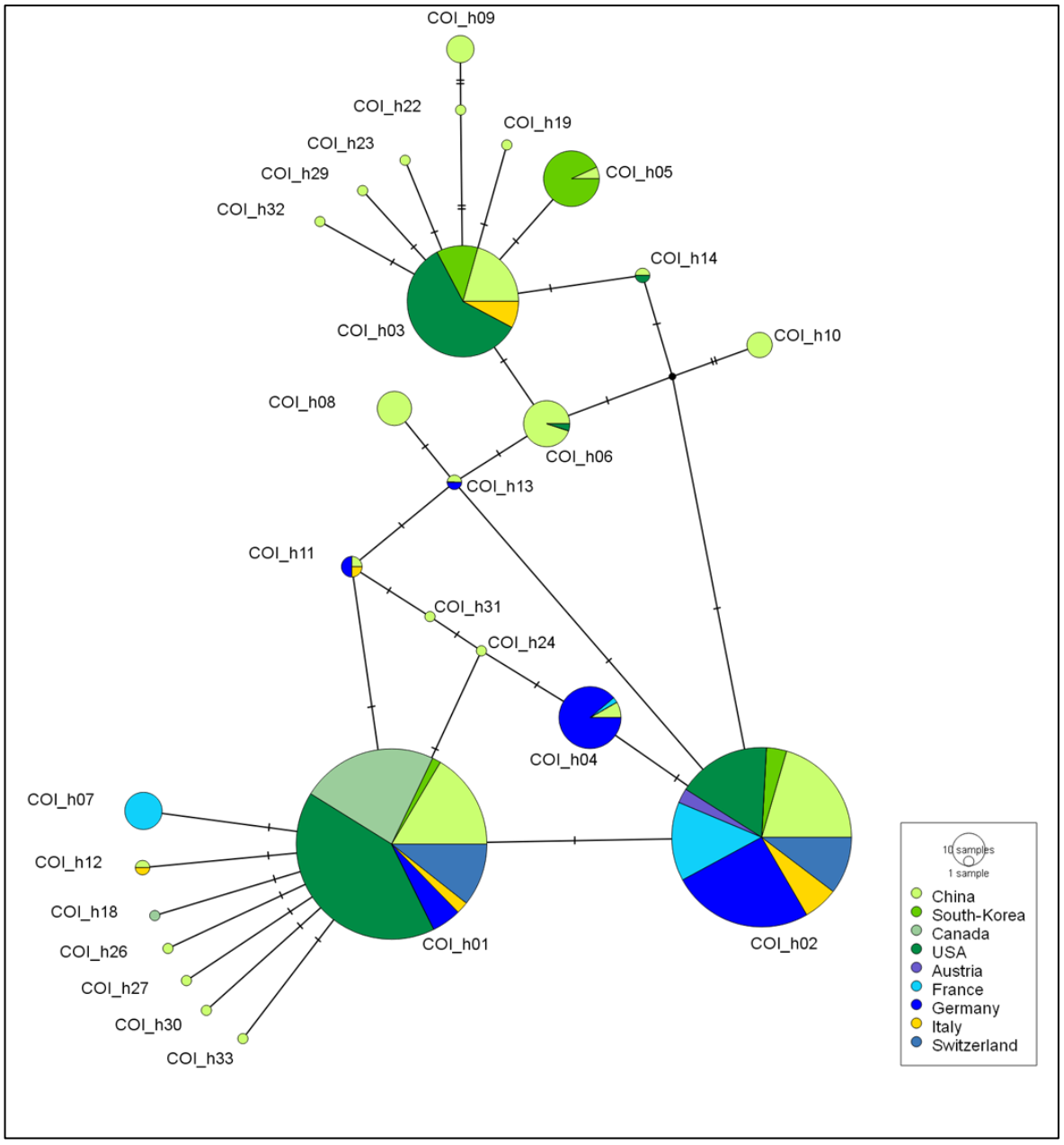
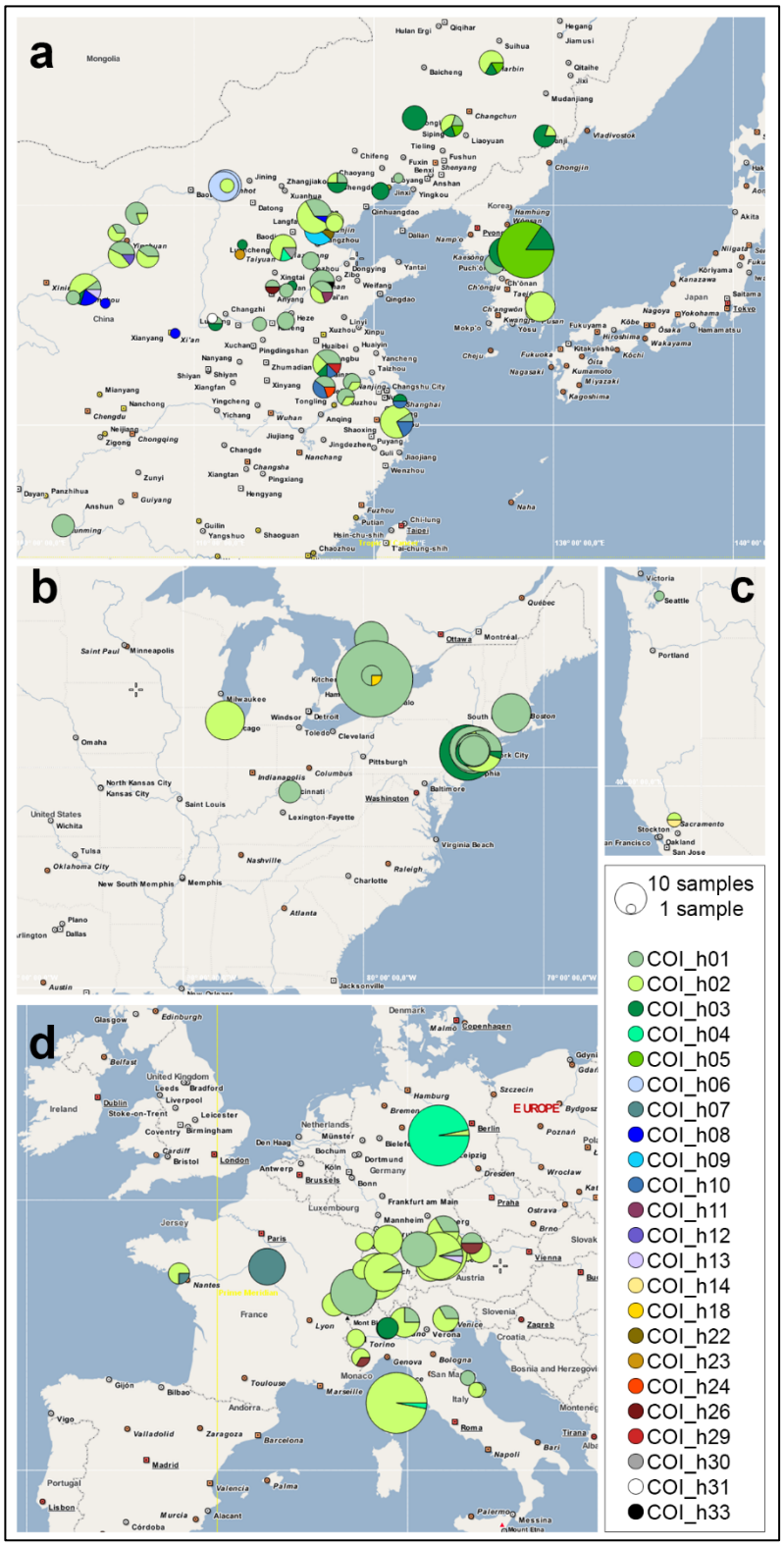


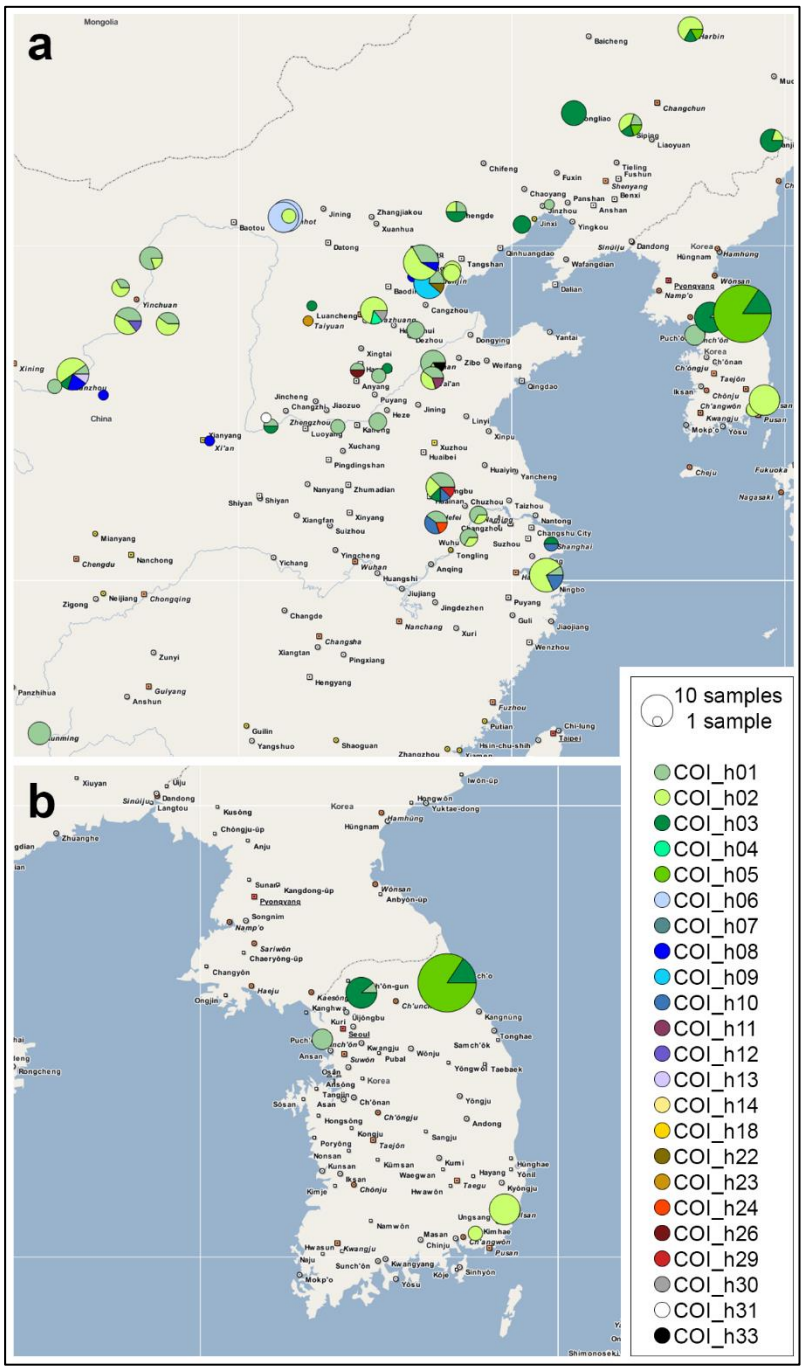
Figure 24. TCS haplotype network of 26 haplotypes of mtCOI-360 bp from 902 ALB specimens from Asia, Europa and North-America (PopART v. 1.7).
Trait colours are chosen per country (9). Mutations are represented by hatch marks; the size of the circles is proportional to the number of specimens contributing to the haplotype.



First, **Figure 25** of the haplotype projections shows an overview of the distribution on all three continents at once. It is directly observable that in Asia, especially in China, most of the locations inhabited more haplotypes than only one or two, which is mostly the case in North America and Europe. More than two haplotypes on one location collection site were not observable for Europe.

Figure 25. Haplotype projection to the global map, showing the occurrence of the 33 haplotypes of mtCOI-360 bp from 875 ALB specimens.
a) Asia; **b, c)** North-America; **d)** Europe (PopART v. 1.7).

Figure 26 shows the haplotype distribution on the Asian continent in more detail. This haplotype distribution map displays, that South-Korea only inhabits four different haplotypes. Beside some specimens with COI_h03, COI_h05 is the most dominant haplotype from Seorak mountain national park near Sokcho, whereas the South of South Korea only showed COI_h02 in Gimhae and Ulsan. COI_h01 and COI_h03 occurred in Northwest of South Korea in Incheon and Pocheon.



In China, most of the locations exhibit several haplotypes in the same collection site, often with some of the unique haplotypes h18-h33. The haplotype COI_h06 was only detected in Hohot in Inner Mongolia. North-East of China (Harbin, Tongliao, Lishu, Yanji, Linhai, Xingcheng) showed solely the same haplotypes that were also abundant in South-Korea. In other more Western parts of China the haplotype COI_h03 was rare, while h01 and h02 are most dominant in all parts of China.

Figure 26. Haplotype projection to the map of Asia, showing the occurrence of the 33 haplotypes of mtCOI-360 bp from 875 ALB specimens.
a) China and South Korea; **b)** South Korea (PopART v. 1.7).

In **Figure 27** it is obviously notable, that most of the North-American sites showed the haplotypes COI_h01, h02 and h03, with three exceptions. In Amityville in New York the COI_h06 haplotype was found once (abundant in Hohot, China) and the haplotype h14 was found once in Sacramento (abundant in Hubei province, China), and h18 in Vaughan (Northview) in Canada. COI_h18 only occurred one time in the whole dataset but is only one mutation apart from COI_h01. In the rest of Canadian (also including Vaughan) collection sites, only COI_h01 is present. In New York city, most dominant haplotype was h01, but in Eastern parts of New York (Long Island City, Massapequa, Amityville, Bayside) also h02 was very dominant, whereas h02 is the only haplotype found in Chicago and the other finding in Sacramento. The finding in Seattle was also COI_h01. In Linden and Carteret in New Jersey, the haplotype COI_h03 was most dominant.

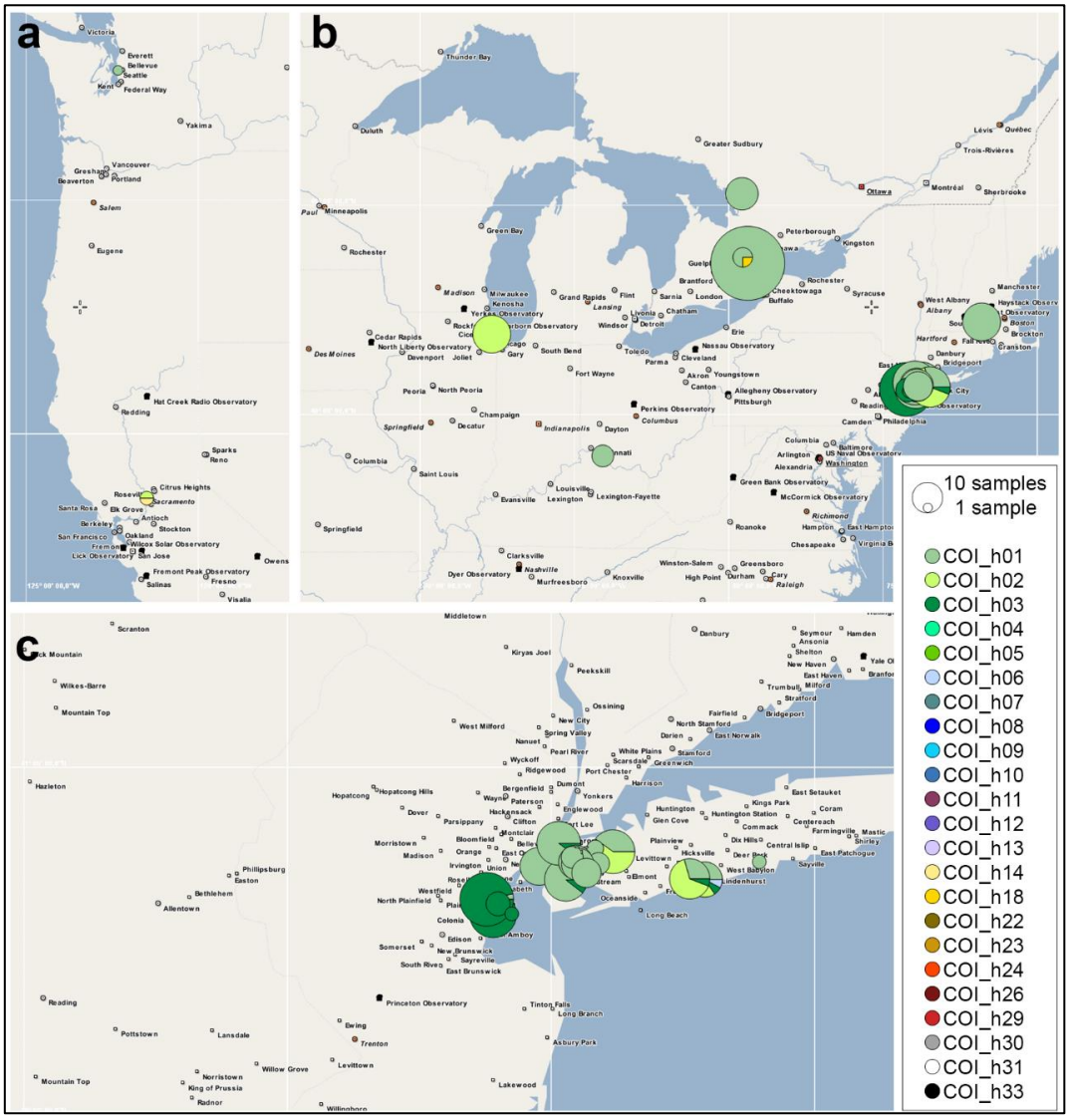


Figure 27. Haplotype projection to the map of North-America, showing the occurrence of the 33 haplotypes of mtCOI-360 bp from 875 ALB specimens.
a) West-USA**; b)** North-East USA, Canada; **c)** New York, New Jersey (PopART v. 1.7).

In Europe (**Figure 29**) the most abundant haplotype was COI_h02, which was found in all five sampled countries. COI_h04 only occurred in Magdeburg and Furiani (Corsica), which is also present in China, Liaoning province (no further information available) and Shijiazhuang (Hebei province). Magdeburg had just one sequence with COI_h02, whereas Furiani had mostly h02 and just one h04. The haplotype h07 which only occurred in Gien in France and one time in Sainte-Anne-sur-Brivet (France). This haplotype was not detected anywhere else in the dataset and is only one mutation apart from h01. The rare haplotype h11 was found in Neukirchen am Inn (Germany) and Cuneo (Italy), while in China it is present in Tai'an, Shandong province. Haplotype COI_h03 appeared only in Milano, Corbetta and Vittuone, as well as in South Korea and Northeastern China. COI_h12 occurred only in Milano and Qingtongxia, China. Another rare haplotype, COI_h13, was solely present in Neubiberg (Germany) once and in Lanzhou (China) (among five other haplotypes).

The Venn diagram **Figure 28** identified that the only haplotypes shared between all three continents are COI_h01, h02 and h03. Asia has the highest number of unique haplotypes, whereas North America only showed h18 in Canada and in Europe only h07 in France. Asia and Europe shared h04 (Magdeburg, Furiani, Liaoning province, Shijiazhuang), h11 (Neukirchen am Inn, Cuneo, Tai'an), h12 (Milano, Qingtongxia) and h13 (Neubiberg, Lanzhou), whereas Asia and North America shared h06 (Amityville in New York, Hohot) and h14 (Sacramento, Hubei province). There were no haplotypes shared solely between North America and Europe.

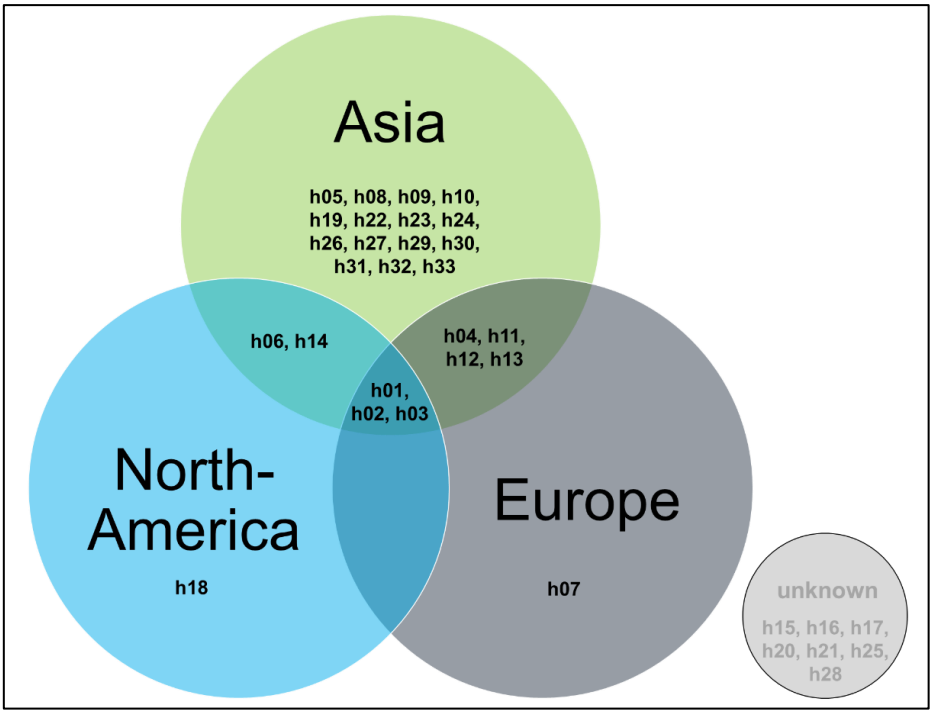


Figure 28. Venn diagram on the distribution of the 33 mtCOI-360 bp DNA-marker haplotypes on the three continents where *A. glabripennis* is abundant - Asia, North-America and Europe, as well as the haplotypes with unknown origin.

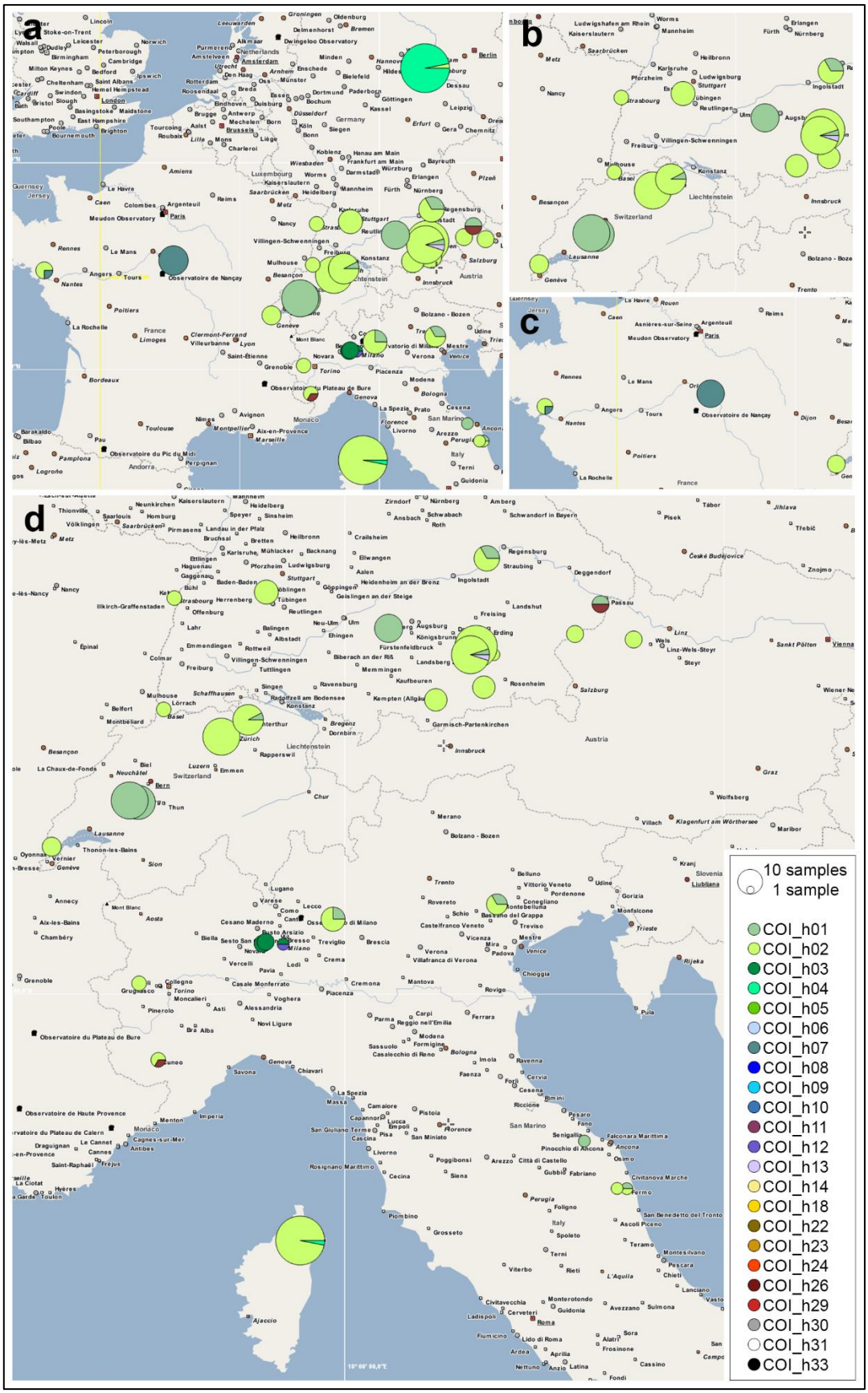


Figure 29. Haplotype projection to the map of Europe, showing the occurrence of the 33 haplotypes of mtCOI-360 bp from 875 ALB specimens.
a) total; **b)** Mid-Europe**; c)** France without Corsica; **d)** Europe without Magdeburg (PopART v. 1.7).

The genetic structure of the populations was also investigated by AMOVA, like it was done with the whole mtCOI-658 bp DNA-marker. Here, the two genetic structure scenarios panmixia and a restricted mating between the different continents were tested. The AMOVA **Table 15** showed very high molecular variance of >70% in both tested structures within the populations, whereas the variation among the populations was estimated to be lower in total (24.27%), also within the continents (21.28%). The lowest molecular variance was detected among the three continents Asia, North-America and Europe. The measures to determine the degree of differentiation, the Fixation indices F_{ST}, F_{SC} and F_{CT} are also shown in **Table 15**. The highest degree of differentiation was estimated in the hierarchical approach with F_{ST}-values to show the differentiation among the demes with >0.26. These were lower in the approach without structure assumption with 0.24. In general, the differentiation measured was high, but less than with the whole mtCOI-658 bp DNA-marker (**Table 12**). The differentiation among the populations within a continent (F_{SC}=0.22) is similarly high as within the populations. Otherwise, the degree of differentiation among the continents was very low (F_{CT}=0.05).

Table 15. AMOVA design and results on the dataset of 959 sequences from Europe and the whole BOLD and NCBI databases with the DNA marker mtCOI-360 bp conducted in Arlequin v. 3.5.

Source of variation	d. f.	Sum of squares	Variance components	Percentage variation	Fixation indices	p values
(a) AMOVA - No. of groups 1 (959SEQ-COI)						
Among populations	9	325.358	0.39137 (Va)	24.27***		0.00000
Within populations	949	1159.120	1.22141 (Vb)	75.73	[1]0.24267***	0.00000
Total	958	1484.478	1.61279			
(b) AMOVA - No. of groups 3 (959SEQ-COI)						
Among continents	2	161.960	0.08487 (Va)	5.26	[3]0.05257	0.16031
Among populations within continents	6	152.498	0.34357 (Vb)	21.28***	[2]0.22464***	0.00000
Within populations	893	1058.979	1.18587 (Vc)	73.46***	[1]0.26540***	0.00000
Total	901	1373.438	1.61431			

Significance tests (1023 permutations): *** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.1$. [1]F_{ST}, [2]F_{SC}, [3]F_{CT}
(a) panmixia scenario (b) groups: Asia, Europe, North-America; group of unknown origin excluded. Distance method: pairwise difference

The **Figure 30** represents the pairwise differentiation between the subpopulations (F_{ST}) using pairwise difference as distance method. The values for the degree of differentiation are pointed out with grades of colouring from low to high. The exact F_{ST}-values (**Table 27**) and p-values (**Table 28**) are given in Appendix, II Chapter A. Most of the pairwise F_{ST}-values were significant. The comparison between Austria and Italy, Germany and France, as well as between China and Italy and between "unknown" and Italy were not significant. The highest differentiation was observable between Canada and Austria, but Canada was also highly differentiated from South-Korea, Italy, Germany and France. The lowest differentiation was found between "unknown" and China. USA has lower differentiation <0.1 in comparison to China, Italy, Switzerland and "unknown". Moderate differentiation was found between the European countries Germany, France, and Italy, whereas Switzerland was slightly more differentiated from all of them. South Korea showed the highest differentiation to most of the countries, especially compared to Canada.

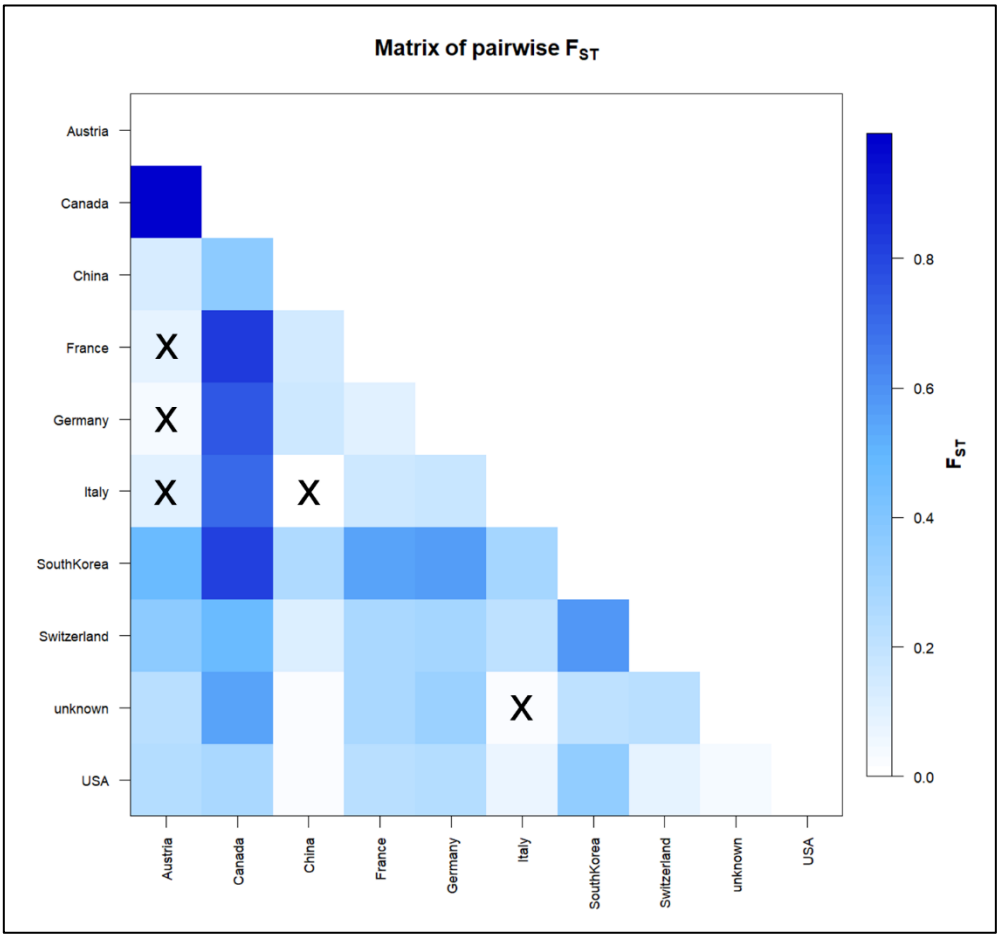


Figure 30. Population differentiation on 33 haplotypes (mtCOI-360 bp) conducted in Arlequin v. 3.5.
Pairwise F_{ST}-values among countries. x = not significant (significance level=0.05).

The NeighborNet phylogenetic network (**Figure 31**) drawn with Splitstree v. 4.17.1 gives a small overview about the phylogenetic connections of the haplotypes. Most of the haplotypes are cross-linked to each other, and very few are more divergent.

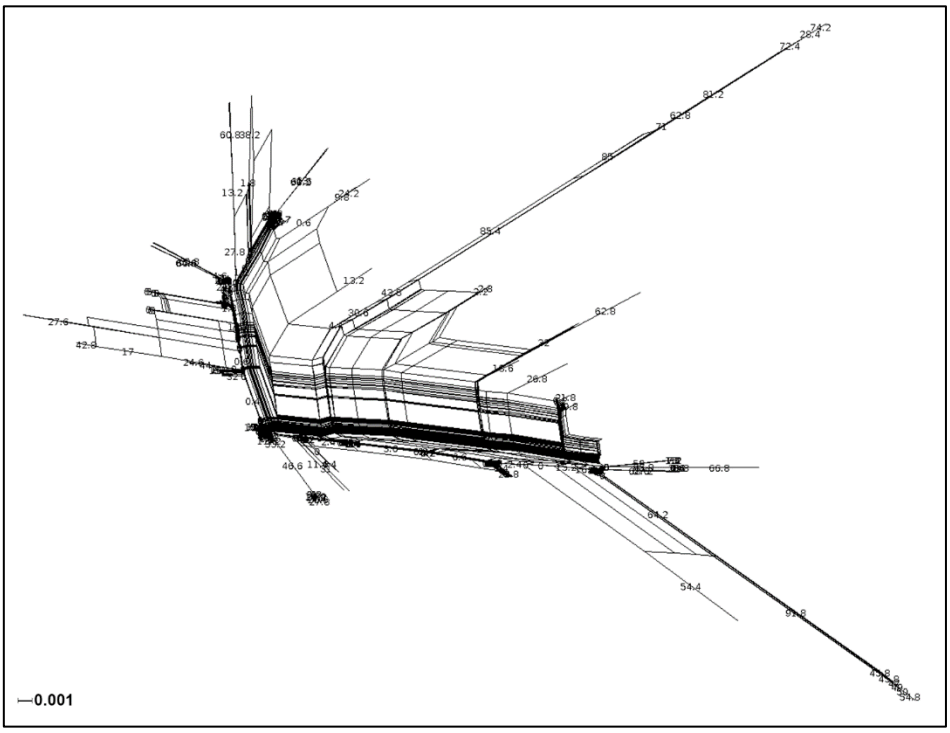


Figure 31. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on 959 sequences representing 33 haplotypes of mtCOI-360 bp.
The ɣ-distributed HKY (+G 0.37) with a proportion of invariable sites (+I 0.76) distance and Equal Angle were used to draw the network with 567 splits; Fit = 93.396; 500 bootstraps. 959 Taxa and 360 characters are shown.

5.2.3 Divergence within mtCOI-marker in the genus Anoplophora

To have an evolutionary perspective on the sequence divergence in mitochondrial COI sequences within the genus *Anoplophora*, all available *Anoplophora* spp. sequences for mtCOI were obtained from BOLD. After the alignment, the sequences of *A. freyi, A. malasiaca* and *A. nobilis* needed to be removed completely, as well as some single sequences of the other species, since they did not cover enough of the COI-5P barcoding marker region. Subsequently, 74 sequences with a length of 618 bp could be used for the evolutionary analyses, from which nine were the haplotypes COI_Type1-8 and the reference of ALB.

In Mega X (Kumar et al. 2018) the estimates of evolutionary divergence were calculated using the Tamura 3-parameter model (Tamura 1992). The results are depicted in **Table 16**. The sequence of *A. flavomaculata* is most distinct from the other species (~18-22%). With 20.188% different nucleotides, this species is most distinct from *A. glabripennis*, while ALB has ~8-10% differences to the others. With 8.08% different nucleotides *A. glabripennis* is from all

represented *Anoplophora* species most similar to *A. davidis*. In comparison to the other species, where the differences are between ~7-22%, most similarities are found between *A. macularia* and *A. davidis* (2.762%), as well as between *A. chinensis* compared with *A. davidis* (3.345%) and *A. macularia* (3.513%).

Table 16. Estimates of evolutionary divergence between 74 sequences from *Anoplophora* spp. of mtCOI DNA-marker conducted in MEGA X converted into percentage.

	Anoplophora spp.	1	2	3	4	5	6	7	8	9	10	11
1	*A. glabripennis*											
2	*A. davidis*	8.087%										
3	*A. flavomaculata*	20.188%	18.072%									
4	*A. horsfieldii*	8.573%	11.177%	19.557%								
5	*A. macularia*	8.390%	2.762%	18.562%	11.532%							
6	*A. chinensis*	8.670%	3.345%	18.494%	11.358%	3.513%						
7	*A. beryllina*	9.664%	10.388%	21.914%	10.799%	10.476%	10.412%					
8	*A. stanleyana*	8.316%	7.424%	19.431%	10.423%	7.731%	7.235%	8.999%				
9	*A. elegans*	9.465%	8.085%	19.844%	12.095%	8.406%	8.424%	10.273%	7.552%			
10	*A. lurida*	10.022%	11.148%	19.178%	11.819%	11.422%	11.797%	9.842%	9.937%	11.308%		
11	*A. granata*	10.286%	11.872%	21.847%	12.366%	12.715%	12.184%	9.696%	9.386%	11.174%	9.946%	

The number of base substitutions per site from averaging over all sequence pairs between groups are shown. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992). This analysis involved 74 nucleotide sequences with a total of 618 positions in the final dataset.

The evolutionary tree drawn by the maximum likelihood method is shown in **Figure 33** as bootsrap consensus tree. Prior model testing of the alignment (74 sequences, 618 bp) measured the lowest BIC score on the nucleotide substitution model GTR+G+I (General Time Reversible model, Gamma distribution, some sites evolutionarily invariable), so this model was chosen to describe the substitution pattern best. In the tree, all branches with bootstrap replicates lower than 90% were collapsed. In the grey boxes on **Figure 33** with question marks, *Anoplophora* species where marked which occurred in clades that are distant to the other mtCOI DNA-marker sequences of their respective species, including sequences of *A. chinensis*, *A. macularia*, *A. davidis* and *A. horsfieldii*. In light green, all nine haplotypes of this study from *A. glabripennis* are represented, also showing COI_Type6 as being most distant from the other ALB haplotypes. In this evolutionary tree *A. horsfieldii* exhibits one branch that is nearest to the ALB haplotype COI_Type6, but this species also showed another branch closer to the bottom of the tree. Since in the original tree (not shown) the highest

number of substitutions per site was observed between *A. flavomaculata* and the other species, the bootstrap consensus tree was rooted by *A. flavomaculata*.

The NeighborNet phylogenetic network (**Figure 32**) drawn with Splitstree v. 4.17.1 supports the findings from the Maximum likelihood tree and gives more detail insight in the phylogenetic connections of the *Anoplophora* spp. First, it is observable that *A. chinensis* showed an own branch cluster including two sequences from *A. macularia*. *A. macularia* also had its own cluster but included sequences of *A. chinensis* and *A. davidis*. Another *A. davidis* sequence was detected to be in between *A. chinensis* and *A. macularia*. *A. glabripennis* forms its own cluster far distant from the others, together with one *A. horsfieldii* sequence mentioned before, which is split from ALB with a high confidence. All the other species are far distant from the others with high confidence and only have cross-links directly after splitting from the rest.

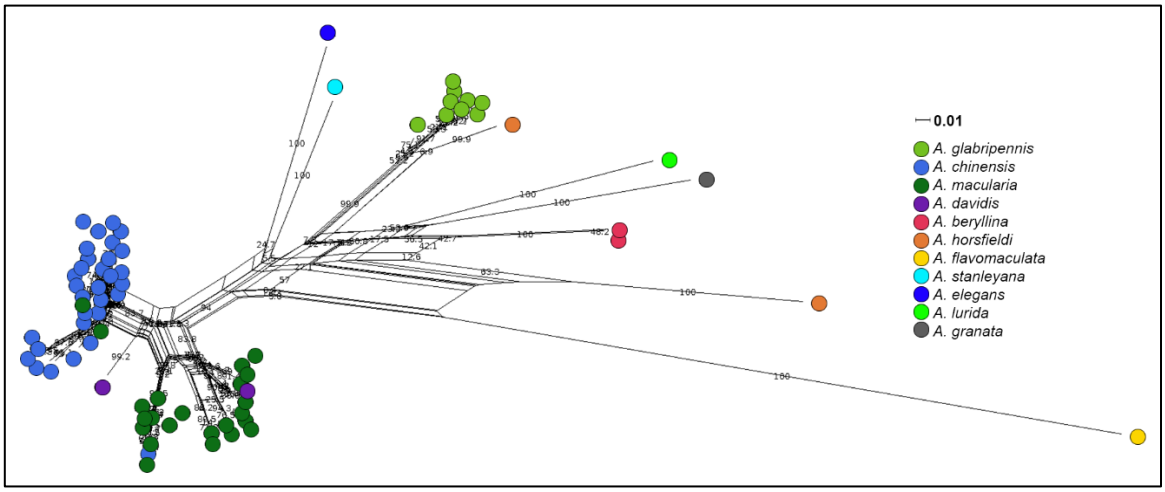


Figure 32. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on 74 sequences of *Anoplophora* spp. and the ALB haplotype sequences COI_Type1-8 and the reference from NCBI.
The ɣ-distributed (+G 0.23) HKY distances and Equal Angle were used to draw the network with 189 splits; Fit = 96.896; 1000 runs Bootstrap. 74 taxa and 618 characters are shown.

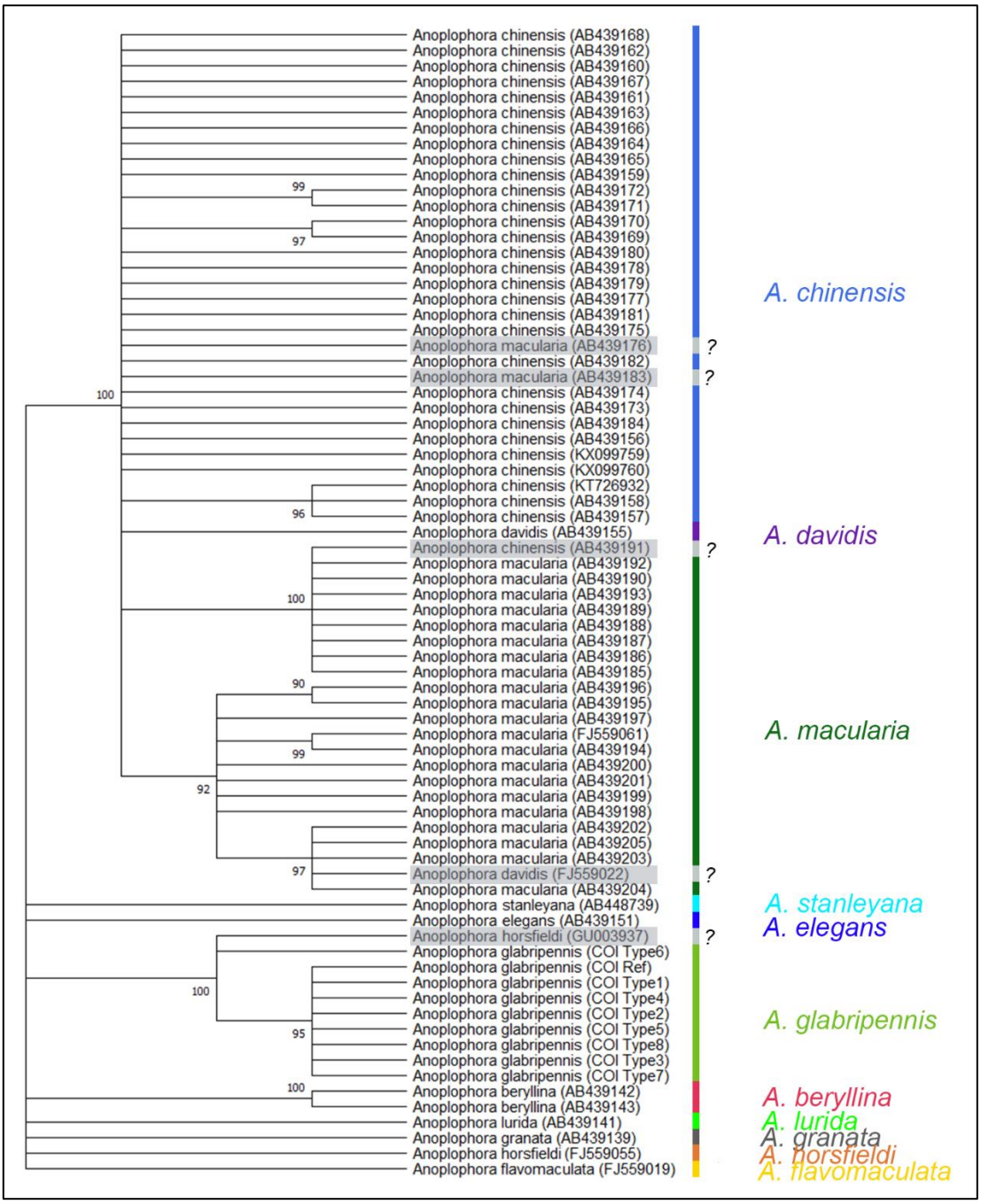


Figure 33. Evolutionary history inferred by using the Maximum Likelihood method on 74 sequences of mtCOI *Anoplophora* spp. conducted in MEGA X (Kumar et al. 2018).
The General Time Reversible model (Nei and Kumar 2000) was used for the pairwise distance estimation to draw the tree and bootstrap method to test the results. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The bootstrap consensus tree is shown here, and branches reproduced in less than 90% bootstrap replicates are collapsed. This analysis involved 74 nucleotide sequences. There was a total of 618 positions in the final dataset.

5.3 Chapter B: results

Chapter B

"Mitochondrial insertions in the genome of the invasive long-horned beetle *Anoplophora glabripennis* – genetic diversity of NUMTs and their impact on DNA barcoding"

5.3.1 ALB-COI-II fragment as potential DNA-marker

The intention of the study was to develop an easy hands-on-tool for diagnosis to make it easier to distinguish between different populations of ALB with a DNA-marker of more length and more specificity to ALB than the COI barcoding DNA-marker (5.2 Chapter A: results). Consequently, ALB specific primers were designed for a DNA-marker from the same region but extending the size to the beginning of Cytochrome oxidase subunit 2 (COII) and the tRNA sequence in between to have more possibilities of population specific variations. The partial COI and COII CDS with the tRNA sequence in between were also used in other studies of ALB before (Carter et al. 2009c; Carter et al. 2010; Lee et al. 2020). The proportion of the specimens sequenced for this purpose is pictured in **Figure 34**.

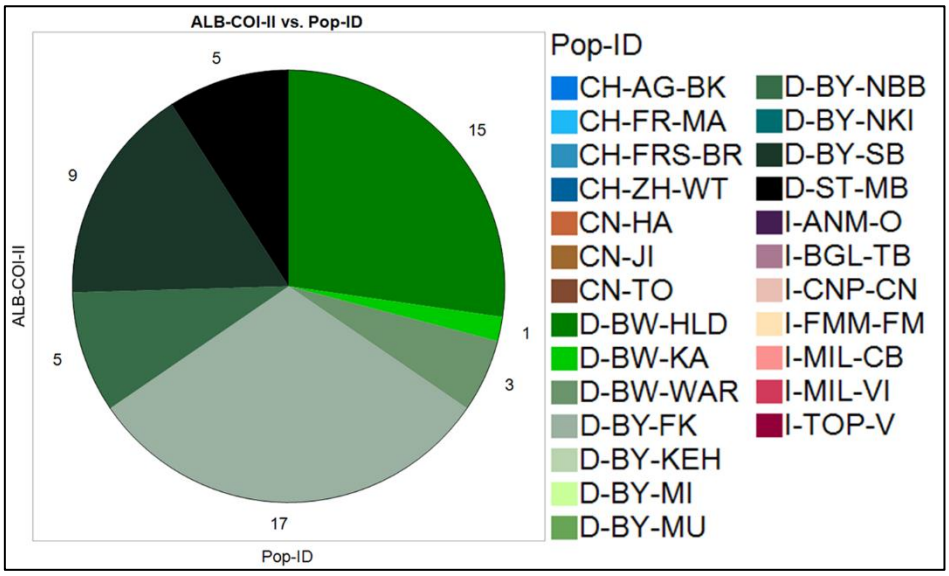
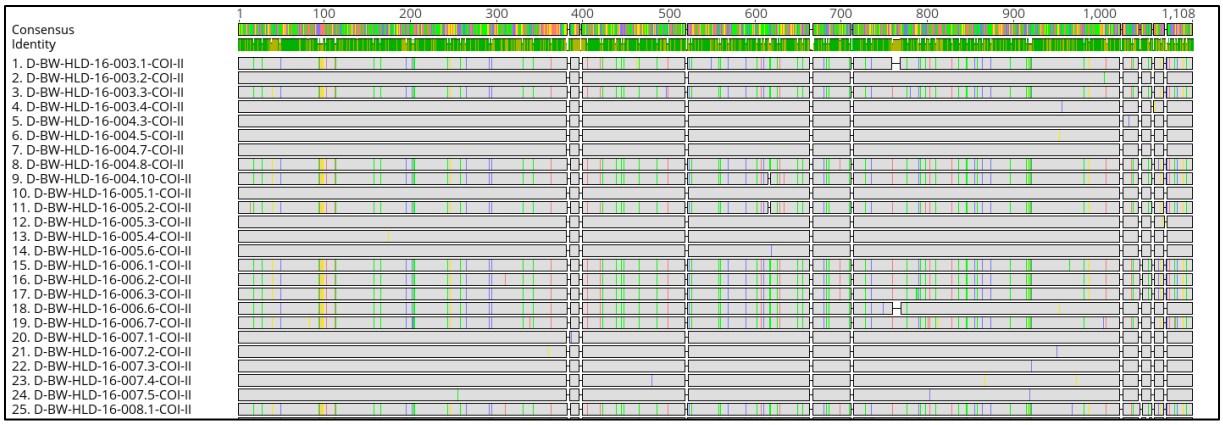


Figure 34. Represented ALB-populations (Pop-ID) from ALB-COI-II DNA-segment.

All efforts of direct PCR product sequencing of ALB-COI-II failed because of aborts or overlapping peaks, so all PCR products were cloned before sequencing. Surprisingly, most sequenced clones showed several sequences for the same specimen, which were also very different to each other. In **Figure 35** huge differences in nucleotides as well as in the length between the clones for a specimen are emphasized on a few examples of the whole alignment compared to the mitochondrial reference of ALB-COI-II, as well as to a nuclear NUMT sequence, which will be described later (5.3.2, 5.3.3). Despite the huge differences, all the 198 sequences were still similar enough to align.



[…]

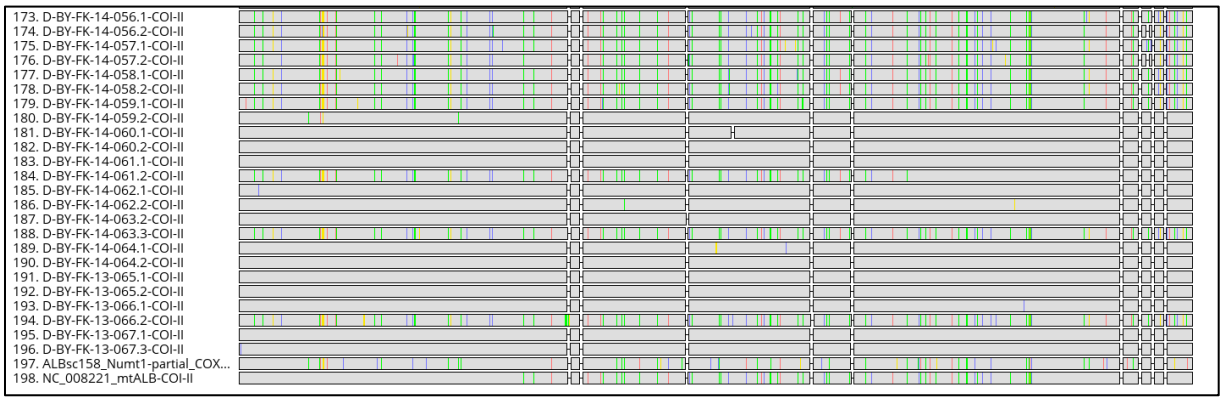


Figure 35. Partial alignment of 196 clone sequences of ALB-COI-II sequences from 55 ALB specimens of seven German populations together with the partial reference NC_008221 in Geneious Prime® v. 2021.0.1 and a nuclear mitochondrial-like sequence (NUMTs). Disagreements to the consensus sequence were marked here coloured according to their base (G = yellow, C = blue, T = green, A = red, - = gap).

The high genetic diversity of the ALB-COI-II sequences was also reflected by the translation to amino acid sequences presented on the DNA-alignment. Frame +3 was used as suggested by the mitochondrial reference of ALB-COI-II. The translation into amino acid emphasized even more severe disturbances due to indels and the high amount of nucleotide substitutions. In the alignment sections displayed in **Figure 36**, it is clearly visible that the amino acid differences in some sequences were quite high in comparison to the consensus, including the reference of ALB-COI-II and a NUMT sequence. For some sequences the amino acid translation showed also early abortions due to stop codons, while other sequences did not show any stop codons. Even the NUMT sequence showed the first stop codon on DNA sequence position 629 like the mitochondrial reference (not shown).

[…]

[…]

Figure 36. Partial view on amino acid translation on the alignment **Figure 35** in Geneious Prime® v. 2021.0.1.
Amino acid translation (+3 frame) according to mtALB-COI-II frame of mitochondrial reference sequence by using the genetic code: invertebrate mitochondrial (translation table 5). Disagreements to the amino acid consensus sequence were marked here colored according to their amino acid, where black bars are stop codons.

A maximum likelihood evolutionary tree (T92 model, according to BIC, **Figure 37**) of all 196 cloned ALB-COI-II sequences, together with a NUMT and the reference sequence surprisingly expressed that the sequences form roughly two haplotype clusters, named Cluster A (green) and Cluster B (blue) from here on. Both clusters inhabited one group with the most sequences in them, which were less distant within and a smaller one which was more distant to the main clusters ("Cluster A" and "Cluster B"). Neither the NUMT sequence, nor the mitochondrial reference showed low distance to the clone sequences of ALB-COI-II, but the mitochondrial reference is less distant to Cluster A/"Cluster A", while the NUMT sequence is less distant to Cluster B/"Cluster B".

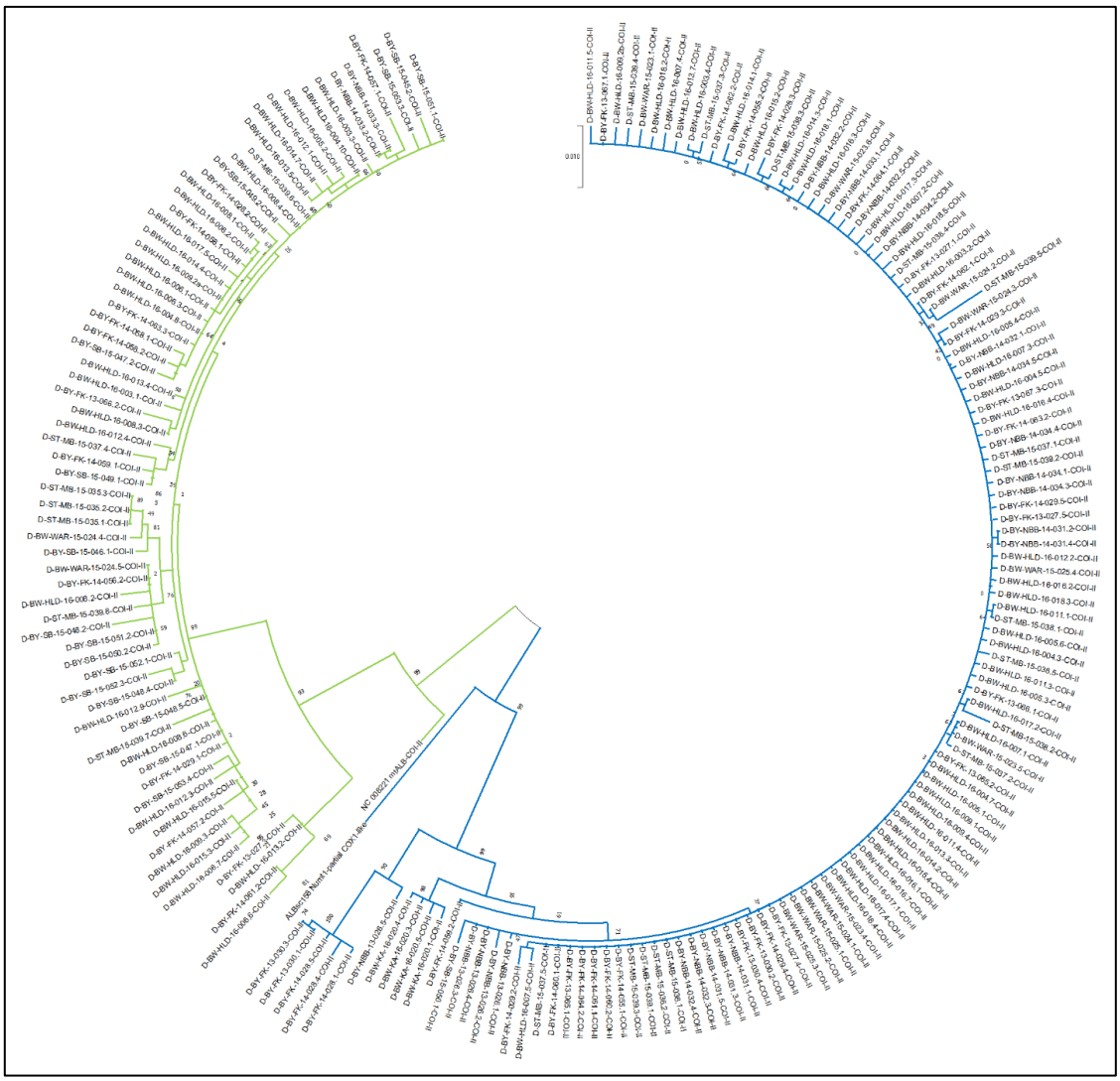


Figure 37. Evolutionary history inferred by using the Maximum Likelihood method on 196 clone sequences of ALB-COI-II, the reference and a NUMT sequence conducted in MEGA X (Kumar et al. 2018).
The Tamura 3-parameter model (Tamura 1992) was used for the pairwise distance estimation to draw the tree and bootstrap method was used to test the results. The branch lengths measure in the number of substitutions per site. This analysis involved 198 nucleotide sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 1096 positions in the final dataset.

The NeighborNet phylogenetic network (**Figure 38**) drawn with Splitstree v. 4.17.1 supports the clusters found by the maximum likelihood tree. But, the Splitstree emphasizes more detailed, that there are no cross-links between the clusters, the mitochondrial ALB-COI-II and the NUMT sequence. Accordingly, the phylogenetic connections are all distant to each other. Also, for the smaller "clusters" it is shown more explicitly in the network, that "Cluster B" is genetically closer to Cluster B, than "Cluster A" to Cluster A, which is genetically nearer to the mitochondrial ALB-COI-II reference.

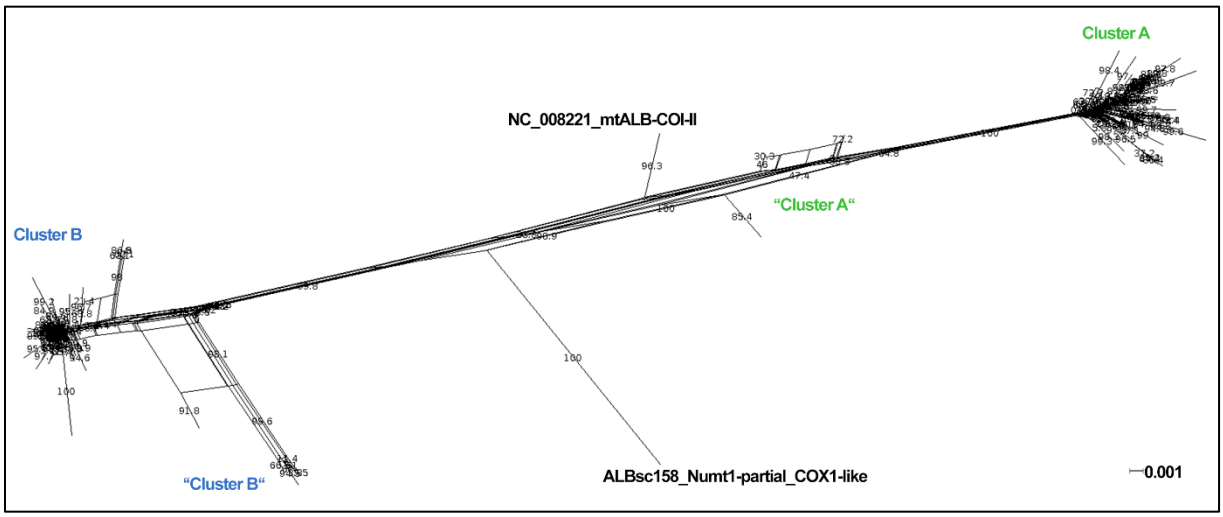


Figure 38. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on 198 sequences representing the alignment **Figure 35** of ALB-COI-II.
The ɣ-distributed (+G 0.7) HKY distances (Hasegawa et al. 1985) and Equal Angle were used to draw the network with 409 splits; Fit = 98.765; 1000 runs of Bootstrap method. Cluster labels were chosen according to the Maximum likelihood tree (**Figure 37**). 198 taxa and 1108 characters are shown.

5.3.2 Discovery of NUMTs in ALB genome

Considering the unexplained high genetic variation detected for ALB-COI-II, the circumstances were studied further to find the background for these results. The length differences of some of the clone sequences as well as some of the sudden aborts of the amino acid translations gave clues to watch out for nuclear mitochondrial-like sequences (NUMTs). This was first explored by megaBLAST of some example ALB-COI-II sequences as described in 4.5.1. The BLAST results of 6 different ALB specimens are demonstrated in Appendix, III Chapter B, **Table 29** for clone sequences of ALB-COI-II. For these six specimens, one clone sequence of ALB-COI-II Cluster A and Cluster B (**Figure 37**, **Figure 38**) were used respectively. The six cloned sequences of Cluster A had the most similarities (~96%) with the mitochondrial ALB genome (NC_008221.1) and less (~91-92%) to sequences of the complete COI CDS or COI-tRNA-Leu-COII partial CDS sequences. Some of the sequences of COI or COI-II from the nucleotide collection with ~92% identity, were not from *A. glabripennis*, for instance EU914812.1 (*A. chinensis*) and KF737826.1 (*A. lurida*). When the search was limited to

representative genomes (GCA_000390285.2, NC_008221.1), the Cluster A sequences showed similarities (~92-93%) with ALBscaffold158 (NW_019416265.1) and matches (~91%) with ALBscaffold1356 (NW_019417445.1). The cloned Cluster B sequences exhibited ~94-95% identity with the mitochondrial ALB genome, >99% identity with COI complete CDS or COI-tRNA-Leu-COII partial CDS, ~94% identity with scaffold158 and ~92% identity with scaffold1356.

The second step on the discovery of putative NUMT sequences in the genome of ALB was an iteration of the megaBLAST limited to the representative genomes with the whole mitochondrial reference genome and several partial sequences of NC_008221.1. This megaBLAST revealed also hits with very low E-values and higher % identities with scaffold1709 (NW_019417922.1) and scaffold27 (NW_019416292.1) on ALB's genome. The results of the BLAST of selected partial mitochondrial sequences can be looked up in Appendix, III Chapter B, **Table 30**, but more sections of NC_008221.1 have been blasted and are not shown here. The alignments from the megaBLAST of the partial scaffold sequences and the partial mitochondrial genome sequence were repeated in Geneious Prime® v. 2021.0.1, confirming the high similarities. For the ALB scaffolds 158 and 1709 (sc158, sc1709) at least five nuclear-mitochondrial-like regions were found (**Figure 39**, **Figure 41**). For three of them primers were designed to amplify them with their flanking regions via PCR for subsequent Sanger-sequencing (**Figure 40**, **Figure 43**, **Figure 44**, 4.4.4). The other putative NUMTs from scaffold 1709 and scaffold 1356 (sc1709, sc1356), were also summarized and depicted in **Figure 40** and **Figure 45**. These were not used for PCR amplification and therefore not tested if they are real NUMTs.

Figure 46 illustrates the whole mitochondrial genome NC_008221.1 map. The annotations point out, beside other annotations, the positions found with high similarities to the scaffold sections of sc158, sc1709 and sc1356 which were found via megaBLAST and were confirmed by Geneious alignment. These sections are often called "mitochondrial counterpart" in the following.

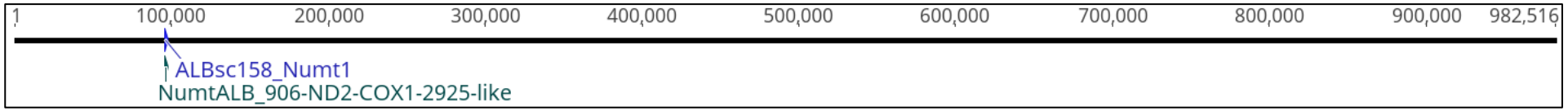


Figure 39. Position of ALBsc158-Numt1 or NumtALB_906-ND2-COX1-2925-like (= 1982 bp) on scaffold158 of ALB's genome (NW_019416265.1; 982516 bp) from (+) 96624-97605 (= 1982 bp).

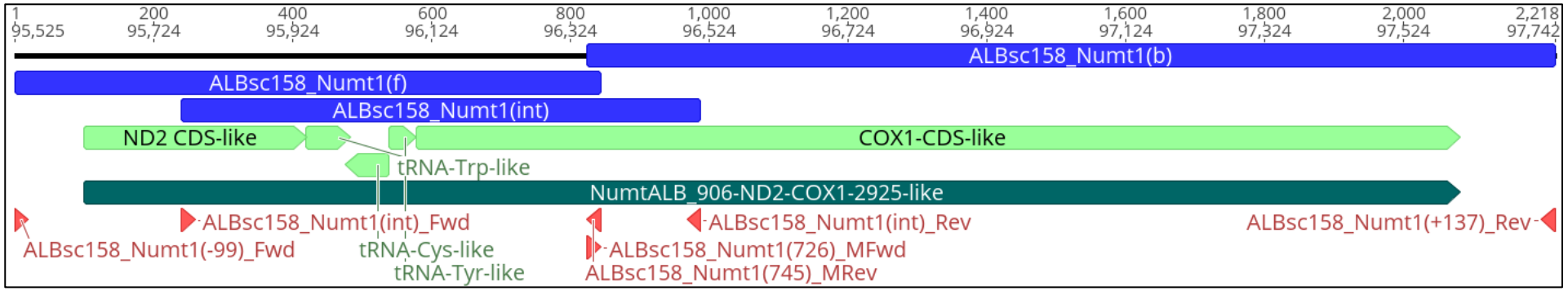


Figure 40. PCR products to obtain ALBsc158-Numt1 sequence, including the NumtALB_906-ND2-COX1-2925-like sequence as well as the flanking regions of ALBsc158.
The ALBsc158-Numt1 sequence is covered by three overlapping PCR products (royal blue). ALBsc158-Numt1(f) covering the flanking region and front end of the NUMT-1, ALBsc158-Numt1(int) covering the overlapping internal part in the NUMT-1 and ALBsc158-Numt1(b), which covers the back end of the NUMT-1 as well as the flanking region of scaffold158. In light green the mitochondrial counterparts are annotated here, while the pine marks annotate the whole NUMT-region. In red, the primers for the amplification of the PCR-products are pictured.

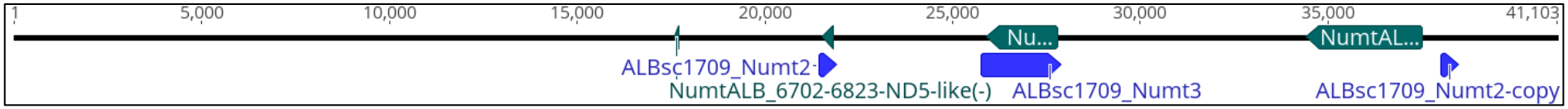


Figure 41. Position of four putative NUMTs (pine) on scaffold1709 on ALB's genome (NW_019417922.1; 41103 bp), from which NUMT-2 and NUMT-3 were sequenced.
ALBsc1709-Numt1: NumtALB_6702-6823-ND5-like(-), position 17594-17715 (= 122 bp) on ALBs' scaffold1709; This fragment was neither amplified via PCR nor Sanger-sequenced. ALBsc1709-Numt2 (or NUMT-2): NumtALB_3555-COX2-ATP8-3856-like(-), position 21519-21828(-) (= 310 bp) on ALBs' scaffold1709; ALBsc1709-Numt3 (or NUMT-3): NumtALB_374-ND2-COX1-2273-like(-), position 25910-27809(-) (= 1900 bp) on ALBs' scaffold1709. ALBsc1709-Numt4: NumtALB_15360-2555-DL-tRNA-ND2-COX1-mashup-like(-), position 34431-37525(-) (= 3095 bp) on ALBs' scaffold1709; This fragment was neither amplified via PCR nor Sanger sequenced. ALBsc1709-Numt2-copy: after PCR product amplification and sequencing of ALBsc1709-Numt2, BLAST of haplotypes of NUMT-2 (**Table 32**) revealed a copy of NUMT-2 on ALBsc1709 on position 38021-38489.

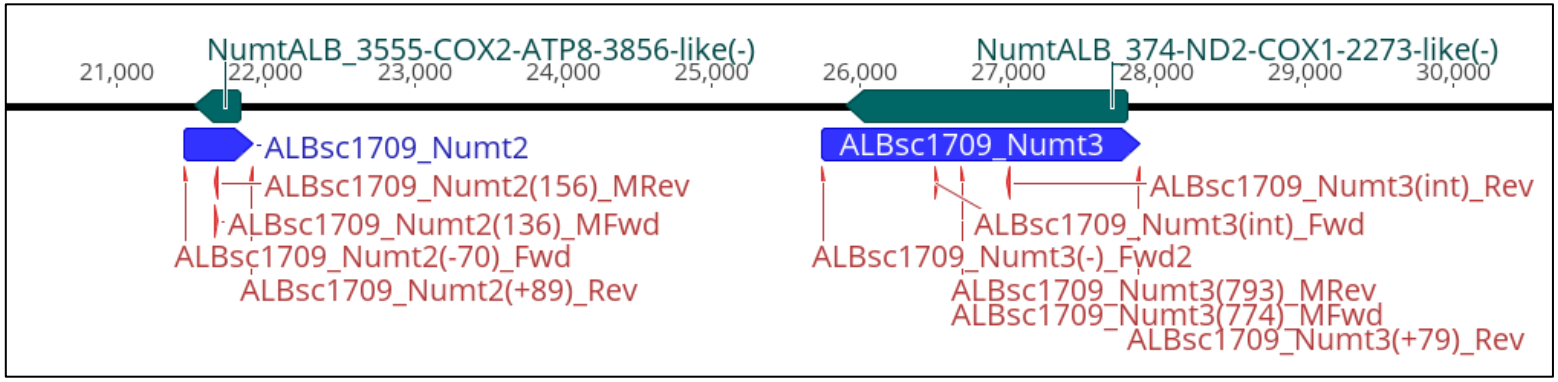


Figure 42. Position and distance between ALBsc1709-Numt2 and ALBsc1709-Numt3 on scaffold1709 of ALB's genome.
The distance between the two mitochondrial-like sequences of ALBsc1709-Numt1 (NUMT-2) and ALBsc1709-Numt3 (NUMT-3) is 4081 bp. The pine colour annotation stands for the NUMT-regions and the royal blue one for the amplified PCR-products.

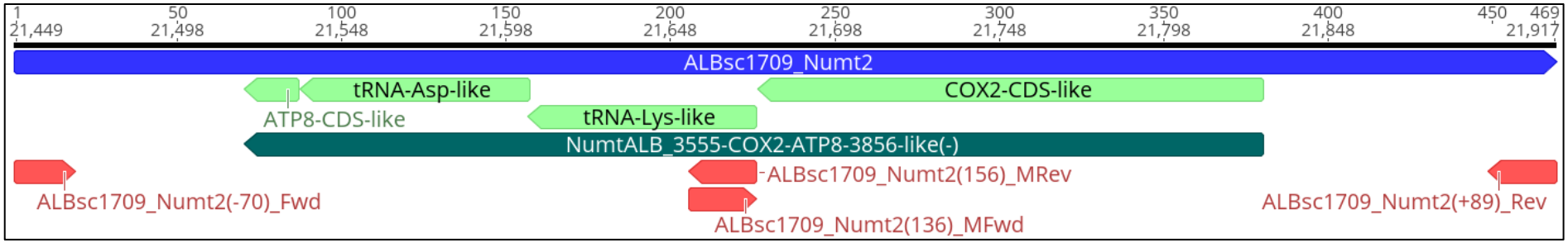


Figure 43. PCR products to obtain the ALBsc1709-Numt2 sequence, including NumtALB_355-COX2-ATP8-3856-like(-) sequence as well as the flanking regions of ALBsc1709.
The ALBsc1709-Numt2 (NUMT-2) sequence is covered by the complete PCR product (royal blue). The internal primers were used to confirm the existence of this NUMT region, but these PCR-products were not sequenced. The mitochondrial counterparts are annotated here in light green, while dark green marks the whole NUMT-region. In red, the primers for the amplification of the PCR-products are pictured.

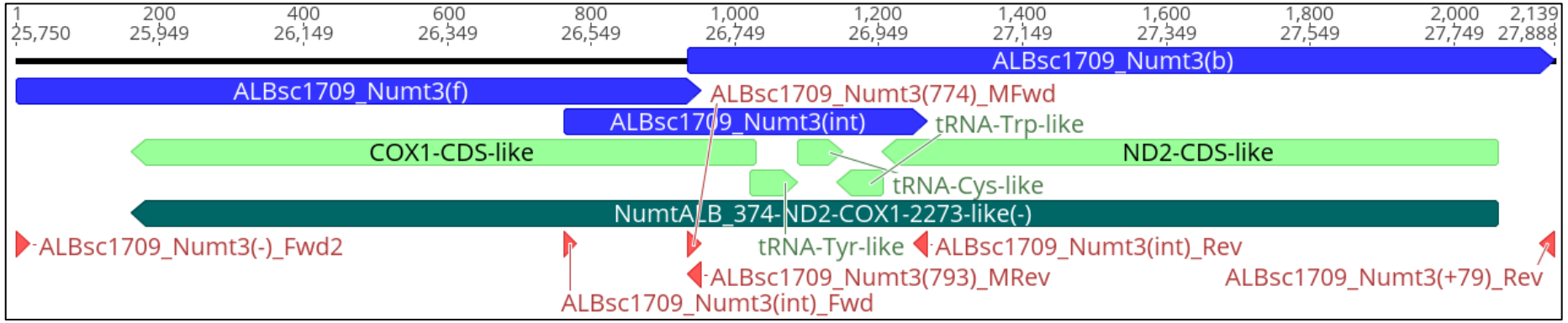


Figure 44. PCR products to obtain the ALBsc1709-Numt3 sequence, including NumtALB_374-ND2-COX1-2273-like(-) sequence and the flanking regions of ALBsc1709.
The ALBsc1709-Numt3 (NUMT-3) sequence is covered by three overlapping PCR products (royal blue). ALBsc1709-Numt3(f) covering the flanking region and front end of NUMT-3, ALBsc1709-Numt3(int) covering the overlapping internal part in NUMT-3 and ALBsc1709-Numt3(b), which covers the back end of NUMT-3 as well as the flanking region of ALBsc1709. The mitochondrial counterparts are annotated here in light green, while the dark green marks the whole NUMT-region. In red, the primers for the amplification of the PCR-products are pictured.

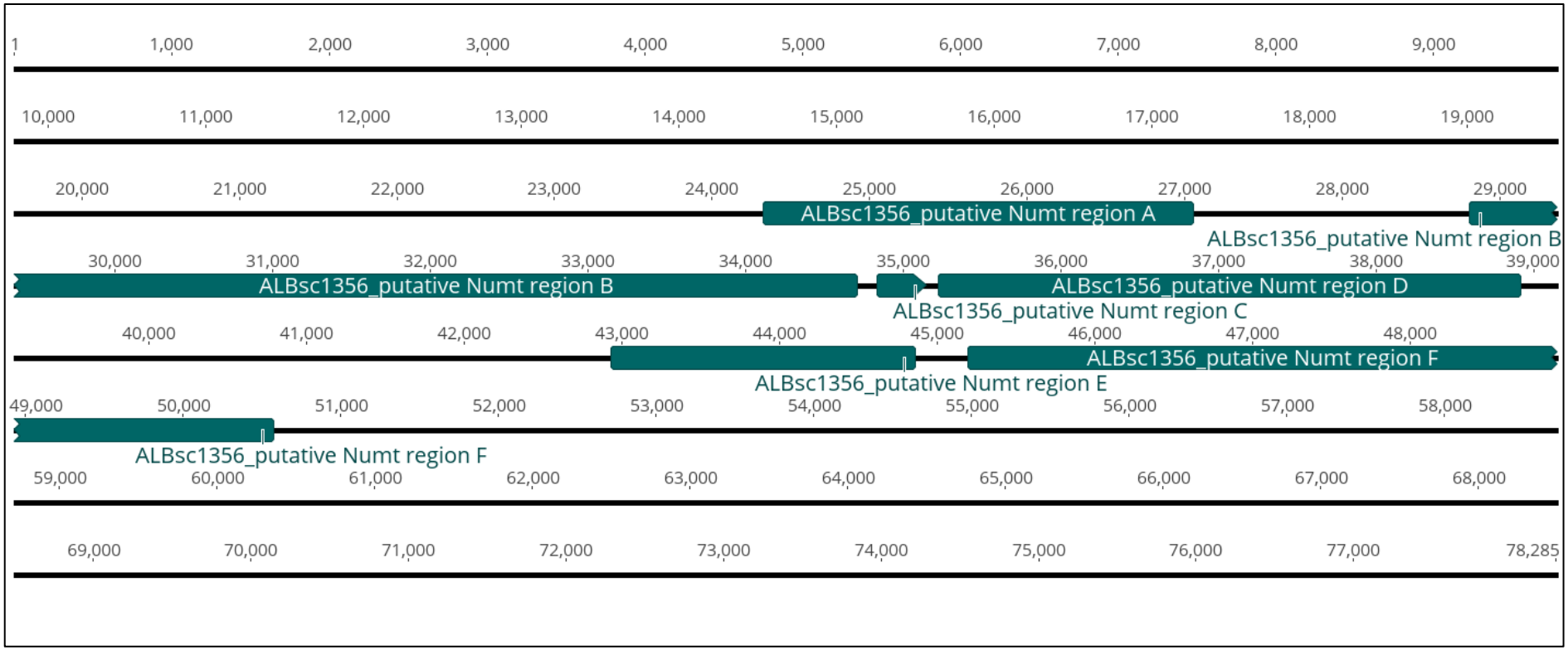


Figure 45. Position of six putative NUMT-regions on scaffold1356 of ALB's genome (NW_019417445.1; 78285 bp).
The putative NUMT regions (pine), from which none were sequenced, were discovered via megaBLAST of mitochondrial genome sequences like the other NUMTs. The regions were not as continuous as pictured here, the components were sometimes scattered, in different order and/or in different directions.

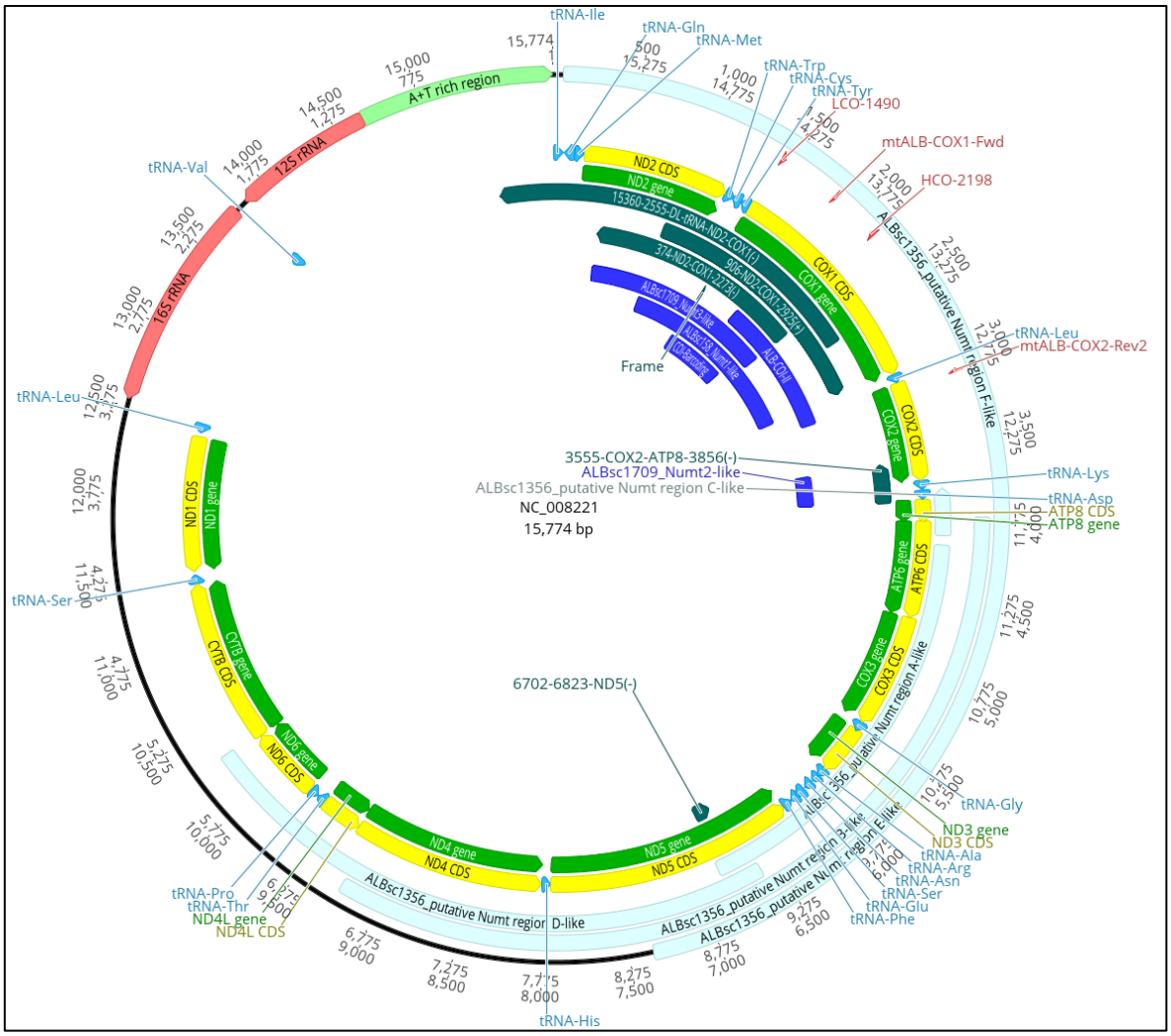


Figure 46. Map of ALB's mitochondrial genome (NC_008221; 15774 bp) and its features. In this map, all annotations from the natural features like genes (green), CDS (yellow), tRNAs (blue) etc. are pictured, as well as primer positions (red), PCR products (royal blue) and counterpart regions that aligned to NUMTs or putative NUMT regions in the nuclear genome of ALB (pine and light blue).
The royal blue annotations show the PCR products of COI-barcoding, ALB-COI-II and the mitochondrial counterparts of the NUMTs that have been Sanger-sequenced. The similar mitochondrial sequences of the NUMT-regions on sc158 and sc1709 of ALB's genome are annotated in dark green. ALBsc158-Numt1 has similarities to the mitochondrial counterpart from position 906-2925 including ND2, COX1 CDS sequence parts and t-RNAs in between. ALBsc1709-Numt1 has similarities with position 6702-6823 including partial ND5 gene of NC_008221; ALBsc1709-Numt2 shows similarities with position 3555-3856 including partial CDS of COX2 and ATP8 genes, as well as tRNAs in between of ALB's mitochondrial genome. ALBsc1709-Numt3 possesses similarities with position 374-2273 including partial ND2 and COX1, as well as tRNAs in between of NC_008221. The ALBsc1709-Numt4 had similarities with position 15360-2555 (including partial D-Loop, tRNAs and CDS of ND2 and COX1) of the mitochondrial ALB genome, but not in the same order ("mashup"). The annotations in light blue represent the resemblances with putative NUMT regions on sc1356 of ALB's genome.

5.3.3 Genetic diversity of NUMT sequences of ALB

22 specimens from Germany and Italy were used for PCR product amplifications, Sanger-sequencing and analyses of population genetic and evolutionary relationships on three different NUMT regions as described in 4.4.4. In this chapter either the location shortcuts (e.g. SB, HLD, KEH) or sample-ID numbers (e.g. 047, 013, 104) will be used, not the whole sample-IDs. The names of the three sequenced NUMTs from sc158 and sc1709 were shortened to NUMT-1, NUMT-2 and NUMT-3 in the following chapters.

Some of the sequences showed double peaks in their trace data, which was visually checked at the de novo assembly from direct PCR product sequencing. Consequently, this resulted in two consensus sequences for a specimen, which were named with _A or _B in the end. Not more than two sequences per specimen were tolerated. When considering the multiple abundance of similar NUMTs in ALB's genome, even two sequence variants with ambiguous nucleotides for one specimen were still tolerable when occurred by clone sequencing. However, sequence variants with ambiguities were only used in TCS networks and evolutionary trees, if no polymorphic sites were lost due to masking of ambiguous sites. Thus, sequences with too many ambiguous sites were also removed from the dataset and excluded from the analyses. For all the sequences, whereever it was possible, cloning of the PCR products was used to help with phasing of the sequence variants. Identical sequences of the NUMT-segments were defined by PopART v. 1.7 (Leigh and Bryant 2015) since it was not always possible to use DnaSP v. 6.12.01 due to indels and ambiguities. The PopART v. 1.7 software masks all ambiguous nucleotides and gaps. In the haplotype sequences, all unsure positions from the assigned specimens were replaced with N. Therefore, there were some ambiguous haplotypes used for the analyses, but these ambiguous sequences only occurred for the NUMT-fragment Numt1(f) and Numt3-complete sequence.

Of the Numt1(f) PCR-products, 17 sequences were generated, from which the sequences of three specimens included ambiguous base calls (013, 095, 174). Not all ambiguities could be cleared up by two non-ambiguous sequence variants per specimen, because four of the final sequences were also ambiguous (013_A, 013_B, 095_A, 174_B). Two of the 17 sequences were excluded, since they were only represented in a single clone sequence (095_B, 174_A). There were no sequences generated from the ALB specimens 074, 194, 085, 038, 129, 124, 120 and 181. Hence, 15 sequences could be used in the further analyses. The alignment of these sequences with the NUMT-1(f)-reference sequence had a length of 748 bp. A second alignment, which was used in the further alignments, additionally included the NC_008221-906-ND2-COX1-2925(+)-f2 partial mitochondrial ALB genome sequence. The NUMT-1(f2)-alignment had a length of 726 bp, since there were no flanking regions of NUMT-1 included here. Of the Numt1(int) PCR products, 19 sequences could be sequenced, from which the sample WAR-189 was the only one with two sequence variants. However, the sequence 189_A was removed from the dataset, since it was solely found in a single clone. The sequences of the NUMT-1(int)-reference and the sequence of 189_B were shorter than all the others. No sequences of NUMT-1(int) were obtained by the specimens 104, 038, 149, 180 and 181. The alignment of the remaining 18 sequences and the NUMT-1(int)-reference was 705 bp long. The second Numt1(int2)-alignment, which was used for subsequent analyses, included the NC_008221-906-ND2-COX1-2925(+)-int2 partial mitochondrial ALB genome sequence with a length of 706 bp. Of the Numt1(b) PCR products, 14 sequences could be generated, from which none had ambiguous base calls. There were no Numt1(b) sequences available for the specimens 189, 024, 194, 129, 180 and 181. Other sequences from the samples 013, 074 and 058 were sorted out, because the sequences were much shorter than the other ones. After NUMT-1(b)-alignment together with the NUMT-1(b)-reference, all sequences had a length of 1258 bp here. Also, for the back end of the NUMT-1 sequence a second alignment was conducted (NUMT-1(b2)-alignment) including the NC_008221- 906-ND2-COX1-2925(+) partial mitochondrial ALB genome sequence with a length of 1193 bp with no flanking regions of NUMT-1.

Figure 47 pictures the specimen count of represented Pop-IDs with their three Sanger-sequenced PCR products of ALBsc158-Numt1.

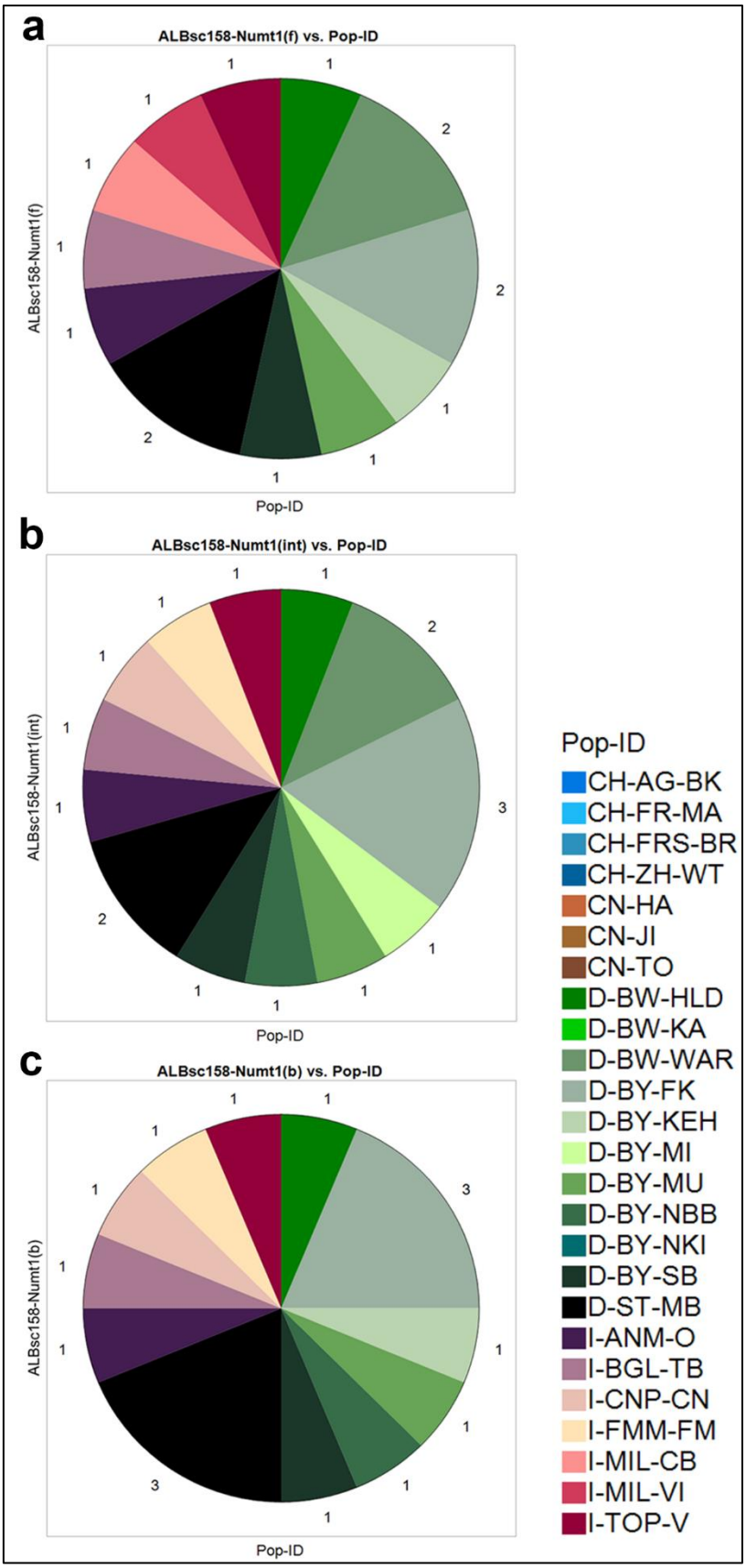


Figure 47. Represented ALB-populations (Pop-ID) by ALBsc158_Numt1 sequencing. **a)** ALBsc158-Numt1(f); **b)** ALBsc158-Numt1(int); **c)** ALBsc158-Numt1(b).

The results obtained by the estimates of evolutionary divergence from Mega X (Kumar et al. 2018) are depicted in **Table 35** (front end), **Table 36** (internal segment) and **Table 37** (back end) for all three PCR products of ALBsc158-Numt1 resulting from the alignments NUMT-1(f2), NUMT-1(int2) and NUMT-1(b2) in Appendix, III Chapter B. An overview of the mean % divergences is illustrated in **Figure 48**. Because there were length differences and ambiguous nucleotides found in the data, all positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). The number of base substitutions per site from between sequences are shown converted into percentage. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992), since model testing of all three alignments measured the lowest BIC for this model. In all three tables, not the whole matrix of the divergence estimates is shown. The pairwise evolutionary divergence as % of substitutions per site are shown for the NUMT-1-reference sequences of partial ALBsc158 for Numt1(f2), Numt1(int2) and Numt1(b2), as well as the respective mitochondrial counterparts NC_008221-906-ND2-COX1-2925(+) for f2, int2 and b2 compared to the respective Sanger-sequenced sample sequences. The 15 ALBsc158-Numt1(f2) sequences generated in this study and the NUMT-1(f2)-reference showed a genetic divergence of ~6% to the mitochondrial sequence, whereas the NUMT-1(f2)-reference had explicitly lower values with divergences of ~0-0.6% to the 15 sequences. The 18 ALBsc158_Numt1(int2) sequences sequenced in this study showed contradicting results compared to the front end. Here, the genetic divergence between the mitochondrial NC_008221-906-ND2-COX1-2925(+)-int2 sequence to the 17 sample-sequences was comparatively lower with ~0.3-0.7%, while the divergence to WAR-189_B and NUMT-1(int2)-reference was ~5%. The NUMT-1(int2)-reference showed higher genetic differences of ~4-5% to all other sequences, except to the sequence of WAR-189_B were ~0.3% divergence was measured. The 14 ALBsc158-Numt1-(b2) sequences showed similar results compared to the front end. Again, the mitochondrial NC_008221-906-ND2-COX1-2925(+)-b2 showed high genetic divergence of ~5-6% to all the other NUMT-1(b2) sequences, while the genetic divergence of the NUMT-1(b2) sequences to their nuclear NUMT-1(b2)-reference was in a range of ~0.2-1.2% genetic differences. The higher genetic differences of ~1.2 to the NUMT-1(b2)-reference sequence was only observed in the sequences of O-170 and FM-120, while the others were in the range of ~0.2-0.3% divergence.

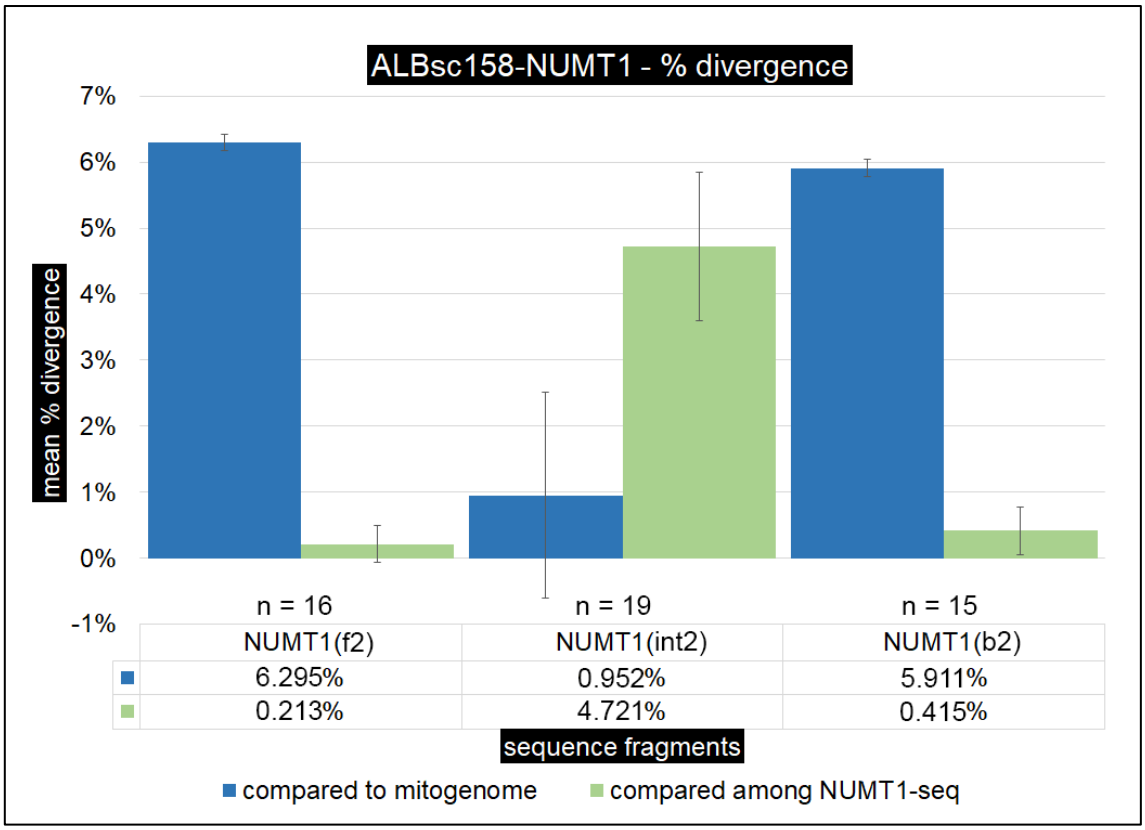


Figure 48. Estimates of evolutionary divergence in % between ALBsc158_Numt1 sequences and their respective mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018).
The mean % divergences of 16 sequences of ALBsc158_Numt1(f) with 660 positions, 19 sequences of ALBsc158_Numt1-(int) with 666 positions, 15 sequences of ALBsc158_Numt1-(b) with 1193 positions to the respective mitochondrial counterparts (blue), as well as among the different haplotypes (green) are shown.

In **Figure 49** the TCS haplotype network of all haplotypes of the ALBsc158-Numt1(f2), ALBsc158-Numt1(int2) and ALBsc158-Numt1(b2) sequences are pictured in comparison to the NUMT-1-reference and the mitochondrial counterpart mtALB_ND2-COX1. Consequently, this analysis was based on the shortened alignments, which included only the NUMT-sections and no flanking regions. Haplotypes were formed by PopART v. 1.7 as described before (p. 115). For the front-end part of NUMT-1, the haplotype sequences included some ambiguous nucleotides (N), since the sequences 013_A, 013_B, 095_A and 174_B for Numt1(f) were ambiguous and were assigned each to a different haplotype. For the internal and the back-end parts of NUMT-1 sequence, all haplotype sequences were non-ambiguous.

For the front end of ALBsc158-Numt1, no TCS network including all the sequenced haplotypes was possible, due to the length difference of 36 bp. The alignment with mtALB_f2_ND2-COX1 shows that Numt1(f2)_Type3 and Type4 are more conform with the length of the mitochondrial sequence. Hence, one network was drawn with the NUMT-1(f2) reference and the haplotypes Numt1(f2)_Type1 and Type2, while another one was drawn with haplotypes Numt1(f2)_Type3

and Type4 with the mitochondrial counterpart. Numt1(f2)_Type2 is one mutation different to both, Type1 and the NUMT-1(f2)-reference, while Numt1(f2)_Type1 had two mutations to the NUMT-1(f2)-reference. The other TCS network drawn with haplotypes Numt1(f2)_Type3 and Type4, together with the mitochondrial counterpart, showed 39 mutations to the node between haplotype 3 and 4. Numt1(f2)_Type4 showed two mutations to the node and Numt1(f2)_Type3 one mutation to the node. Numt1(f2)_Type3 solitary occurred in Italy and the haplotypes 4 and 2 only in Germany.

For the internal DNA-segment Numt1(int2), most of the haplotypes were only a few mutations apart from the mitochondrial counterpart, while there were 32 mutations from the most common Numt1(int2)_Type1 to the NUMT-1(int2)-reference sequence. Numt1(int2)_Type5 from Weil am Rhein in Germany is the only haplotype with two mutations apart from the NUMT-1(int2)-reference.

The haplotypes of the back-end of NUMT-1 are most genetically distant to the mitochondrial counterpart with 64 mutations difference to the NUMT-1(b2)-reference. The Numt1(b2)_Type3 from Ostra and Fermo in Italy is 12 mutations apart from the most sequenced Numt1(b2)_Type1 in this dataset. Numt1(b2)_Type1 is only one mutation apart from Numt1(b2)_Type2 and three mutations distant from the NUMT-1(b2)-reference.

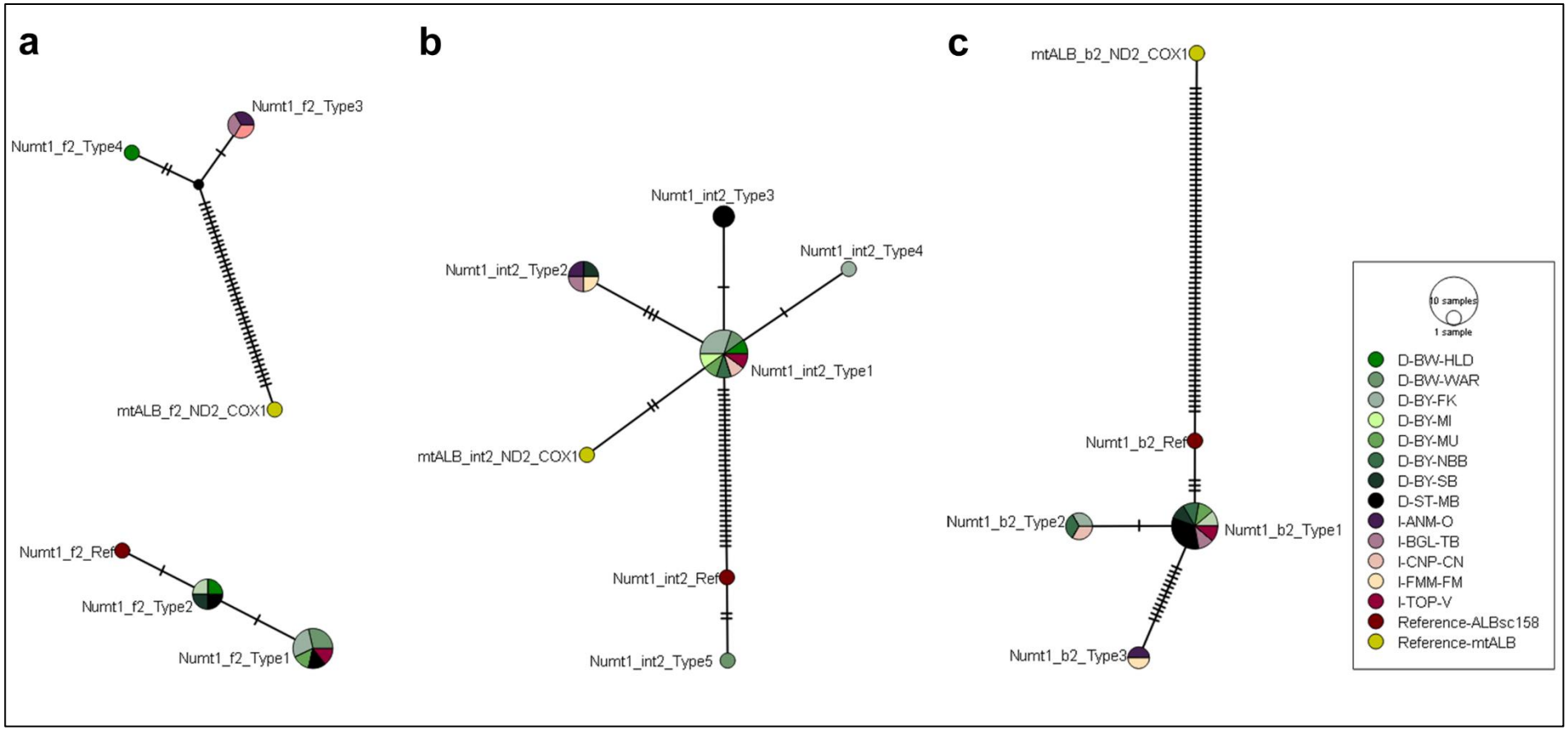


Figure 49. TCS haplotype network of **a)** ALBsc158-Numt1(f2), **b)** ALBsc158-Numt1(int2) and **c)** ALBsc158-Numt1(b2) sequences from Germany and Italy in comparison to the NUMT-1-reference and the mitochondrial counterpart mtALB_ND2-COX1 (PopART v. 1.7).
Mutations are represented by hatch marks; the size of the circles is proportional to the number of specimens contributing to the haplotype.

In Appendix, III Chapter B, **Table 31**, the first three hits of the result from megaBLAST of all the haplotypes of ALBsc158-Numt1, limited to representative genomes of ALB, are shown in **Table 34** of the reference sequences against the nucleotide collection. All hits for ALBsc158-Numt1 haplotypes had an E-value of ~0.0. The NUMT-reference, Numt1(f2), Numt1(f2)_Type1 and 2 showed additional to sc158, comparable similarities with sc1709 (~95%) and the mitochondrial genome of ALB (94%). The haplotypes Numt1(f2)_Type3 and 4 were different and had higher conformities with ~93% identity to sc1709, sc1356 and the mitochondrial genome of ALB than to sc158. The results of Numt1(int2)_Type5 were similar to NUMT-1-(int2)-reference with ~99% identity to sc158 and only ~93% identity to NC_008221 and sc1709. Otherwise, the haplotypes 1-4 from Numt1(int2) had >99% identity with the mitochondrial reference genome of ALB and lower identity to sc1709 and sc1356. The back-end of NUMT-1 showed with the haplotypes Numt1(b2)_Type1-4 the highest % identity (~98-99%) to sc158.

The evolutionary history of the NUMT-1 haplotypes is depicted in three maximum likelihood trees shown in **Figure 50** on the alignments of all three PCR-products for ALBsc158-Numt1 with their respective mitochondrial counterparts and NUMT-1-reference. The results are quite similar to the results from the TCS networks, but show the distance in more detail. For instance, despite the length difference, all haplotypes of Numt1(f) could be illustrated here. Likewise, the haplotype network, Numt1(f2)_Type1 with seven sequences (MU, MB, FKR, FK, WAR, V) and the NUMT-1-reference are not very distant from Type2 with four sequences (SB, MB, HLD, KEH), but clearly separated. Different than suggested in the network, the NUMT-1-reference is least distant to Type1. Clearly separated are Numt(f2)_Type3 (O, TB, CB) and the single Type4 from HLD. Undoubtedly, the mitochondrial counterpart forms an outgroup in this tree. For the internal Numt1(int2) sequence, Type1 with ten sequences (FK, MI, MU, NBB, CN, V, HLD, WAR) and the slightly apart two sequences of Type2 (MB), cluster together with the mitochondrial counterpart. The Numt1(int2)_Type2 with four sequences (SB, O, TB, FM) forms its own group, while Type5 (WAR) and the NUMT-1-reference form together an outgroup. Lastly, the haplotypes of Numt1(b2) show noticeably the three groups of Type1 with eight sequences (MB, TB, SB, NBB, MU, KEH, V) together with the NUMT-1-reference, Type2 with three sequences (FKR, NBB, CN) directly next to Type1 and obviously distant Type3 with two sequences (O, FM). The mitochondrial counterpart forms again the outgroup of this evolutionary tree.

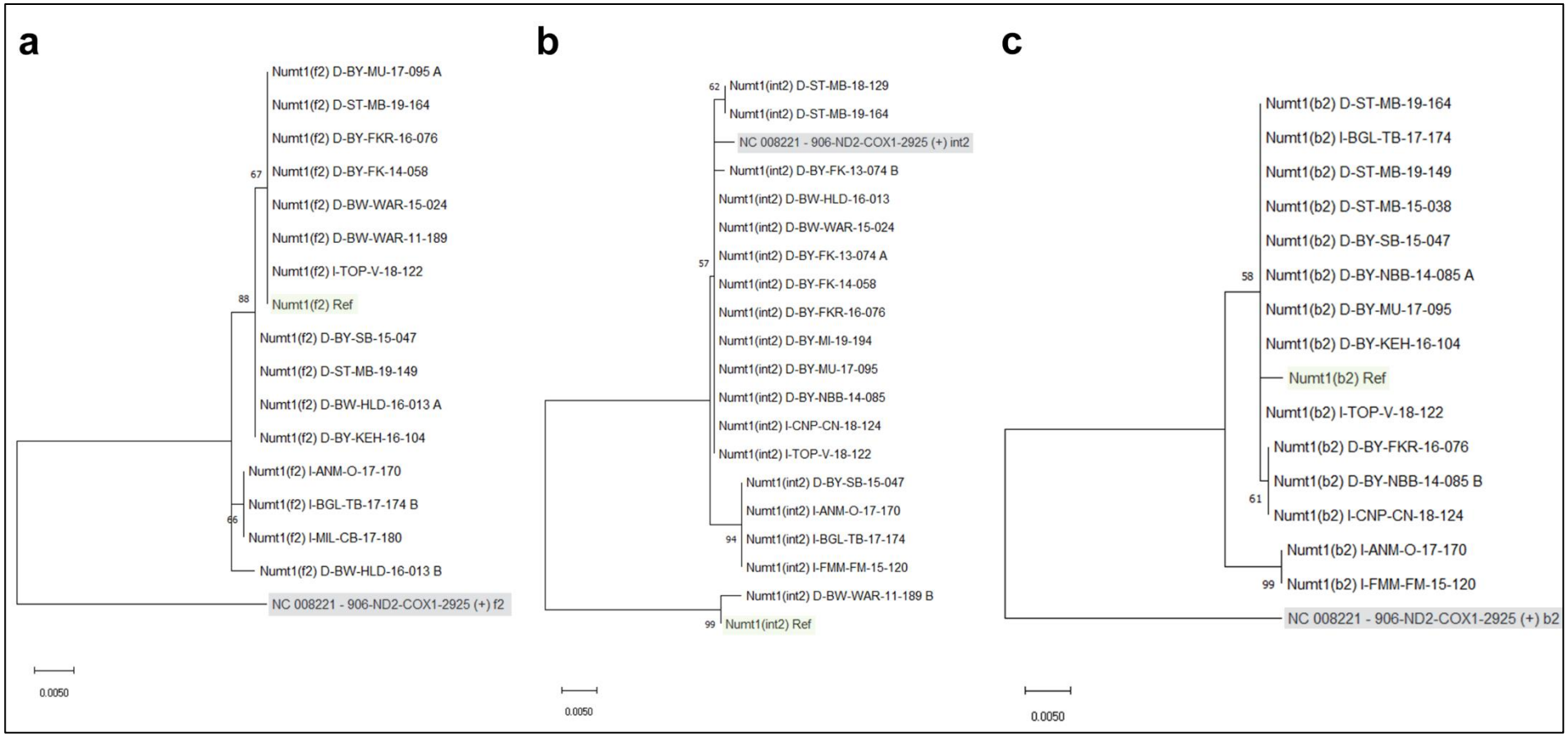


Figure 50. Evolutionary history inferred by using the Maximum Likelihood method on the alignments of all three PCR-products for ALBsc158-Numt1 with their respective mitochondrial counterparts (grey) and NUMT-reference (green) in MEGA X (Kumar et al. 2018).
The Tamura 3-parameter model (Tamura 1992) was used for the pairwise distance estimation to draw the tree, and the bootstrap method was used to test the results. The branch lengths measure in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). **a)** ALBsc158-Numt1(f2) sequence: This analysis involved 17 nucleotide sequences and a total of 660 positions in the final dataset. **b)** ALBsc158-Numt1(int2) sequence: This analysis involved 20 nucleotide sequences and a total of 666 positions in the final dataset. **c)** ALBsc158-Numt1(b2) sequence: This analysis involved 16 nucleotide sequences and a total of 1193 positions in the final dataset.

For the ALBsc1709-Numt2 DNA-segment, the PCR products from the front-end, the back-end of the NUMT and its flanking regions were used to proof that the NUMT is not just an artefact. This was validated for all 22 samples for the front- and the back-end of NUMT-2 without subsequent Sanger-sequencing.

For all 22 samples, the PCR products of the whole ALBsc1709-Numt2 were directly Sanger-sequenced. Some direct PCR product sequencing resulted in a homozygous sequence (eight from Germany and seven from Italy), while some resulted in a heterozygous sequence exhibiting at least one or more obvious ambiguous peaks (seven from Germany). A new megaBLAST revealed another very similar potential NUMT on unplaced genomic scaffold164 of ALB (NW_019416480.1) which is maybe co-amplified. Hence, all 22 PCR products were cloned and sequenced. At least 3-5 clones were generated and sequenced. From all 132 sequences of ALBsc1709-Numt2, there were three haplotypes observable, originating from variants of the ambiguous sites W (position 142), R (position 200) and S (position 312). The three observed combinations were named A, B and C, according to the frequency in the dataset. The cloning of the PCR-products allowed phasing of these haplotypes to the phased haplotypes A (TGG), B (AAC) and C (TAC) on the positions 142, 200 and 312. The reference sequence is representing the haplotype A for ALBsc1709-Numt2 and possesses another SNP on position 149 not represented in any of the sequenced samples. On this position, NUMT-2-reference has G, while the other haplotypes have an A. From the heterozygous specimens, the combinations AB and AC occurred equally often three times, while BC only occurred once. The alignment of ALBsc1709-Numt2 haplotypes A, B and C together with the NUMT-2-reference measured a length of 429 bp, while NUMT2-alignment-2 included the mitochondrial sequence of 3555-COX2-ATP8-3856(-) without any flanking regions (311 bp).

In **Figure 51** the proportion of sequenced specimens for each Pop-IDs for the complete ALBSc1709-Numt2 sequence is illustrated.

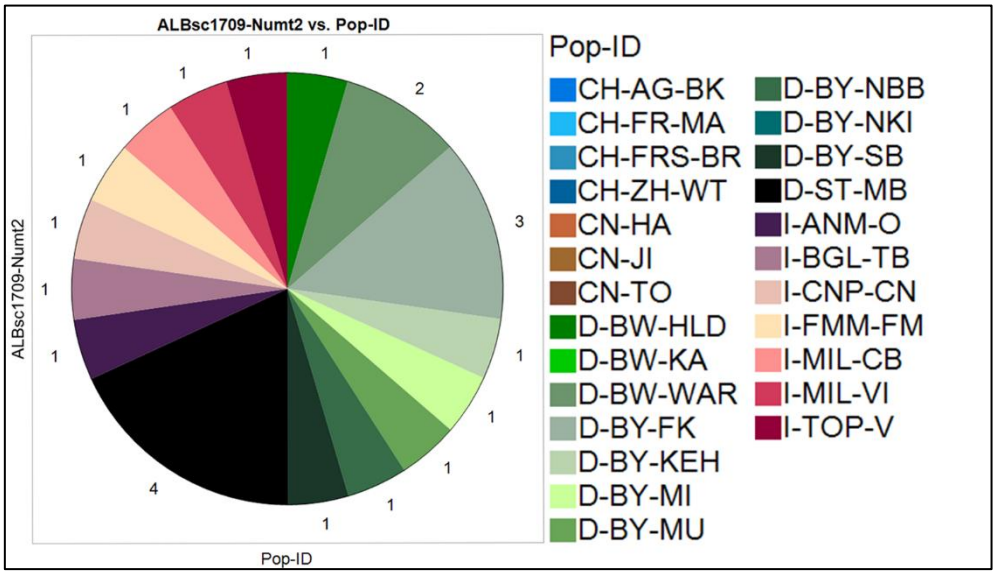


Figure 51. Represented ALB-populations (Pop-ID) from ALBsc1709-Numt2 sequencing.

The results obtained by the estimates of evolutionary divergence from Mega X (Kumar et al. 2018) are depicted in **Table 38** in Appendix, III Chapter B and in **Figure 52** as mean % divergence. This analysis was performed on an alignment of all 22 specimens and their phased sequences, hence the heterozygous specimens were represented twice here. The model testing calculated the lowest BIC for the Tamura-3-parameter model. Not the whole matrix of the divergence estimates is shown here. The pairwise evolutionary divergence as % of substitutions per site are shown for the NUMT-2-reference sequence of partial sc1709 and the respective mitochondrial counterpart NC_008221-3555-COX2-ATP8-3856(-) to the 29 sample sequences. The divergence of all 29 sequenced sequences and ALBsc1709-Numt2 reference was with ~10% quite high to the mitochondrial counterpart. The divergence of the NUMT-2-reference to the 29 sequences of the 22 specimens were ~0.3% for the A haplotype, ~0.9% for the C haplotype and highest genetic difference with ~1.3% between haplotype B and the NUMT-2-reference.

The megaBLAST results of the three haplotypes and the NUMT-2-reference limited to representative genomes of ALB are shown in Appendix, III Chapter B, **Table 32** and in **Table 34** of the reference sequence against the nucleotide collection. All four haplotypes had only 87-88% identity with the mitochondrial genome section 3555-COX2-ATP8-3856(-), but still a significantly low E-value. The first two hits were for all sequences sc164 and sc1709 of ALB's genome with E-values of 0.0. The NUMT-2 haplotype A, B, C and NUMT-2-reference had similar high similarities to sc164 and sc1709 (~99-100%).

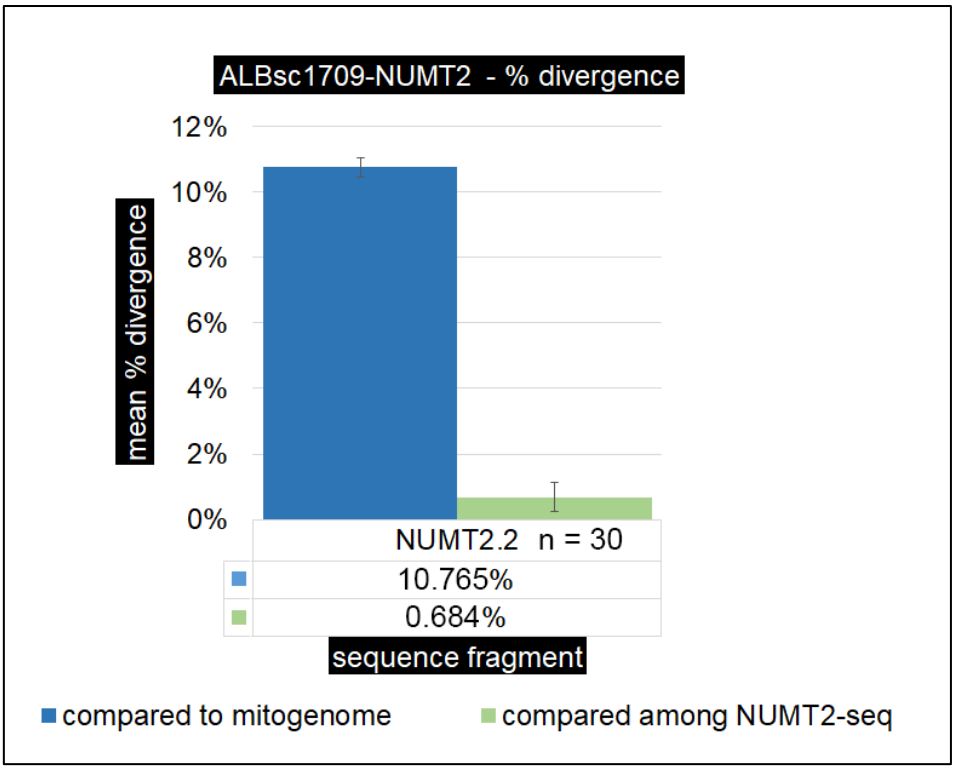


Figure 52. Estimates of evolutionary divergence in % between ALBsc1709_Numt2 sequences and their respective mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018).
The mean % divergences of 30 sequences of ALBsc1709_Numt2 with 310 positions to the respective mitochondrial counterparts (blue), as well as among the different haplotypes (green) are shown.

In the TCS haplotype network (**Figure 53**) it can be observed that the haplotype of the NUMT-2-reference is next to the NUMT-2 haplotype A2 with a divergence of one mutation only. The rarest haplotype in this dataset C2 has two mutations from A2 and is with 28 mutations the closest haplotype to the mitochondrial counterpart. The haplotype B2 is one mutation different from C2, and three mutations different from A2.

As also visualized in the haplotype projection (**Figure 54**), the NUMT-2 haplotype A was found in most sequenced locations, except for Hildrizhausen in Baden-Wuerttemberg, Miesbach, Schoenebach (both Bavaria, Germany) and Fermo (Marche, Italy). The haplotype B occurred with exception of Fermo solely in Germany in Magdeburg, Schoenebach, Neubiberg, Miesbach, Weil am Rhein and Hildrizhausen. The rarest haplotype C was only observed in Schoenebach, Murnau and Magdeburg in Germany.

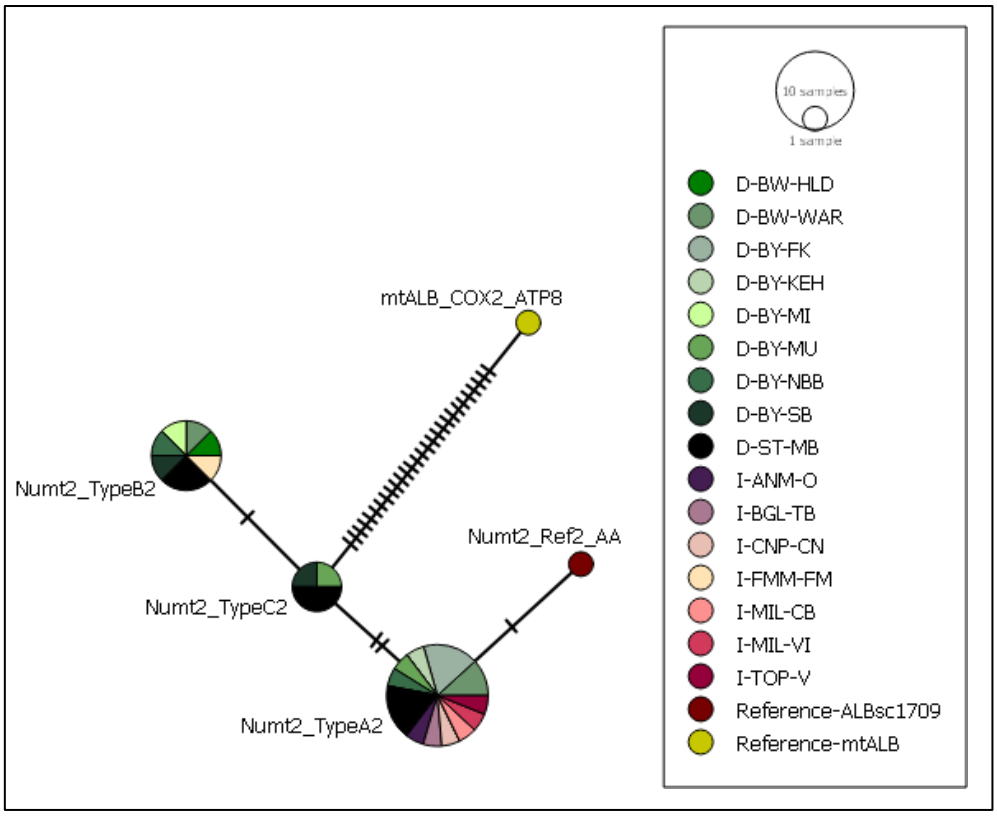


Figure 53. TCS haplotype network of haplotypes of ALBsc1709-Numt2 without flanking regions from Germany and Italy in comparison to the NUMT-2-reference and the mitochondrial counterpart mtALB_COX2-ATP8 (PopART v. 1.7).
Mutations are represented by hatch marks; the size of the circles is proportional to the number of specimens contributing to the haplotype.

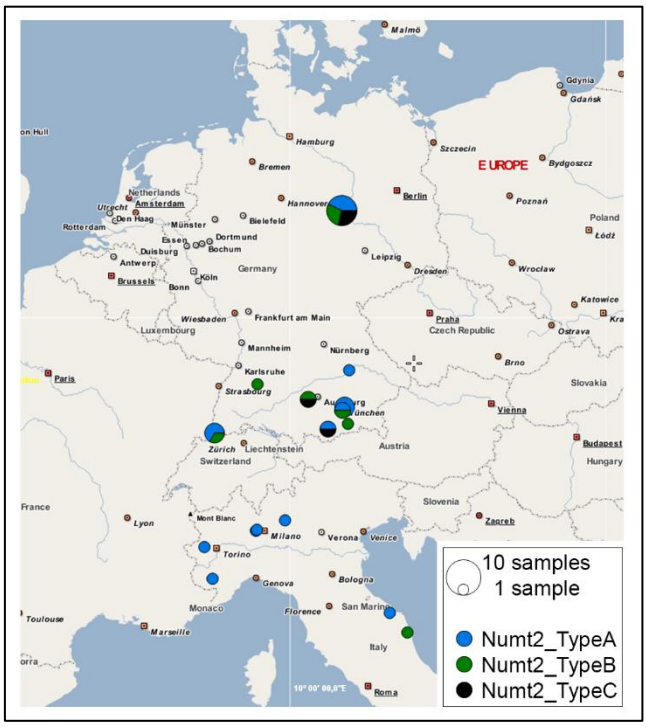


Figure 54. Haplotype projections of 3 ALBSc1709_Numt2 haplotypes in Germany and Italy (PopART v. 1.7).

The evolutionary history of the NUMT-2 sequences of ALBsc1709-Numt2 is depicted in a maximum likelihood tree (**Figure 55**) on the alignments of the sequences for ALBsc1709-Numt2 with their respective mitochondrial counterpart and NUMT-2-reference. The results reflect exactly the results of the TCS networks, but show the relationships between the haplotypes in more details. Here it is more explicitly demonstrated, that haplotype C is the oldest haplotype for NUMT-2 in this dataset, and that haplotypes A, NUMT-2-reference and B emerged from C. Again, the mitochondrial counterpart 3555-COX2-ATP8-3856(-) forms the outgroup with the next branch ending in haplotype C. The distribution of the haplotypes among the sequenced specimen is as described before, while AA and BB reflects the homozygous specimen and A, B or C the single sequences of the heterozygous ones.

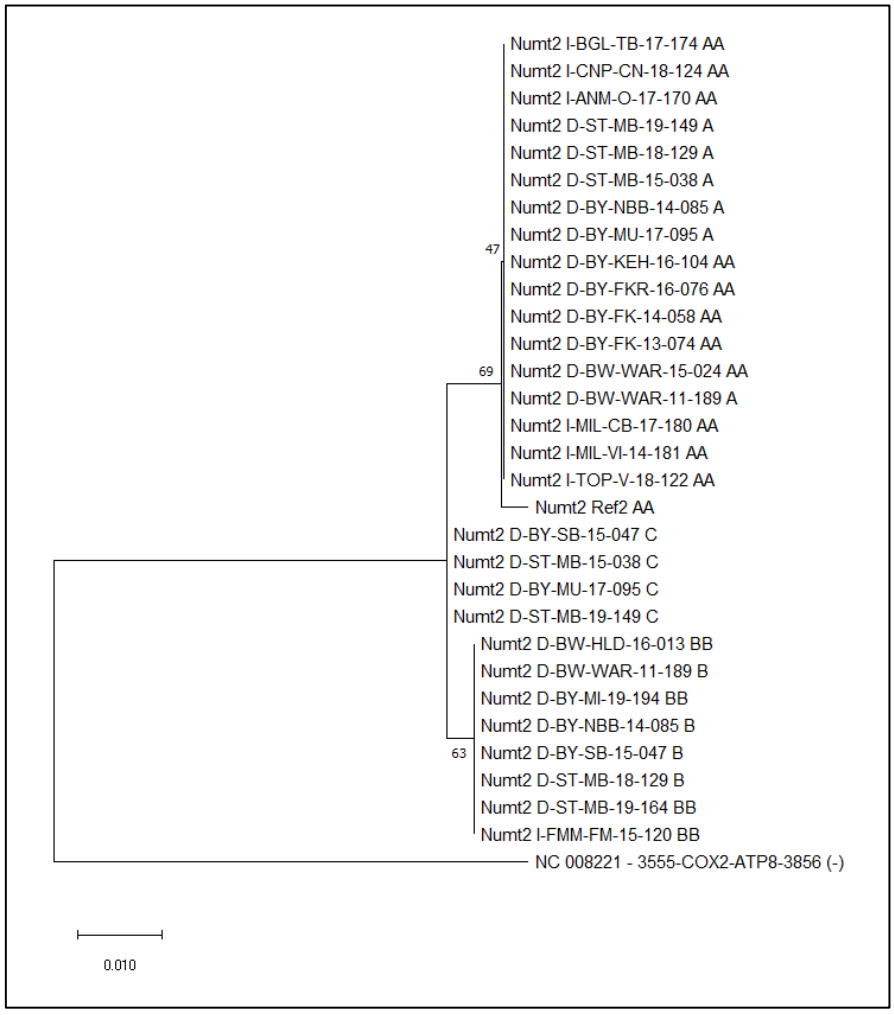


Figure 55. Evolutionary history inferred by using the Maximum Likelihood method on the alignment of ALBsc1709-Numt2 with the respective mitochondrial counterpart and the NUMT-2-reference in MEGA X (Kumar et al. 2018).
The Tamura 3-parameter model (Tamura 1992) was used for the pairwise distance estimation to draw the tree, and the bootstrap method to test the results. The branch lengths measure in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). This analysis involved 31 nucleotide sequences and a total of 310 positions in the final dataset.

Of ALBsc1709-Numt3, also three PCR products were sequenced for the front-end, the internal part and the back-end. For 18 of the specimens it was also possible to form a consensus sequence out of their respective sequences Numt3(f), Numt3(int) and Numt3(b). The alignment of the complete sequence was 1945 bp long. When aligned with the mitochondrial sequence of 374-ND2-COX1-2273(-), the length of the sequences was cut to 1810 bp since there were no flanking regions of NUMT-3 included here. The NUMT-3-complete DNA-sequences had some samples (WAR-024, FK-058, KEH-104, TB-174, CN-124, FM-120) with two sequence variants (_A and _B). The sequences KEH-104_A and _B of NUMT-3-complete had on the same position a Y (Pyrimidine, T or C) due to an overlap conflict between Numt3(int) and Numt3(b). But since this did not cover another variable site, the sequences could be included. There were no sequences representing the complete NUMT-3 for WAR-189, MB-149, CB-180 and VI-181. Of all ALBsc1709-Numt3 front-end sequences, only WAR-189 sample was not represented. The alignment of the 28 available Numt3(f) sequences together with the NUMT-3(f)-reference for Numt3(f) had a length of 914 bp. The alignment Numt3(f2) together with NC_008221-374-ND2-COX1-2273(-) partial mitochondrial ALB genome sequence without flanking regions of NUMT-3 had a length of 774 bp. The heterozygous specimens with two sequence variants of Numt3(f) were WAR-024, FK-058, KEH-104, TB-174, CN-124, FM-120 and VI-181. Of all ALBsc1709-Numt3 internal sequences only for CB-180 there was no sequence obtained. All the other 21 samples were sequenced for ALBsc1709-Numt3(int) with no heterozygous specimens for this fragment, except for KEH-104 and VI-181. The Numt3(int) sequence of VI-181 showed two ambiguous nucleotides, one of them masking a variable site from the sequences of FK-074 and O-170. Hence, VI-181 was excluded in the haplotype network, but not in the evolutionary analyses. The Numt3(int)-alignment including 21 sequences and the NUMT-3(int)-reference had a length of 469 bp. The second alignment Numt3(int2) together with NC_008221-374-ND2-COX1-2273(-) partial mitochondrial ALB genome sequence had a length of 473 bp. Of all ALBsc1709-Numt3 back-end sequences there were no sequences obtained for the samples VI-181, MB-149 and WAR-189. Two sequence variants for Numt3(b) were obtained from WAR-024, FK-058, KEH-104, TB-174, CN-124 and FM-120. NUMT-3(b)-alignment of the 21 sample sequences and the NUMT-3(b)-reference had a length of 980 bp here. The second variant, NUMT-3(b2)-alignment, was aligned additionally to NC_008221- 374-ND2-COX1-2273(-) partial mitochondrial ALB genome sequence with a length of 984 bp without the downstream flanking region of the NUMT-3.

The specimens count of represented Pop-IDs for the three PCR products of ALBsc1709-Numt3 are shown in **Figure 56**. These sequences together represent NUMT-3 and its flanking regions on sc1709.

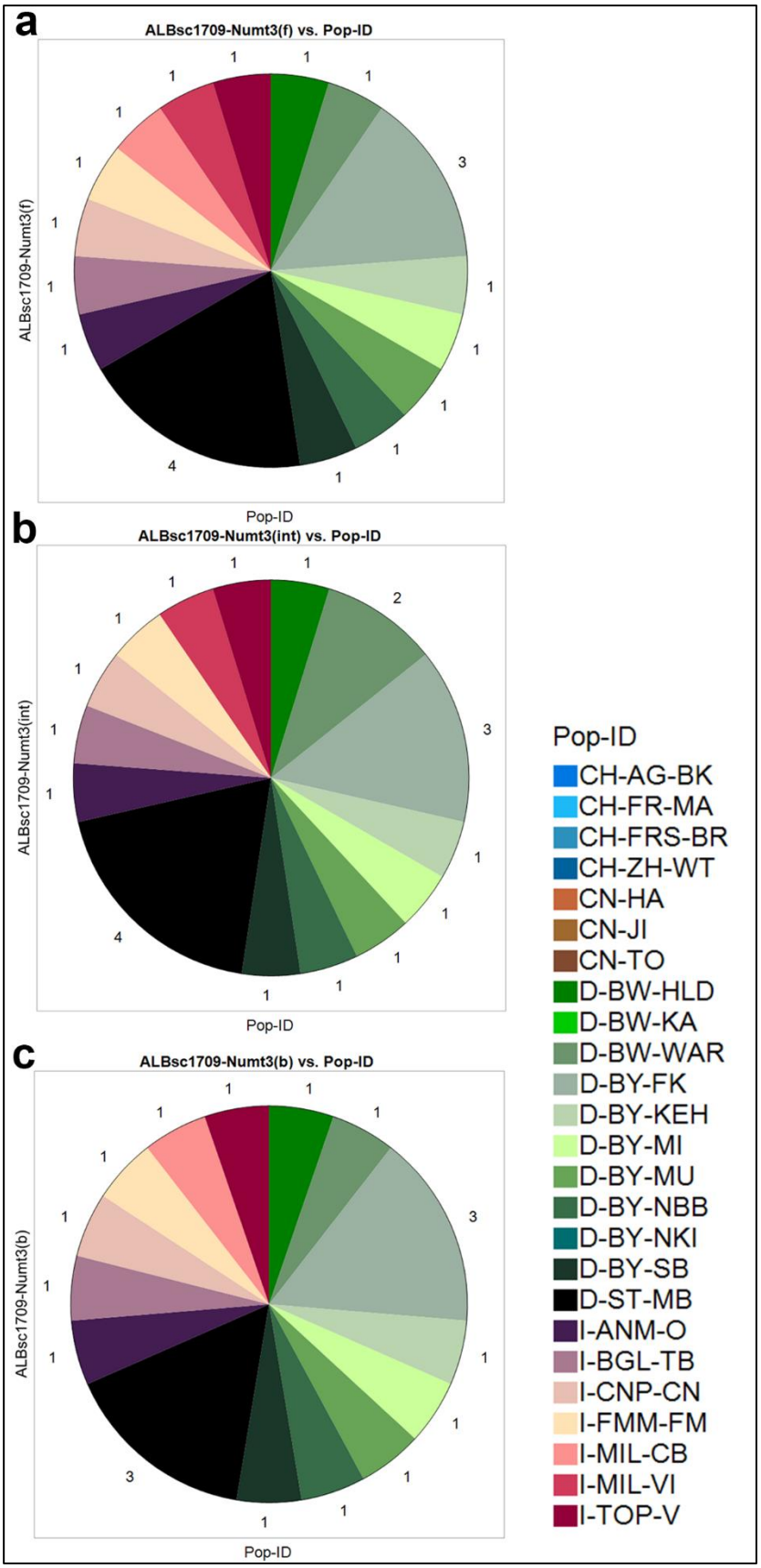


Figure 56. Represented ALB-populations of ALBsc1709-Numt3 sequencing.
a) ALBsc1709_Numt3(f); **b)** ALBsc1709_Numt3(int); **c)** ALBsc1709_Numt3(b).

Just like for the other two NUMTs before, the evolutionary divergence was estimated by Mega X (Kumar et al. 2018) for the front-end (**Table 39**), the internal segment (**Table 40**) and the back-end (**Table 41**) of the ALBsc1709-Numt3 DNA-sequence, without the flanking regions, see in Appendix, III Chapter B. An overview of the mean % divergences is illustrated in **Figure 57**. Again, the partial deletion option was used. The number of base substitutions per site between sequences are shown here converted into percentage. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992), since model testing of Numt3(f) and Numt3(int) resulted in the lowest BIC for this model. For Numt3(b) the lowest BIC was estimated for the Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al. 1985), but in this case Tamura-3-parameter was used as well. In all three tables instead of the whole matrix of the divergence estimates, the pairwise evolutionary divergence as % of substitutions per site for the NUMT-3-reference sequences and their respective mitochondrial counterparts to all sequences of the European specimens is shown. For the front-end sequences of ALBsc1703-Numt3 (774 bp), the divergence of the partial 374-ND2-COX1-2273(-) reference was ~7.5-8% divergent to all 28 sequences of the ALB samples from this study and the NUMT-3(f2)-reference. The NUMT-3(f2)-reference had on the other hand pairwise % of substitutions per cite from ~0.9-1.3% to the 28 ALB NUMT sequences. The highest divergence of ~1.3% to the NUMT3(f2)-reference was found in Corbetta and Vittuone. The 22 sequences of ALBsc1079-Numt3(int2) obtained in this study were less divergent than the front end with ~5.4-5.9% to the mitochondrial reference sequence. Also, the NUMT-3(int2)-reference was slightly less divergent to the 22 sequences of the ALB samples with ~0.6-0.8%. The highest divergence to the NUMT-3(int2)-reference with ~0.8% was detected in FK-074, KEH-104_B and O-170. The back-end ALBsc1709(b2) sequences (980 bp) had similar but slightly higher divergence values like the front-end, with ~8.1-8.4% divergence of the mitochondrial reference to the 25 NUMT sequences of the ALB specimens in Germany and Italy. The % divergence of the NUMT-3(b2)-reference to the 25 sequences were also similar, but slightly lower than in the front-end with ~0.9-1%.

For all haplotypes of the three PCR products as well as the consensus complete ALBsc1709-Numt3 sequences, the results of the megaBLAST against the whole representative genomes of *A. glabripennis* are demonstrated in Appendix, III Chapter B, **Table 33** and in **Table 34** megaBLAST of the reference sequences against the nucleotide collection. All first three alignments of the megaBLAST measured an E-value of 0.0. The seven haplotypes of front-end sequence Numt3(f), the three haplotypes of the Numt3(int) DNA-sequence and the four haplotypes of the back-end of ALBsc1709-Numt3 had very similar results, showing the highest similarity to ALBsc1709.

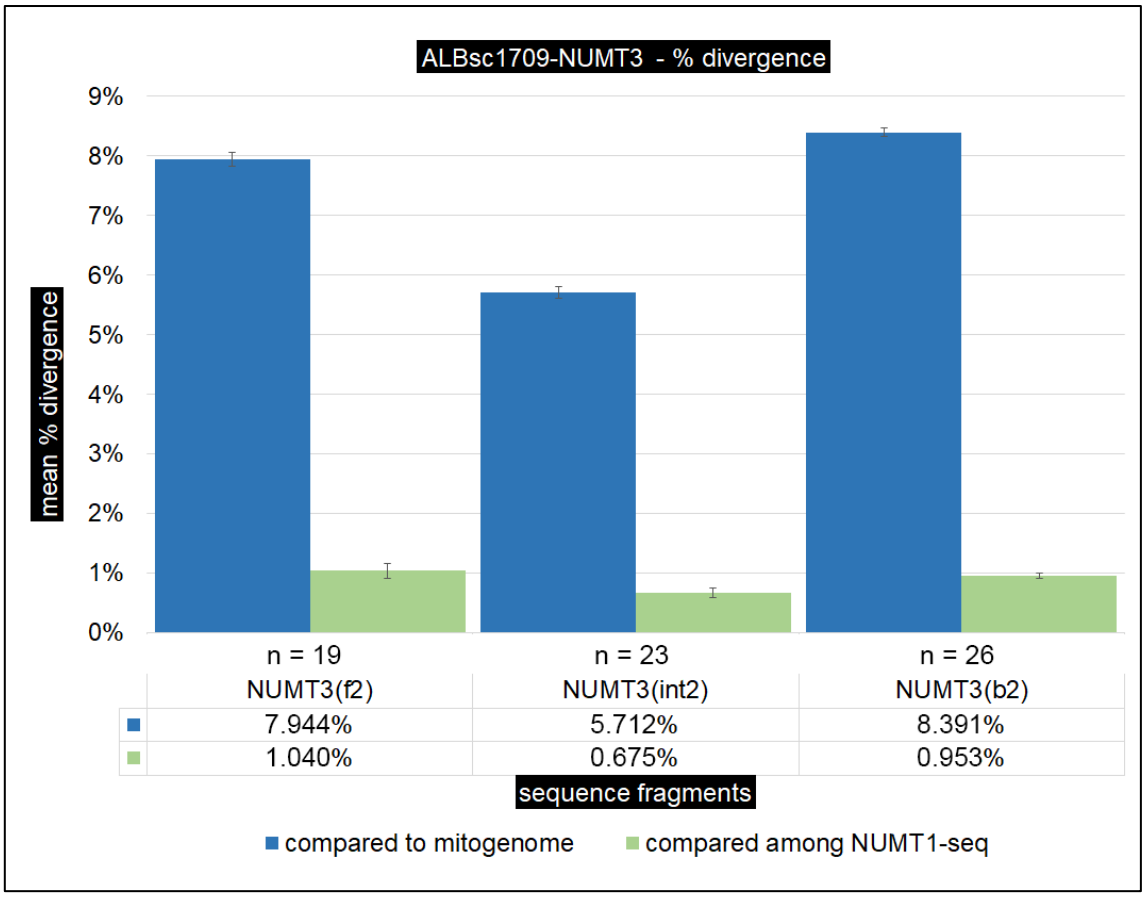


Figure 57. Estimates of evolutionary divergence in % between ALBsc1709_Numt3 sequences and their respective mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018).
The mean % divergences of 19 sequences of ALBsc1709_Numt3(f) with 773 positions, 23 sequences of ALBsc1709_Numt3(int) with 469 positions, 26 sequences of ALBsc1709_Numt3(b) with 980 positions to the respective mitochondrial counterparts (blue), as well as among the different haplotypes (green) are shown.

Within the TCS haplotype networks of **Figure 58** for **-a** ALBsc1709-Numt3(f2), **-b** ALBsc1709-Numt3(int2) and **-c** ALBsc1709-Numt3(b2) sequences the number of mutations that separate the NUMT-3-reference, the haplotypes from Germany and Italy as well as the mitochondrial counterparts of mtALB_ND2-COX1 are visualised. Strikingly, the sequenced DNA-segments for the front- and the back-end were many mutations apart from their respective NUMT-3-reference. For instance, the Numt(f2)_Type1 haplotype is seven mutations distant from the reference and the Numt3(b2)_Type1 haplotype is even nine mutations separated from the reference. The Numt3(f2) TCS network shows that the mitochondrial reference is separated by 55 mutations from the NUMT-3-reference and 58 mutations from the Numt3(f2)_Type1. The seven haplotypes themselves are just 1-3 mutations different from each other. The internal part of NUMT-3 seems to be less distant from the mitochondrial reference as well as within the NUMT-3-haplotypes. So, the mitochondrial ALB_ND2-COX1 sequence is 25 mutations distant from haplotype Numt3(int2)_Type1 and 24 mutations apart from the respective NUMT-3-reference. The three haplotypes of the internal NUMT-3 DNA-sequences are just 1-2 mutations

distant to each other. Even more distant from the mitochondrial counterpart are the back-end sequences Numt3(b2). The mitochondrial counterpart is 76 mutations apart from Numt3(b2)_Type1 and 73 mutations to the NUMT-3(b2)-reference. The four haplotypes of Numt3(b2) are only 1-2 mutations distant to each other.

In **Figure 59** a closer look at the genetic diversity of the whole ALBsc1709-Numt3 is pictured with a haplotype projection (**a**) and a TCS network (**b**) of the complete NUMT-3 sequence, which was formed from the consensus sequences of ALBsc1709_Numt3(f), ALBsc1709_Numt3(int) and ALBsc1709_Numt3(b). In these analyses, the alignments without the flanking regions of NUMT-3 were used for the TCS network, while the whole sequence was used for the haplotype projection. On the map it is observable that Numt3_Type1 is abundant in all three states of Germany with the locations Magdeburg, Hildrizhausen, Weil am Rhein and Neubiberg, as well as in Italy (Fermo). The haplotype Numt3_Type2 had its main occurrence in Bavaria with the locations Feldkirchen, Kelheim, Murnau, Miesbach and Schoenebach and again also in Piemonte, Italy in the locations Cuneo and Vaie. Numt3_Type3 was found only two times in the dataset, represented in the locations Feldkirchen (Germany) and Ostra (Italy). All other haplotypes Numt3_Type4-Type9 were all just sequenced once and are all representing the second haplotype of the heterozygous specimens for ALBsc1709-Numt3. In the TCS network of the whole Numt3 sequence, it is visible that the mitochondrial counterpart 374-ND2-COX1-2273(-) is 138 mutations distant to Numt3_Type1a and 134 mutations distant to the NUMT-3-reference on sc1709. The NUMT-reference ALBsc1709-Numt3 is also 16 mutations apart from the sequenced haplotype Numt3_Type1a. The nine sequenced haplotypes are quite close and only differ 1-3 mutations from each other.

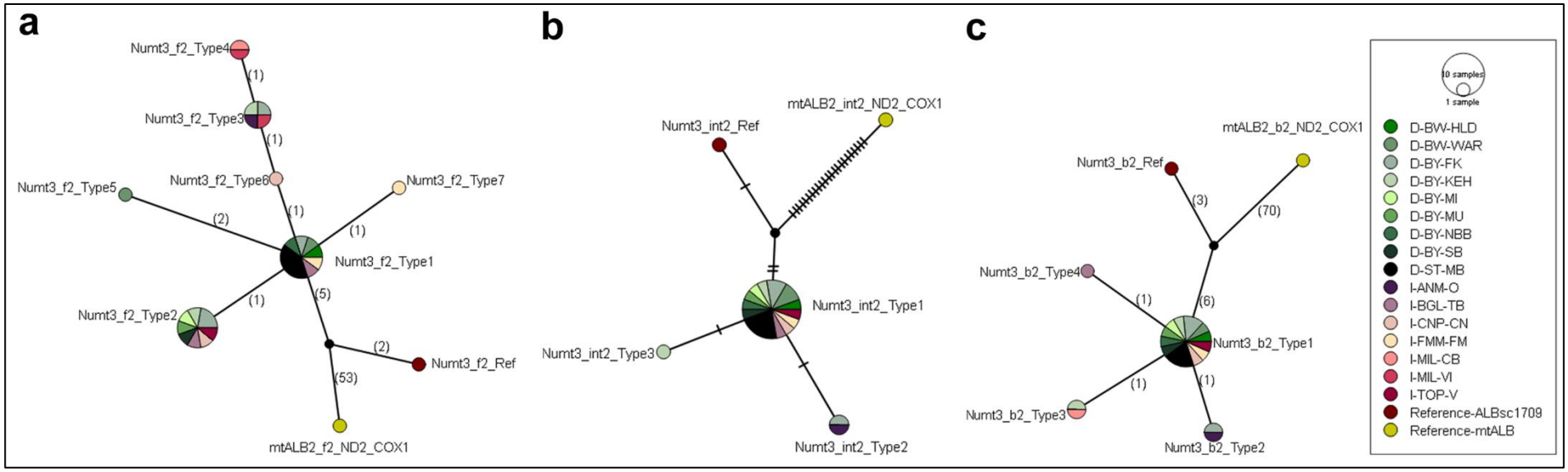


Figure 58. TCS haplotype network of haplotypes of **a)** ALBsc1709_Numt3(f2), **b)** ALBsc1709_Numt3(int2) and **c)** ALBsc1709_Numt3(b2) PCR products from Germany and Italy in comparison to the NUMT reference and the mitochondrial counterpart mtALB_ND2-COX1 (PopART v. 1.7). Mutations are represented by numbers or hatch marks; the size of the circles is proportional to the number of specimens.

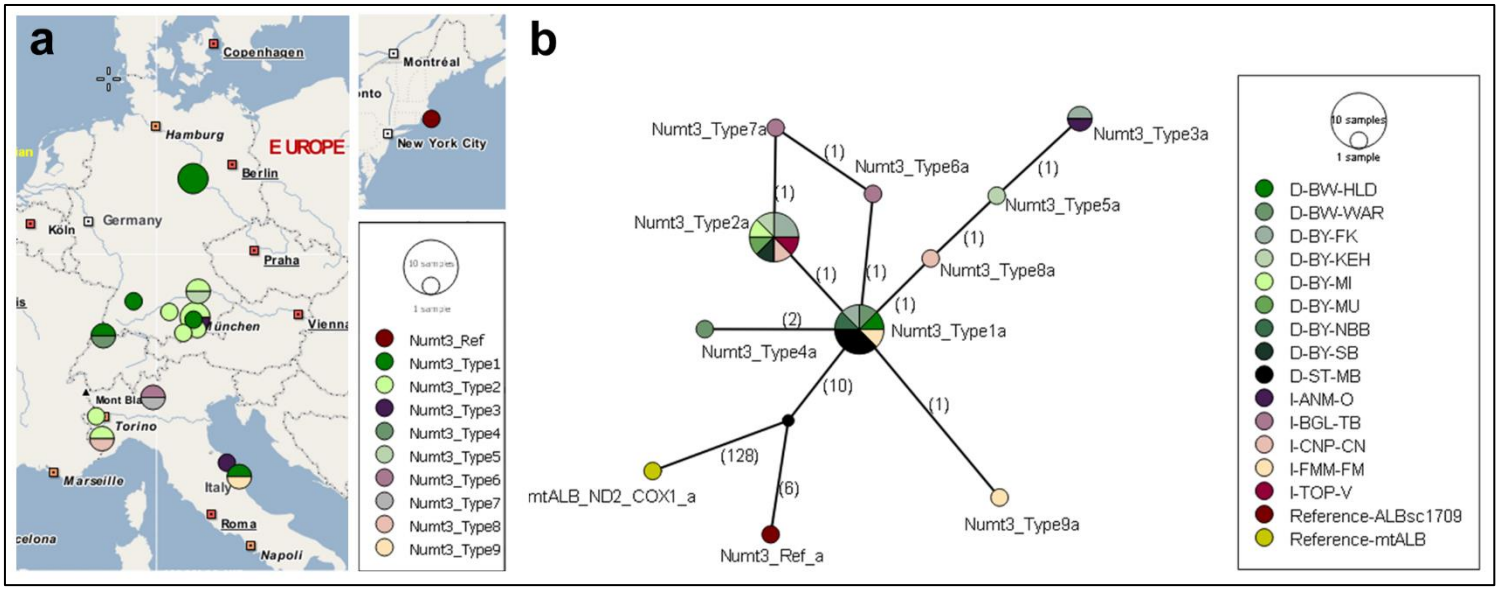


Figure 59. Haplotype projection (**a**) and TCS network (**b**) of the ALBsc1709-Numt3 complete sequence formed from consensus sequences of ALBsc1709-Numt3(f2), ALBsc1709-Numt3(int2) and ALBsc1709-Numt3(b2) (PopART v. 1.7).
a) For the NUMT-3-reference haplotype, the location as deposited on NCBI for the Biosample larva was used (USA: Otis Air National Guard Base, MA). **b)** Number of mutations are shown in brackets; The size of the circles is proportional to the number of specimens contributing to the haplotype.

The evolutionary history of the three NUMT-sequences of ALBsc1709_Numt3 were drawn by using the Maximum Likelihood method on the alignments of all three NUMT-3 sequence parts with their respective mitochondrial counterparts and NUMT-references, **Figure 60**. For all three NUMT-3 sequence parts the reference of the mitochondrial counterparts formed a distant outgroup, and the respective NUMT-references were also clearly separated from the sequenced haplotypes. The haplotypes showed all the same groupings as drawn by the TCS network, but are visualized in more detail with the short branch length and low confidence of the separations.

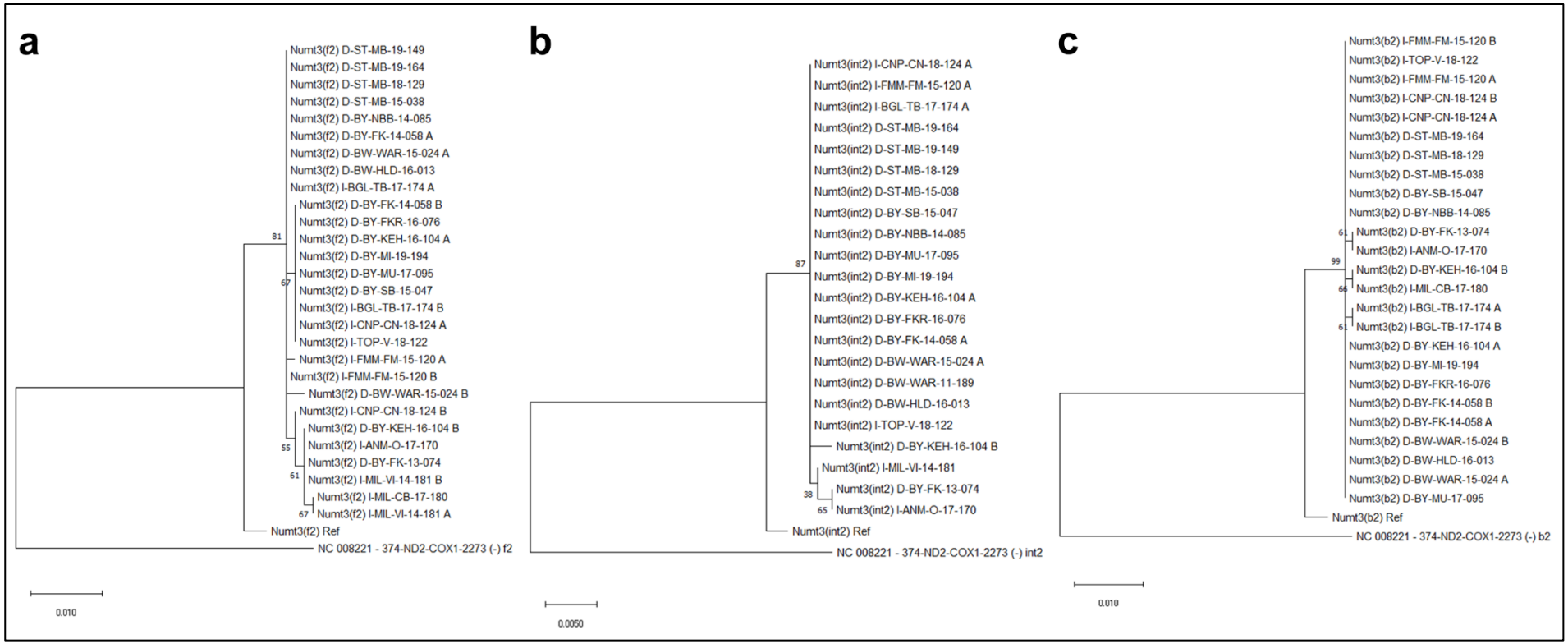


Figure 60. Evolutionary history inferred by using the Maximum Likelihood method on the alignments of all three sequences for ALBsc1709_Numt3 with their respective mitochondrial counterparts and NUMT-references in MEGA X (Kumar et al. 2018).
The Tamura 3-parameter model (Tamura 1992) was used for the pairwise distance estimation to draw the tree in **a)** and **b)**, while Hasegawa-Kishino-Yano model (HKY) (Hasegawa et al. 1985). was used for **c)**. Bootstrap method was conducted to test the results. The branch lengths measure in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). **a)** ALBsc1709_Numt3(f): This analysis involved 30 nucleotide sequences and a total of 773 positions in the final dataset. **b)** ALBsc1709_Numt3(int): This analysis involved 24 nucleotide sequences and a total of 469 positions in the final dataset. **c)** ALBsc1709_Numt3(b): This analysis involved 27 nucleotide sequences and a total of 980 positions in the final dataset.

5.4 Chapter C: results

Chapter C

"Complex European invasion history of *Anoplophora glabripennis* – new insights in its population genomic differentiation using Genotype-by-sequencing"

5.4.1 Discovery of genomic SNP markers for ALB

The genomic SNPs were gained via Genotype-by-sequencing approach. Using this approach, the genome representation and complexity is reduced by using a RE and an appropriate library preparation, so not the whole genome is sequenced. After mapping the sequence tags against the reference genome, the SNPs could be genotyped for each sample respectively (4.4.12). In **Figure 61** it is visualized how many specimens were sequenced with the GBS method per Pop-ID (**Table 4**). Here the only not represented Pop-ID is D-BW-KA.

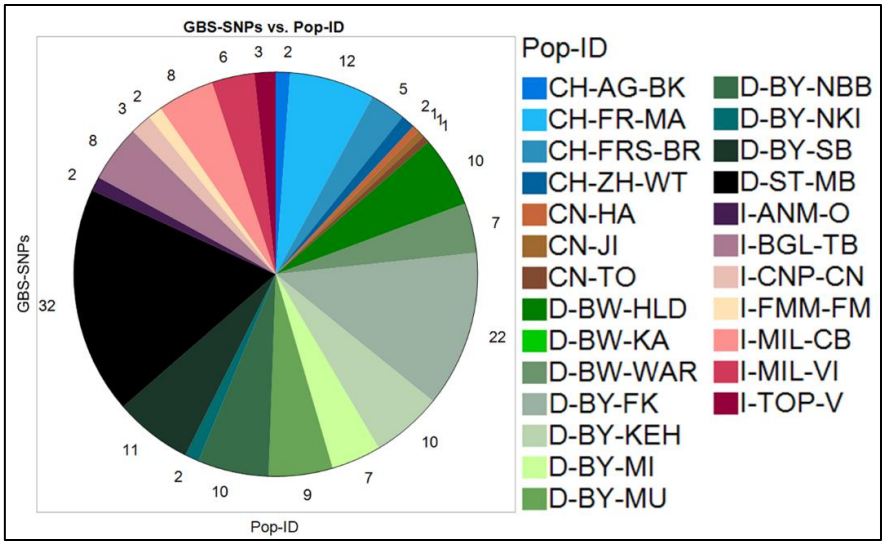


Figure 61. Represented ALB-populations by the SNP-markers obtained via GBS approach.

First the obtained variant called data needed to be processed in order to be used in the analyses described in 4.5.3.2. Initially, the quality assessment of the raw data was made to make a qualified choice on the filtering steps described in 4.5.2.2. The outputs generated with VCFtools v. 0.1.15 on the variant quality, missingness, read depth, allele frequency and heterozygosity were plotted in R. The results can be examined in **Figure 62**, **Figure 63** and **Figure 64**.

The results shown in **Figure 62** were used for the evaluation of the variant quality of the raw SNP data of 185 GBS-sequenced ALB specimens based on phred scores, the missingness per individual and site. Most of the SNPs had high quality genotypes, so the decision was made to only use variants with at least 99.9% base call accuracy (phred score 30). Considering

the missingness per individual, it can be observed that most of the individuals had a quite high relative count of missingness (F_miss value) between 0.6 and 0.7. In the first step it was decided to remove the seven individuals with higher F_miss values than 0.9. Also, the missingness per site in this dataset was very high, so the usual cut-off of 0.25 would not be proportionate for the available data. Therefore, the datasets were filtered on basis of the proportion of missing data to exclude sites with either >50%, >25%, >10% or >1% missingness. The results of all these steps were included in the comparison of outputs on missingness, depth, inbreeding or principal components used to evaluate the data on putative batch effects, which will be described later in this subchapter. A summary on the missingness of the variants (var_miss$fmiss) is shown in **Table 17** which also shows that most sites had high missingness, and there were some extremes with >60% missingness too, which should be excluded in any case. To evaluate the sequencing quality of the raw SNP data of 185 GBS-sequenced ALB specimens, the read depth per site and individuals was calculated. The proportions are depicted in **Figure 63**. Based on the proportions shown in this figure, a minimum depth cut-off of 3 was set to exclude false positive calls. Most of the variants had a read depth between 3 and 5, as visualized in **Figure 63-a** and **-b**. A maximum depth cut-off of 55 was set to remove mapping errors or repetitive sequences, since it is approx. doubled the mean depth (**Table 17**). The very high mean depth values (**c**) per individual will be dropped by this filtering as well.

In **Figure 64** the minor allele frequencies (MAF) per site and the inbreeding coefficient (f) per individual are shown in the raw ALB SNP data. Regarding the MAF, it is conspicuous that there is a high density on very rare alleles with <0.05. Hence, the commonly used filtering cut-off to MAF values above 0.05-0.02 was not proportionate for this dataset. The filter for MAF was set to 0.01 to include some rare alleles, but not too rare ones. In **Table 17** it is shown that this represents the variants from the 1^{rst} quantile. Additionally, one subset with 170 individuals, filter 10, g50 and m50 setting was also used to check if the ancestries calculated based on rare and not so rare alleles do not overestimate the population structure, see ch. 5.4.3. Since the dataset was sampled on more than one population, the inbreeding coefficient which assumes Hardy-Weinberg-equilibrium was only used here to check if some individuals stand out with strongly negative or inflated f-values. Both cases can be used as indicator for problems like contaminations, sequencing errors or duplicates. Most of the individuals in the raw data had strongly negative or high f-values. But, in ch. 5.4.2 (**Figure 71**) it is assured, that the filtering settings strongly diminished the issues indicated by the f-values in the raw data.

Table 17. Summary of the variant missingness, read depth per site and allele frequencies of the ALB SNP raw data.

	Min.	1st quantile	Median	Mean	3rd quantile	Max
var_miss$fmiss	0.000	0.317	0.697	0.612	0.958	0.995
var_depth$meandepth	1.150	3.333	8.738	27.149	32.300	7093.000
var_freq$maf	0.000	0.014	0.143	0.179	0.325	0.500

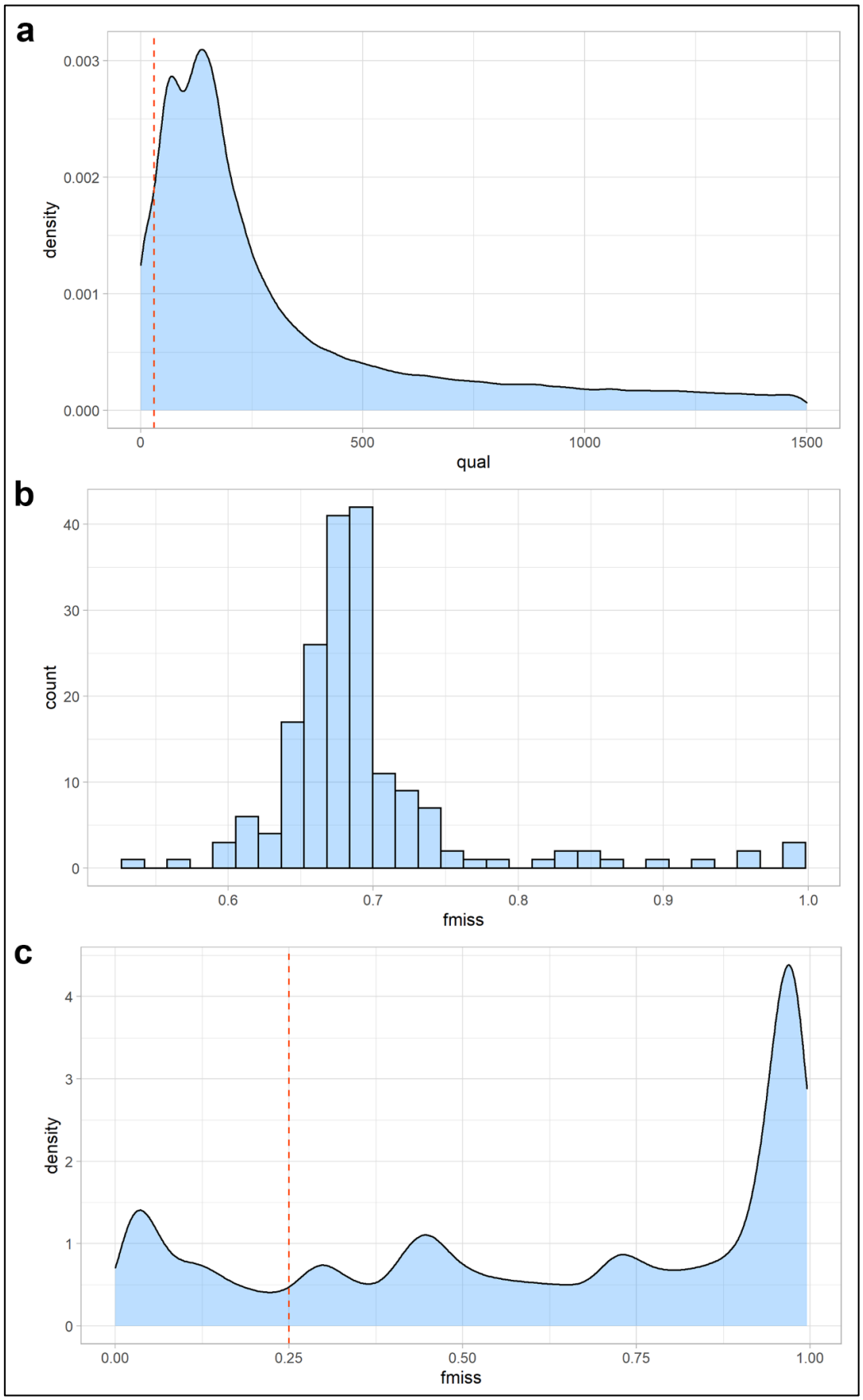


Figure 62. Evaluation of the data quality of the raw SNP data of 185 GBS-sequenced ALB specimens based on **a)** the site quality (red line = phred score 30) **b)** the relative count of missingness per individual (fmiss) and the **c**) proportion of missingness per site (fmiss) (red line = 0.25).

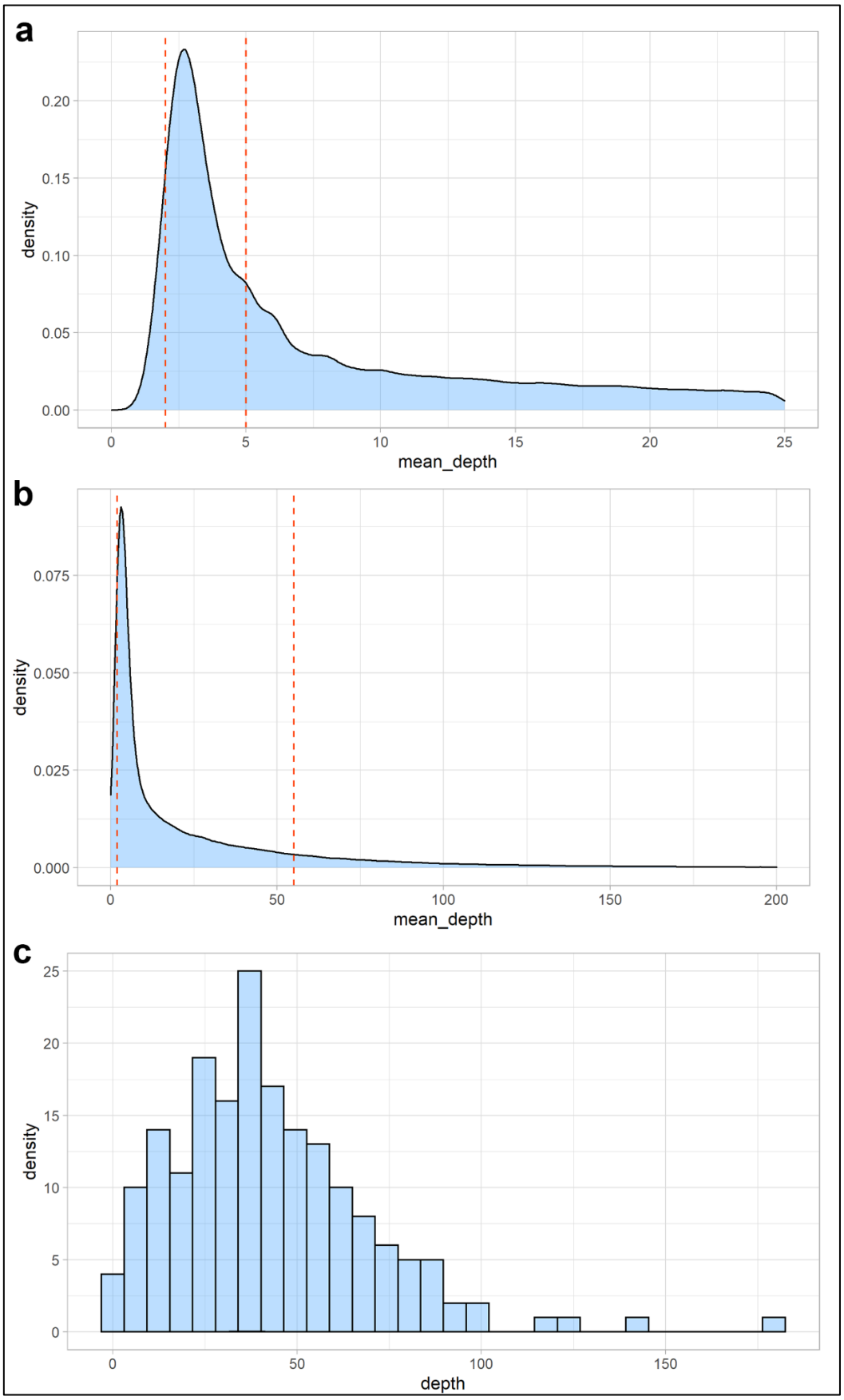


Figure 63. Evaluation of the sequencing quality of the raw SNP data of 185 GBS-sequenced ALB specimens based on the variant mean depth (**a, b**) (red lines = 3; 5, 55) and the mean depth per individual (**c**).

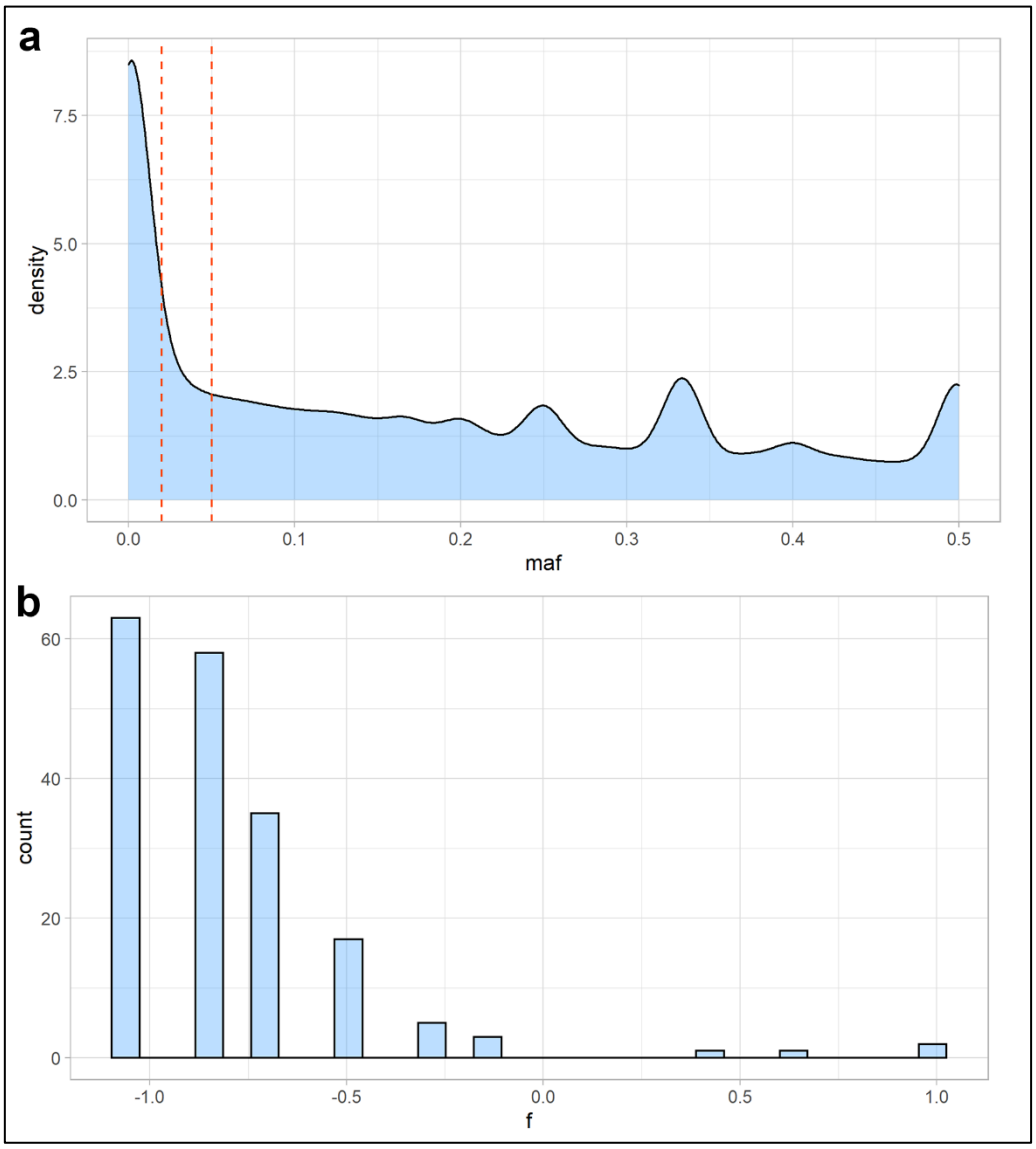


Figure 64. Evaluation of the distribution of the allele frequencies by minor allele frequencies (MAF) per site (red lines = 0.05; 0.02) (**a**) and the inbreeding coefficient f (**b**) of the raw SNP data of 185 GBS-sequenced ALB specimens.

In **Table 18** all filtering steps conducted on the raw data and following recoded VCF-files are represented with their SNP and individual count. The setting describes which filtering steps were applied as described in 4.5.2 and 4.5.3. For both approaches (170 and 178 individuals), the filter steps on the allowed missingness of 10% obviously led to an enormous decrease and 1% to a complete loss of variants for the analyses. The following outputs which test for the presence of effects on the data by non-biological factors (batch effects) in this dataset were also used to decide on more relaxed or stricter filtering along the way. To evaluate the dataset on batch effects, measures for individual depth, missingness and principal components were plotted for the raw data of the merged three GBS runs and the subsets 170f8g50m50, 170f8g75m25 and 170f8g90m10. A description of the abbreviations of the subsets can be found in 4.5.2.2.

Table 18. SNP and individual count at all filter setting described in 4.5.2 and 4.5.3.

VCF-file (.vcf)	SNPs	individuals	setting
GBS-Run1_Variants_raw_All.vcf	239604 out of 239604 sites	72 out of 72	raw data Run1
GBS-Run2_Variants_raw_All.vcf	128391 out of 128391 sites	80 out of 80	raw data Run2
GBS-Run3_Variants_raw_All.vcf	180962 out of 180962 sites	33 out of 33	raw data Run3
Variants_raw_185Indv.vcf	331191 out of 331191 sites	185 out of 185	raw data merged
178Indvf7.recode.vcf	73364 out of 331191 sites	178 out of 185	filter7
178Indvf8.recode.vcf	64830 out of 73364 sites	178 out of 178	filter8
170Indvf8.recode.vcf	64830 out of 64830 sites	170 out of 178	filter8, Fmiss cut-off 0.7
178Indvf9.recode.vcf	55810 out of 331191 sites	178 out of 185	filter9
178Indvf10.recode.vcf	50340 out of 55810 sites	178 out of 178	filter10
170Indvf10.recode.vcf	50340 out of 50340 sites	170 out of 178	filter10, Fmiss cutoff
170Indvf8g50m50.recode.vcf	32432 out of 64830 sites	170 out of 170	max-missing 0.5
170Indvf8g75m25.recode.vcf	10226 out of 64830 sites	170 out of 170	max-missing 0.75
170Indvf8g90m10.recode.vcf	922 out of 64830 sites	170 out of 170	max-missing 0.9
170Indvf8g99m1.recode.vcf	0 out of 64830 sites	170 out of 170	max-missing 0.99
178Indvf8g50m50.recode.vcf	31030 out of 64830 sites	178 out of 178	max-missing 0.5
178Indvf8g75m25.recode.vcf	7738 out of 64830 sites	178 out of 178	max-missing 0.75
178Indvf8g90m10.recode.vcf	200 out of 64830 sites	178 out of 178	max-missing 0.9
178Indvf8g99m1.recode.vcf	0 out of 64830 sites	178 out of 178	max-missing 0.99
170Indvf10g50m50.recode.vcf	27777 out of 50340 sites	170 out of 170	max-missing 0.5
N / A	27599 out of 32432 sites	170 out of 170	--minDP 9 --min-meanDP 9

filter7 (f7) = `--minQ 30 --remove-filtered-all --maf 0.01 --mac 3 --max-alleles 2 --min-alleles 2 --min-meanDP 3 --minDP 3 --max-meanDP 55 --min-alleles 2`, cut-off 0.9 on F_miss for individuals (`--remove <file>`); filter8 (f8) = `--thin10` on f7-filtered subset; 185Indv = specimen count on the unfiltered merged VCF-file; 178Indv = specimen count after cut-off 0.9 on individual F_miss; 170Indv = specimen count after cut-off 0.7 individual F_miss on f8-filtered subset; g50m50 = genotype call rate 50% and tolerated missingness 50% (analogous for higher genotype rates); filter9 (f9) = analogous to f7, but `--maf 0.1`; f10 (f10) = f8 on f9 filtered subset. N/A = no VCF-file available, filter just set on output `--het` on subset 170Indvf8g50m50. All individual counts include D-BY-SB-15-047 as internal control, for GBS-Run2 and GBS-Run3 only once, but in all following VCF-files three times respectively.

The mean depth per individual plotted against the individual missingness (f_miss) depicted in **Figure 65** initially indicated in the subset 170Indvg50m50 a clear separation by the different GBS-runs (library 1-3), which disappears with stricter filtering. But when focusing on **Figure 66**, where the same measures are coloured by location, it is quite obvious that the separation by GBS-run is also caused by the assignment of locations. Since the experimental design was not considering putative batch effects in the first place, all locations, except the internal controls, were only represented in one run. It is clearly observable that e.g. the Swiss locations showed both higher missingness and higher mean depth per individual and seemed thereby separated. This separation in cleared up by more stringent filtering, but it stays obvious that GBS-run3 had a higher missingness than the others while GBS-run2 had the least. In **Figure 67**, the principal components 1 and 2 were plotted of the raw data and the three subsets coloured by location and the GBS-runs as shapes. The noticeably bigger cluster around the zero-point (x-axis 0.0-0.05/y-axis 0.0) clearly stands out, which shows that all three GBS-runs were represented, including the internal controls from Schoenebach, Germany. The two clusters separated in the raw data from GBS-run2 and 3 (Swiss and Italian) were united by filtering. The only clusters that were solely represented by one library were four smaller ones from GBS-run2 from Italy, China and Germany, as well as one from GBS-run1 from Germany. Hence, the results from the PCA are not or only marginally affected by the different GBS-runs merely by location. Also, the results from the PCA stay highly similar even with more strict filtered subsets.

From here on, the choice was made to use subset 170Indvf8g50m50 for most analyses about genetic diversity and the evolutionary relationships with 32432 SNPs. 178Indvf8g50m50 was also used for ancestries estimated by ADMIXTURE v. 1.3.0 and the phylogeny with Splitstree v. 4.17.1 to test the kinship of D-BY-SB-18-110 to earlier findings from 2015 in Schoenebach.

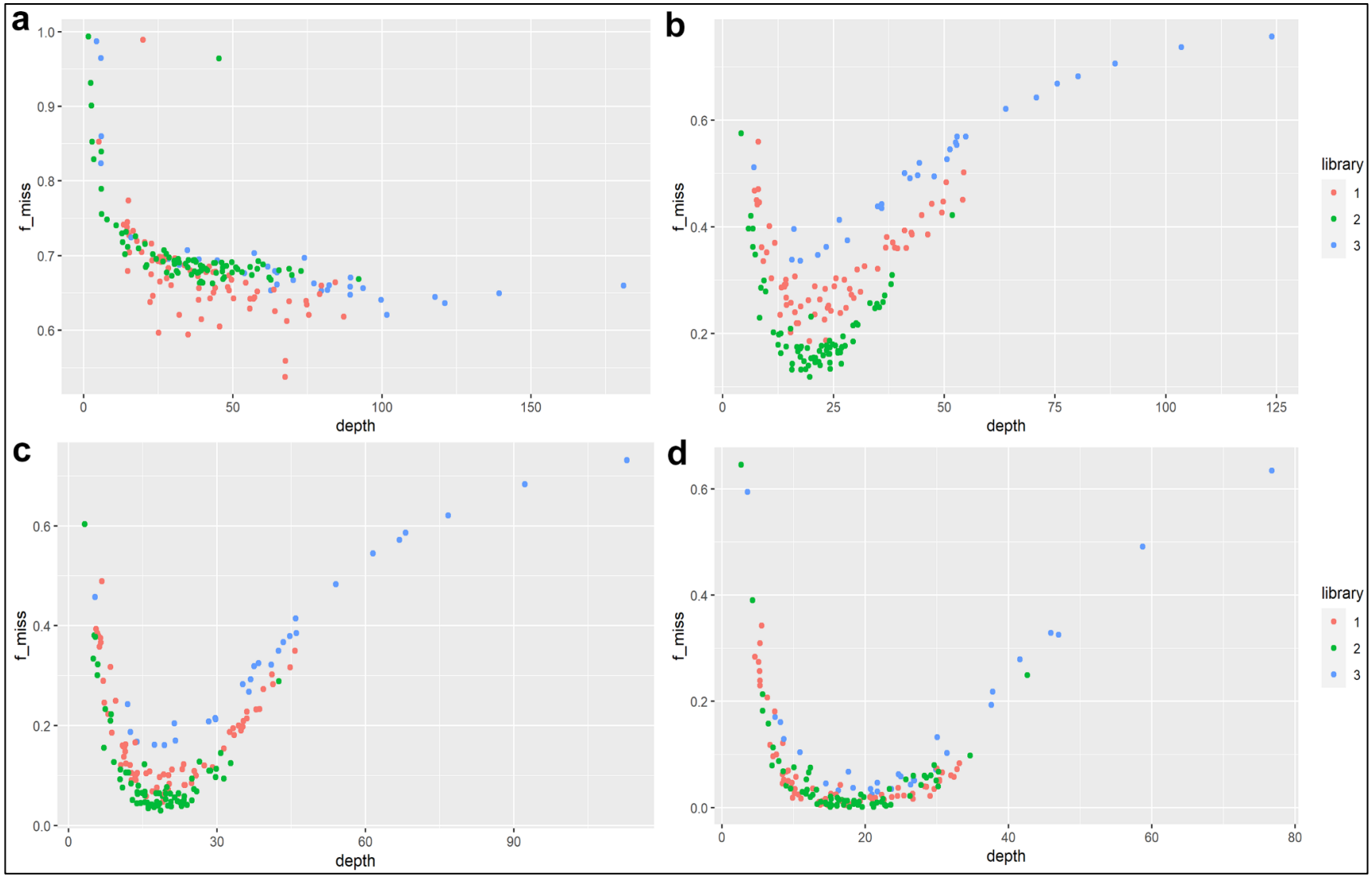


Figure 65. Mean depth per individual (depth) plotted against the individual missingness (f_miss) coloured by library from the raw ALB SNP dataset with 185 individuals (**a**) and the different filtered data subsets **b**) 170Indvf8g50m50, **c**) 170Indvf8g75m25 and **d**) 170Indvf8g90m10.

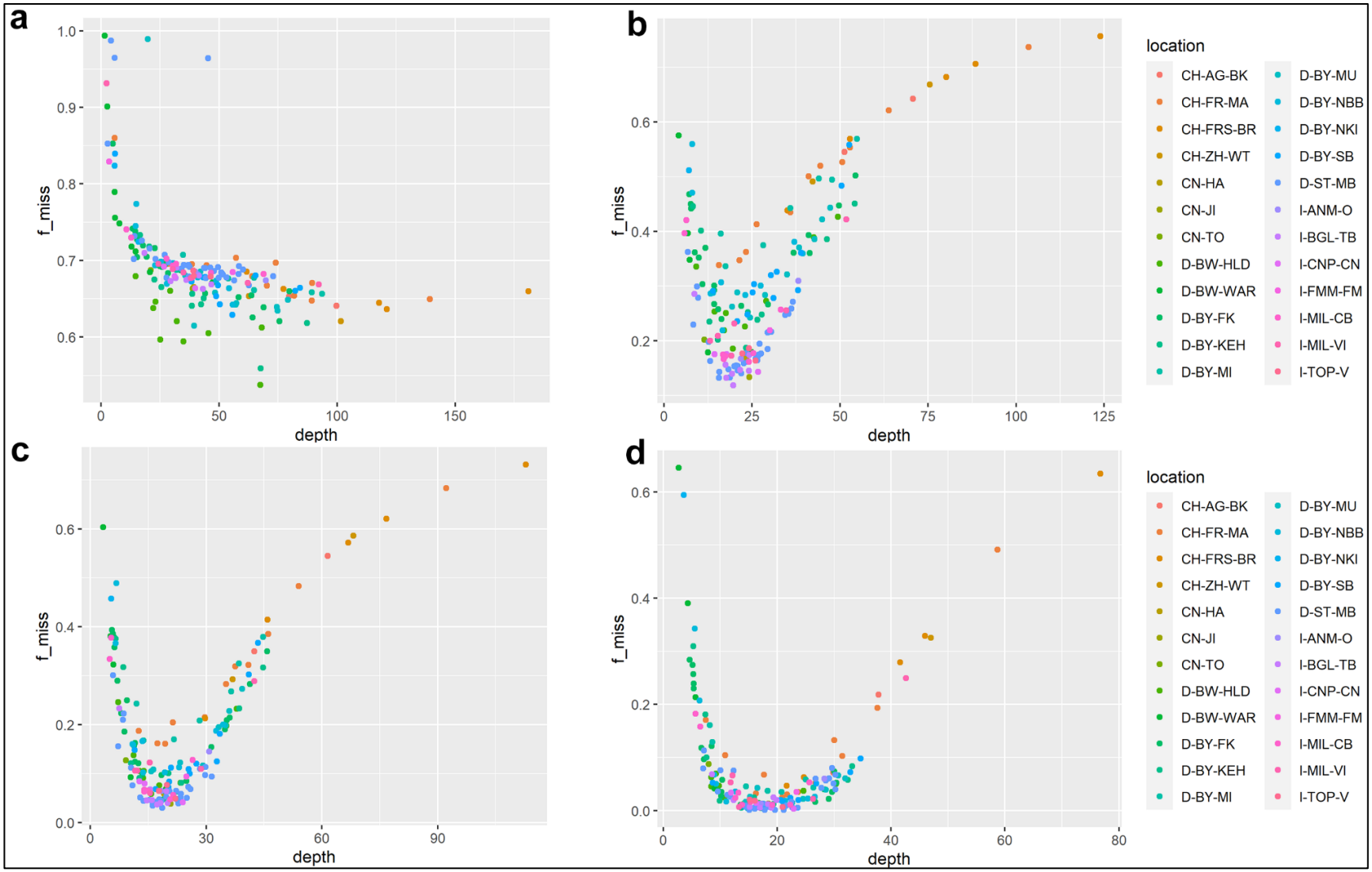


Figure 66. Mean depth per individual (depth) plotted against the individual missingness (f_miss) coloured by location (Pop-ID) from the ALB SNP datasets: raw 185 individuals (**a**) and the different filtered data subsets **b**) 170Indvf8g50m50, **c**) 170Indvf8g75m25 and **d**) 170Indvf8g90m10.

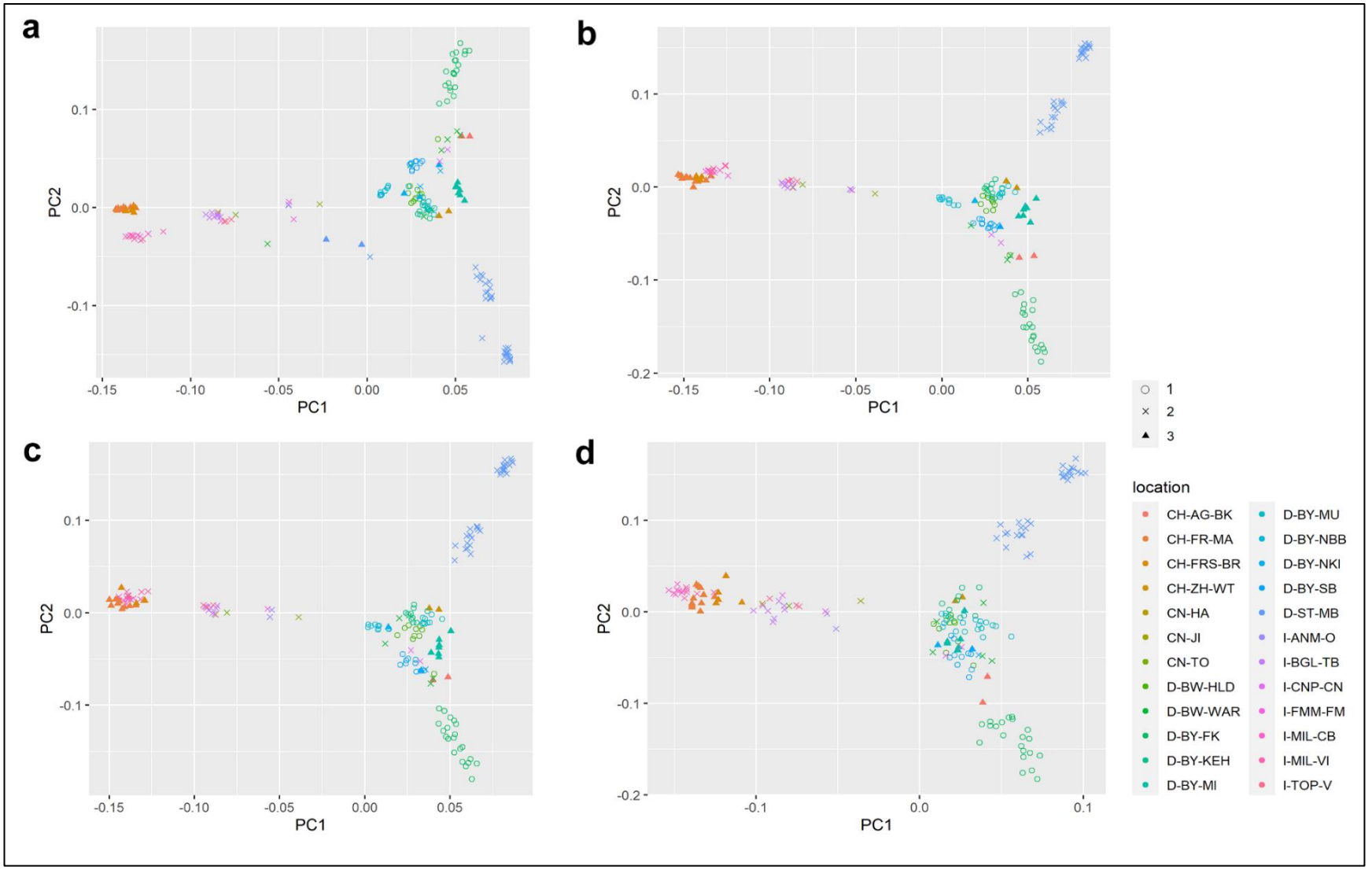


Figure 67. Scatter plot of principal components 1 and 2 coloured by location and shaped by library from the raw ALB SNP dataset with 185 individuals (**a**) and the different filtered data subsets **b**) 170Indvf8g50m50, **c**) 170Indvf8g75m25 and **d**) 170Indvf8g90m10.

5.4.2 Genetic diversity and population structure of invasive ALB populations in Europe

For more representative SNPs for the genetic structure in the principal component analysis (PCA) and the Admixture analysis, SNPs which are in linkage disequilibrium (LD) needed to be pruned beforehand. **Figure 68** pictures the decay of the linkage disequilibrium with the increasing distance of the SNPs to each other, shown with distances from 0-99 kb with bins of 1 kb. The value 0.15 was used as a threshold to linkage equilibrium (LE) to determine the independent SNPs, since this equates roughly half of the maximum decayed (e.g. Vos et al. 2017). Subsequently, the subset 170Indvf8g50m50 was pruned in PLINK v. 1.9 for SNPs exceeding the r^2-threshold of 0.15 in sliding windows of 25 kb and a step size of 5 (variant count). First, pairs of variants in the current window with squared correlation greater than the threshold are noted and in a second step removed. From the 32432 variants which were loaded with a genotyping rate of 0.688608, 7810 independent variants with a total genotyping rate of 0.685353 remained for the PCA and Admixture analyses. The LD-pruned 170Indvf10g50m50 subset had a total genotyping rate of 0.689601 and 6064 variants. The LD-pruned 178Indvf8g50m50 subset had a total genotyping rate is 0.669861 and 7673 variants.

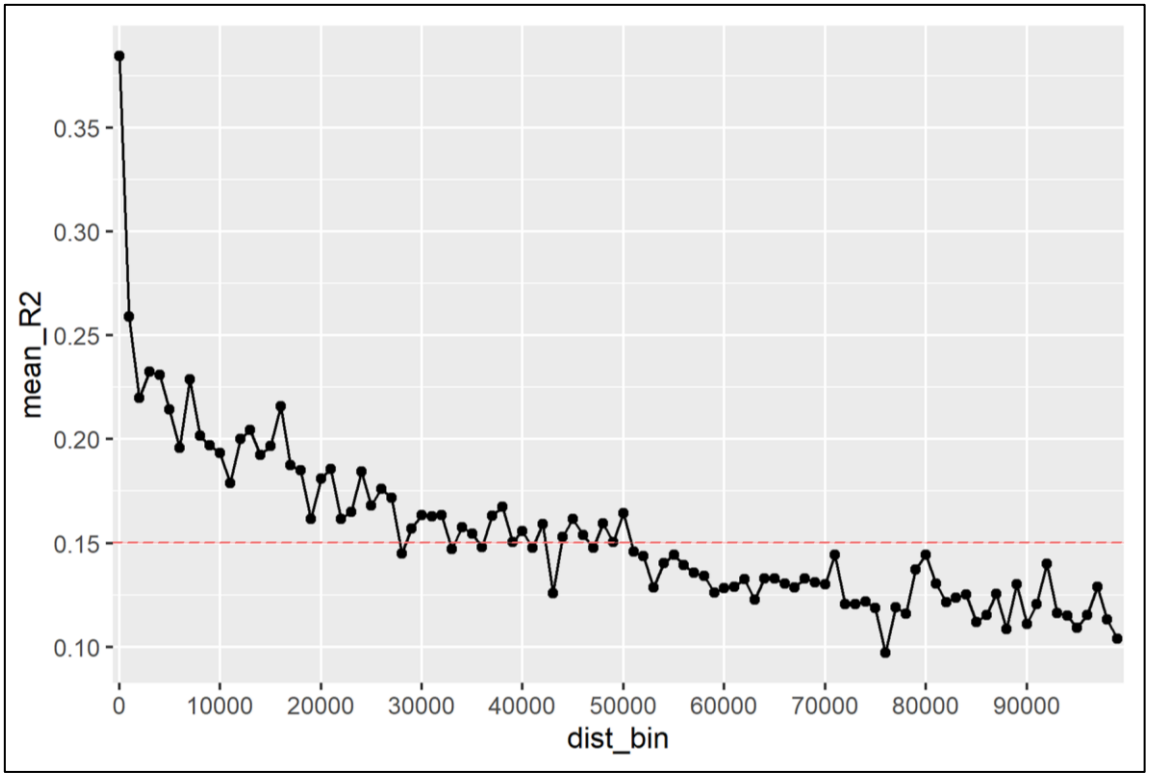


Figure 68. Decay of the linkage disequilibrium (LD) with physical distance (0-9.9 kb) between SNPs of the ALB SNP dataset 170Indvf8g50m50.
The value of 0.15 of mean-r^2 (red line) was set as a threshold for linkage equilibrium (LE).

For analyses with per site measures like for nucleotide diversity (π) or F_{ST}, it was mandatory to reduce the plots with a cut-off for 3 mb for the scaffold size. In **Figure 70** the sequence length of all 9866 scaffolds and contigs of the Assembly Agla_2.0 GCA_000390285.2 of ALB are shown. Hence, a filter for scaffolds >3 mb would only include the scaffolds 1-8 (KZ487478.1, KZ486783.1, KZ486944.1, KZ487026.1, KZ486983.1, KZ487205.1, KZ489286.1, KZ486808.1).

The nucleotide diversity (π) as a measure of genetic variation within the SNP subset 170f8g50m50 is shown in **Figure 69** within 10 kb windows and 10 kb step size on the eight biggest scaffolds. The nucleotide diversity in the 10 kb windows is mostly between 0-0.00005. Less values are between 0.00005-0.00015, and only a few values of the 10k windows are comparatively higher between 0.00015 and 0.00025. The overall nucleotide diversity is pictured here only as an example but was measured to be low within each of the sampled European populations as well (not shown).

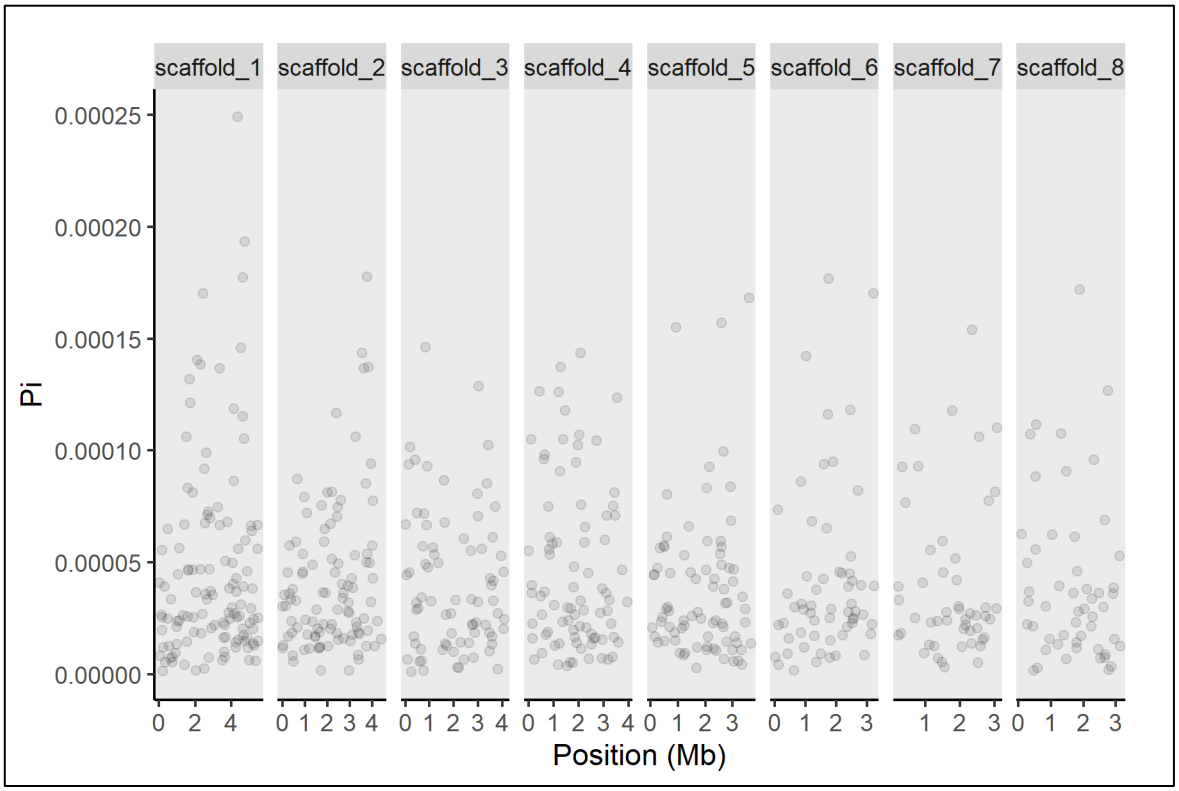


Figure 69. π-bins of 170Indvf8g50m50 subset of ALB SNP data in 10kwindows and 10ksteps on ALB scaffold 1-8.

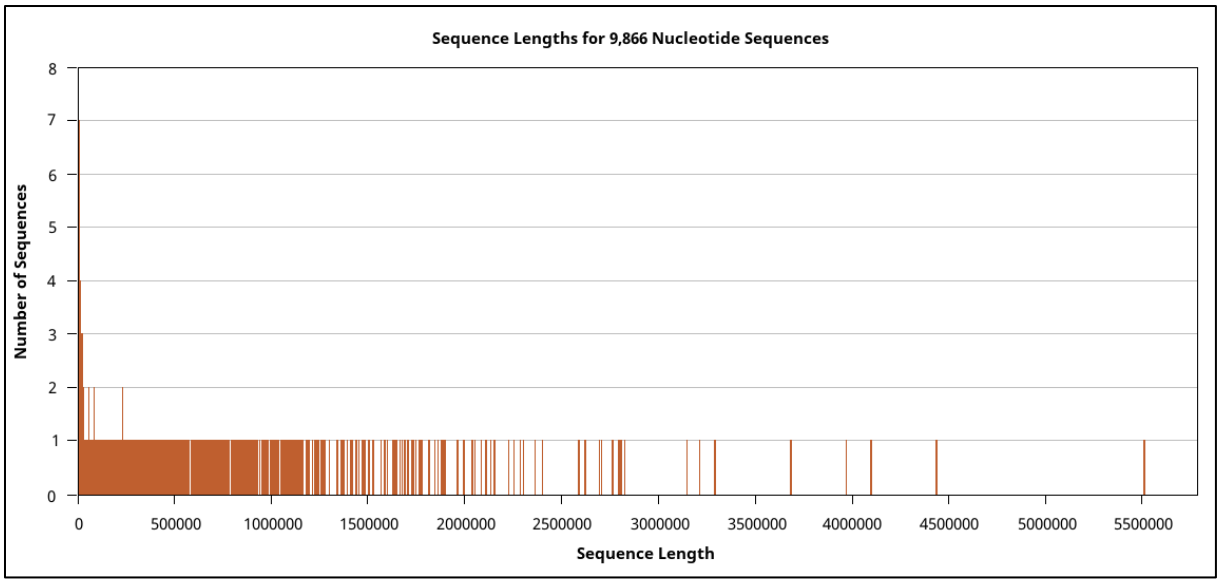


Figure 70. Sequence length distribution for all 9866 scaffolds and contigs of the Assembly Agla_2.0 GCA_000390285.2 of *A. glabripennis* from December 12, 2017.

The values on expected and observed homozygosity, along with inbreeding coefficient on an individual basis were calculated with additional filtering on read depth of at least nine to be sure that a homozygote call is actually homozygous. Thus, the estimates were obtained on a basis of 27599 SNPs on the 170Indvf8g50m50DP9-het subset. Most of the individuals measured an inbreeding coefficient between 0.4 and 0.6, while some individuals from Vittuone, Corbetta, Ostra, Feldkirchen and Marly had f-values >0.7<0.8 and one individual from Corbetta even >0.8. The lowest inbreeding coefficients with f-values <0.2 were found in some individuals from Kelheim, Magdeburg, Neukirchen am Inn, Cuneo and Weil am Rhein. When focus on the distribution of observed and expected homozygosity it stands out, that the observed homozygosity as well as the expected homozygosity were very high in most individuals from Magdeburg and in most Italian populations, except Cuneo and Fermo. The same observation was made for two Chinese specimens. The third Chinese specimen from Tongliao was moderate in comparison to the others. Most other German and Swiss specimens showed not extremely high observed homozygosity, but high expected homozygosity.

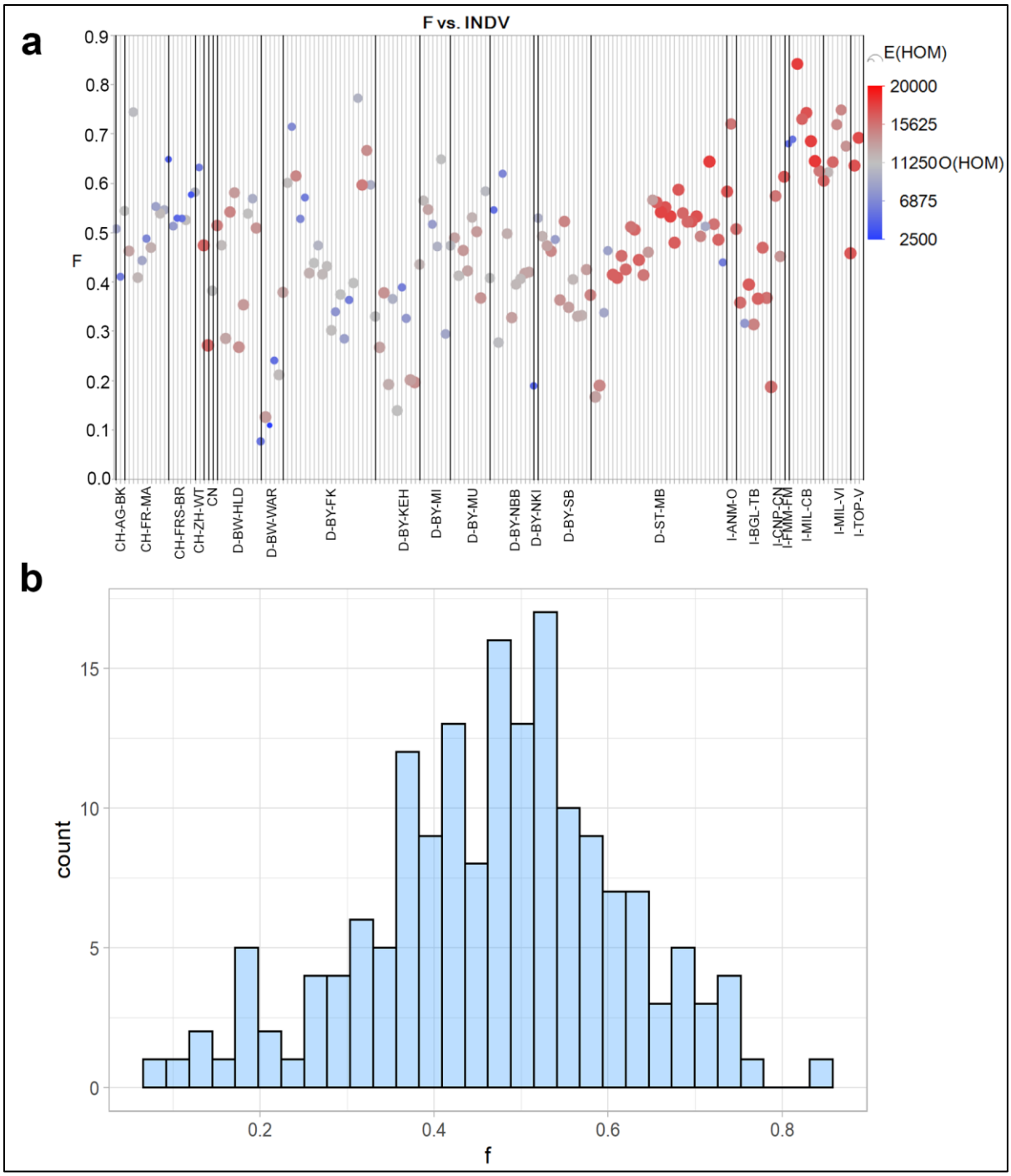


Figure 71. Measures of Heterozygosity on the 170f8g50m50minDP9-het dataset. **a)** f value per individual, observed (colour) and expected homozygosity (size) in JMP **b)** f value distribution in R;

As a measure for the intraspecific genetic differentiation, the pairwise F_{ST}-values between the populations, as well as the Analysis of molecular variance (AMOVA) were conducted in Arlequin v. 3.5. The AMOVA tested no genetic structure and a country-wise genetic structure with the assumption of no differentiation within and between groups. The results are demonstrated in **Table 19**. The locus-by-locus AMOVA revealed very high overall F_{ST}-values of >0.4 between the subpopulations in both tested scenarios. The non-hierarchical analyses estimated ~41% of the variation between the populations and ~58% of the variation within the populations. With a grouping according to the countries, the lowest molecular variation was detected between the countries (~18%), whereas the highest genetic diversity was also within the populations (~53%) and with ~29% a bit lower between the populations within a country. The measures to determine the degree of differentiation, the Fixation indices F_{ST}, F_{SC} and F_{CT} are also shown in **Table 19**. The highest degree of differentiation was estimated in both tested scenarios with F_{ST}-values to show the differentiation among the demes with >0.4. The differentiation between the populations within a country (F_{SC}) is also very high with ~0.35. Otherwise, the degree of differentiation among the countries (F_{CT}) was relatively moderate with a value of ~0.17.

Table 19. Locus by locus AMOVA design and results on ALB SNP subset-170f8g50m50.

Source of variation	Sum of squares	Variance components	Percentage variation	Fixation indices	p values
(a) AMOVA - No. of groups 1 (Dataset-170f8g50m50)					
Among populations	295637.778	1387.92919 (Va)	41.59582***		0.00000
Within populations	406914.609	1948.77426 (Vb)	58.40418	[1]0.41596***	0.00000
Total	702552.387	3336.70345			
(b) AMOVA - No. of groups 3 - CH, D, I (Dataset-170f8g50m50)					
Among groups	94129.356	664.32803 (Va)	17.97123***	[3]0.17971***	0.00000
Among populations within groups	201508.421	1083.51815 (Vb)	29.31105***	[2]0.35733***	0.00000
Within populations	406914.609	1948.77426 (Vc)	52.71773***	[1]0.47282***	0.00000
Total	2552.387	3696.62043			

Significance tests (1023 permutations): *** $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.1$; [1]F_{ST}, [2]F_{SC}, [3]F_{CT} **a**) no hierarchical structure, "out" population not considered **b**) structure to test: populations of Switzerland, Germany and Italy grouped to the respective country, "out" population not considered. AMOVA based on average over 17773 loci.

The pairwise differentiation between the sampled populations was measured in pairwise F_{ST}-values based on genotype frequencies. For this estimation, only populations with at least 5 specimens were considered, also for the locus by locus AMOVA. All other specimens were assigned to the group "out". The pairwise F_{ST} matrix is pictured in **Figure 72** with grades of colouring from low to high, while the exact F_{ST}-values (**Table 42**) and the corresponding p-values (**Table 43**) are offered in Appendix, IV Chapter C. For this matrix, only the pairwise F_{ST} between "out" and Weil am Rhein was not significant. In the following, some comparisons between populations with moderate genetic differentiation, strong or very strong genetic differentiation will be emphasized as detected with the pairwise F_{ST} among populations. Moderate genetic differentiation (~0.05-0.15) was examined between Kelheim (D-BY-KEH) and Trescore-Balneario (I-BGL-TB), Kelheim and Weil am Rhein (D-BW-WAR), as well as between Marly (CH-FR-MA) and Bruensried (CH-FRS-BR). With F_{ST}-values slightly above 0.15 but <0.2 Trescore-Balneario showed considerably moderate differentiation to Marly, Bruensried and Weil am Rhein. Very strong genetic differentiation to all the other subpopulations was detected in Marly and Bruensried. Corbetta (I-MIL-CB) and Vittuone (I-MIL-VI) were also strongly differentiated to the other populations with F_{ST}-values >0.4 and even some with >0.6. The only demes with still very high >0.3 F_{ST}-values but <0.4 to Corbetta/Vittuone were Kelheim and Trescore-Balneario as well as to each other. For Marly/Bruensried the only demes with >0.2 but <0.4 F_{ST}-values to these two populations were Weil am Rhein and Kelheim. In general, for most of the pairwise F_{ST}-values between the sampled demes, great (0.15-0.25) to very great genetic differentiations over all sites (>0.25) were estimated, according to Wright's guidance for the interpretation of F_{ST} (Nielsen and Slatkin 2013).

Beside the overall F_{ST}-values between populations, also the F_{ST}-values per site were considered in addition. But in these cases, not the single populations were used in the comparisons, but the three German states (**Figure 73**) and the three sampled European countries (**Figure 74**). As already done for the nucleotide diversity before, the F_{ST} values in the 50 k windows were restricted to scaffolds with a size >3 mb for plotting. When focusing on the comparison between the German states (**Figure 73**), the comparison between Baden-Wuerttemberg (BW) and Saxony-Anhalt (ST) had sporadic F_{ST}-values per bins up to 0.9. The comparison between BW and ST had most F_{ST}-values per bins >0.5, the comparison between Bavaria (BY) and ST had less F_{ST}-values per bins >0.5 and the comparison BW and BY had least F_{ST}-values per bins >0.5. There were a few negative F_{ST}-values per bins for all pairwise comparisons. In general, in all the comparisons the most F_{ST}-values were regular distributed between ~0-0.5 with an overhang in the range ~0-0.25.

The pairwise F_{ST}-values per 50 k bins between the countries Germany, Switzerland and Italy (**Figure 74**) showed a very similar pattern. In this case the comparison between Germany and

Switzerland showed the most F_{ST}-values >0.5, Germany vs. Italy showed less, but Italy and Switzerland had the least F_{ST} estimated >0.5. A few scattered pairwise F_{ST}-values between Switzerland and Italy were much higher up to >0.9. The distribution of all the F_{ST}-values was again between ~0-0.5 with an overhang between ~0-0.25 and some negative F_{ST}-values were detected too.

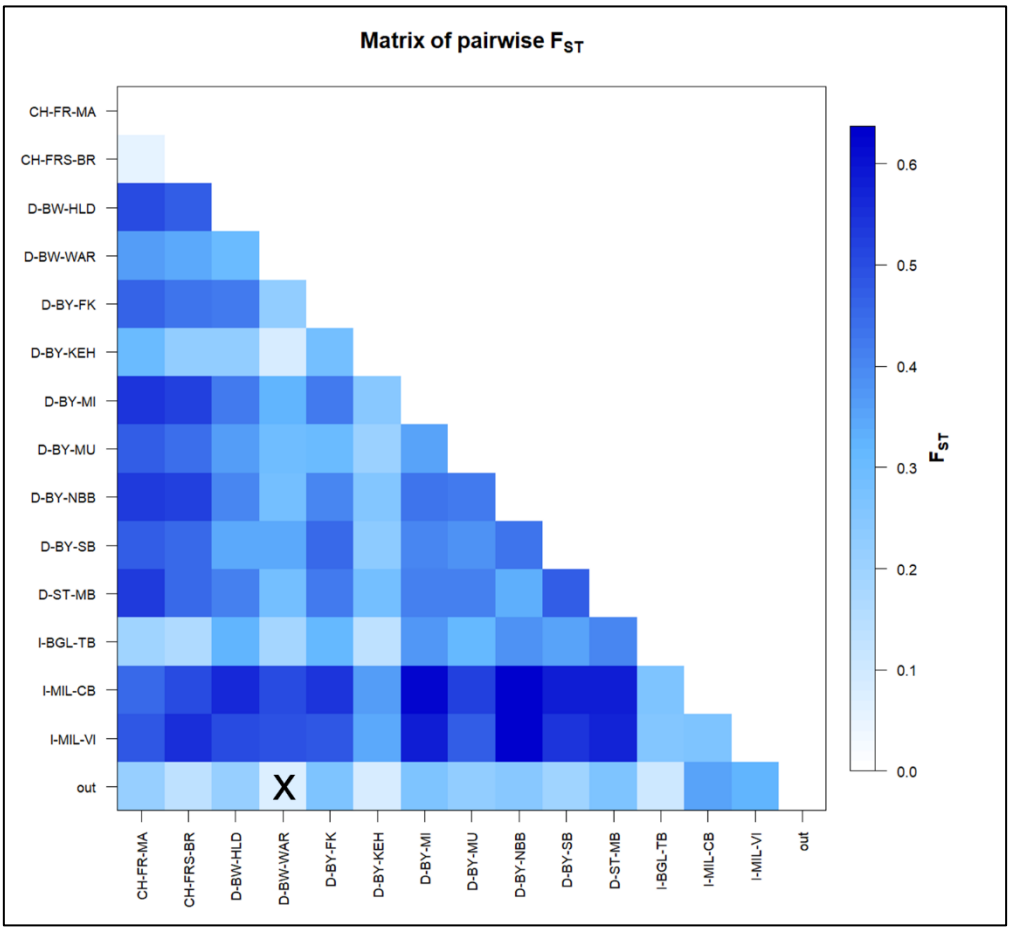


Figure 72. Population differentiation on genomic ALB SNP subset 170Indvf8g50m50 (32432 SNPs) conducted in Arlequin v. 3.5.
Pairwise F_{ST}-value among populations. Population labels were used as set in **Table 4**; x = not significant (significance level=0.05).

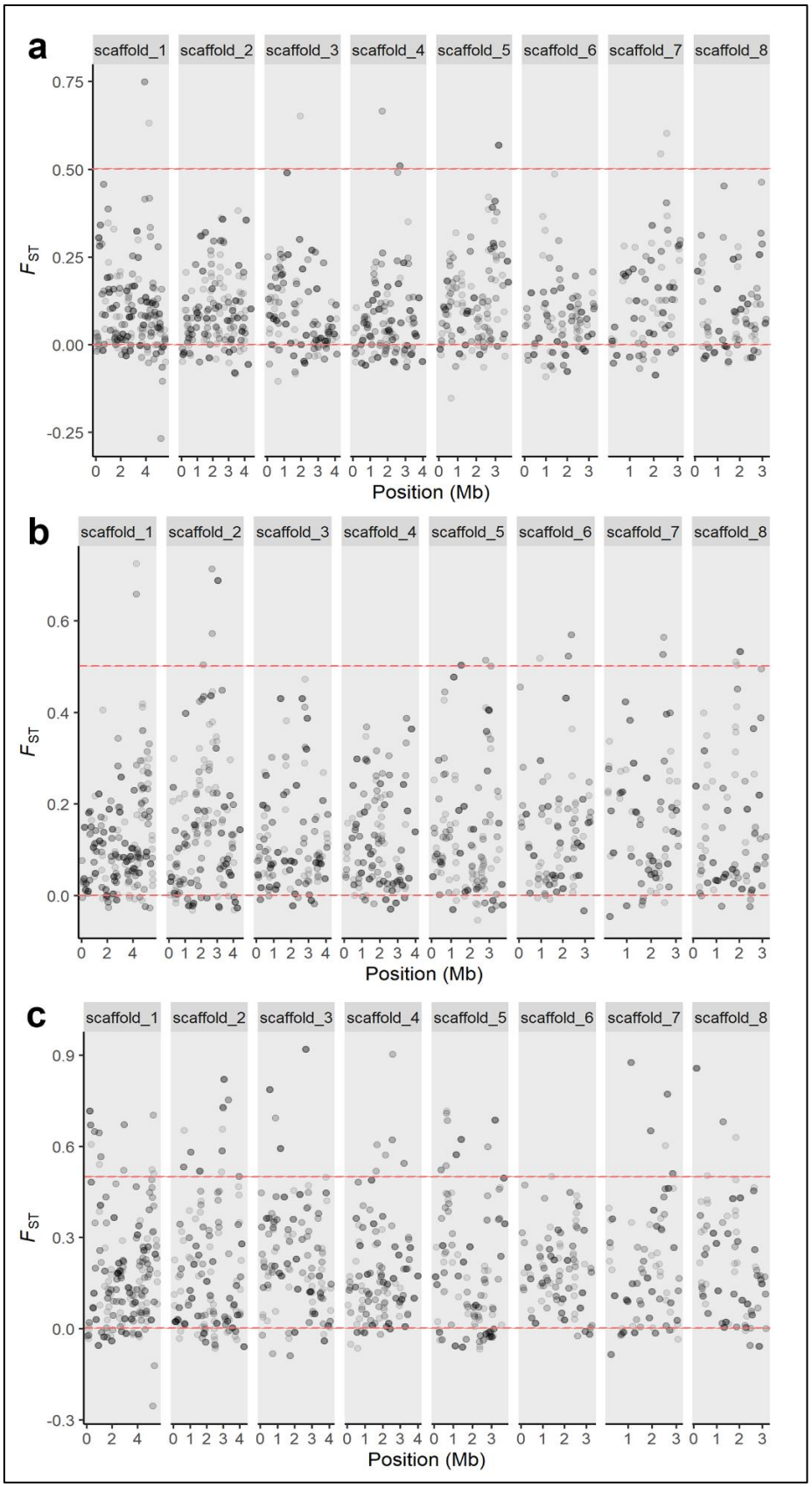


Figure 73. Pairwise F_{ST} estimates per 50 k windows with 10 k steps on ALB SNP subset 170Indvf8g50m50 between German states of ALB SNP data conducted in VCFtools v. 0.1.15.
German states Baden-Wuerttemberg (BW), Bavaria (BY) and Saxony-Anhalt (ST);
a) BW vs. BY **b)** BY vs. ST **c)** BW vs. ST

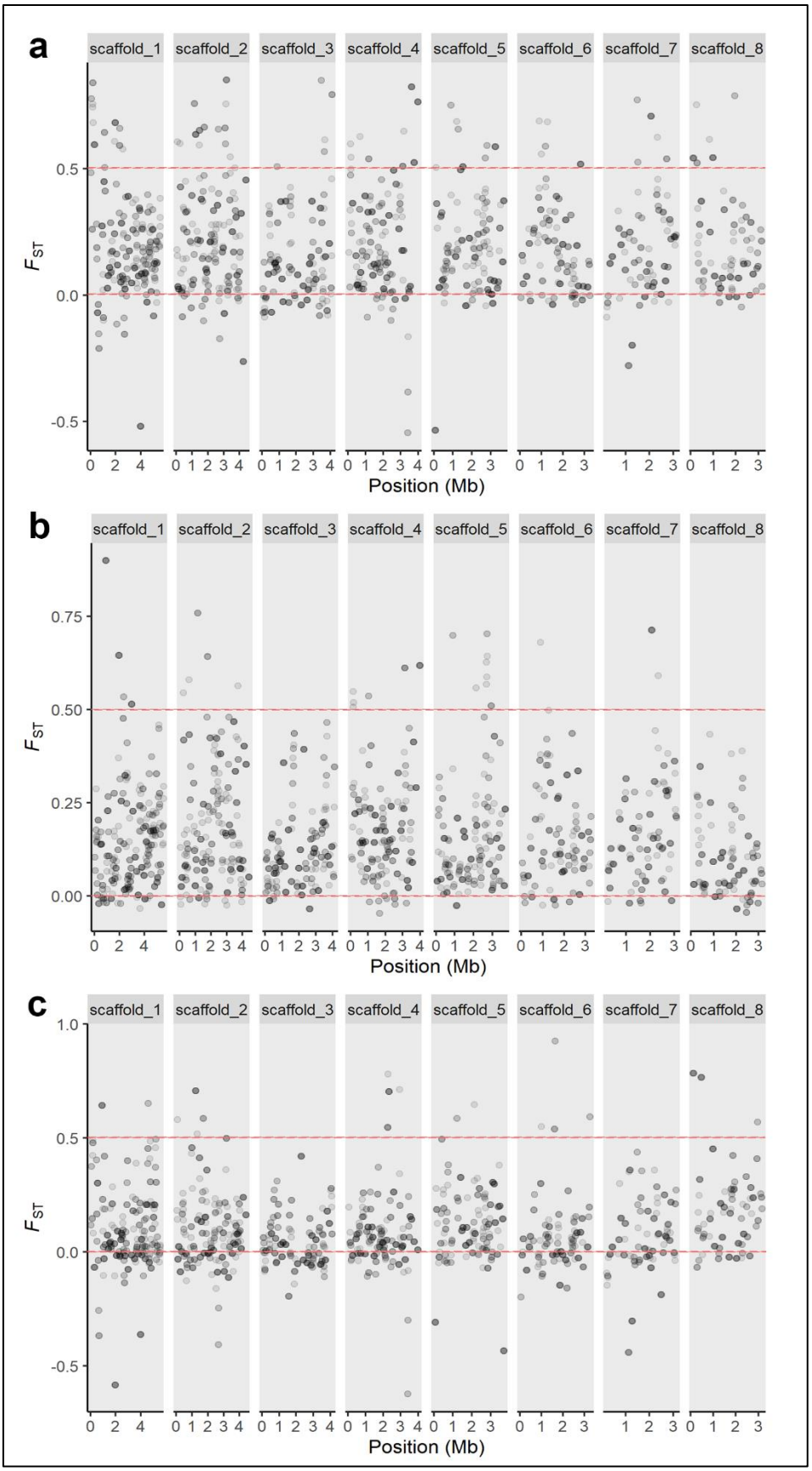


Figure 74. Pairwise F_{ST} estimates per 50 k windows with 10 k steps on ALB SNP subset 170Indvf8g50m50 between European countries of ALB SNP data conducted in VCFtools v. 0.1.15.
a) Germany vs. Switzerland **b)** Germany vs. Italy **c)** Italy vs. Switzerland

As another measure to investigate the population stratification, a principal component analysis (PCA) was performed on the LD-pruned 170Indvf8g50m50 dataset with 7810 independent SNPs. The PCA summarizes the main axes of variation in the allele frequencies on data without correlation to each other, consequently the LD-pruned data needed to be used here. The Eigenvalues obtained by this analysis were used to determine the principal components that explain most of the variation within the dataset. In **Figure 75** and **Table 44** (Appendix, IV Chapter C) the Eigenvalues are displayed for the first 22 PCs. The first three PCs together explain 37.799% of the variation of the data, since PC1 represents 17.513%, PC2 11.002% and PC3 9.284% of the variation.

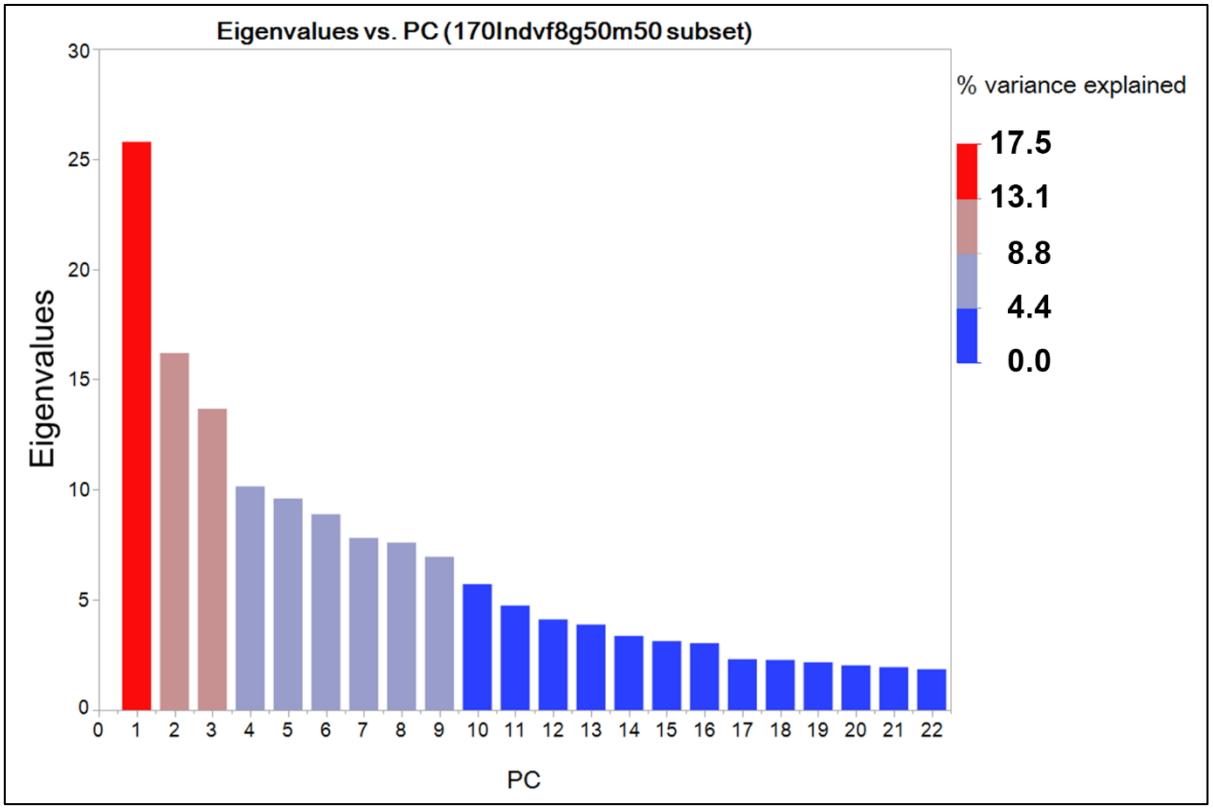


Figure 75. Eigenvalues from PCA analysis performed in PLINK v. 1.9 plotted to the first 22 principal components of the LD-pruned 170Indvf8g50m50 subset of genomic ALB SNPs.

By plotting of PCs to each other, the coordinates built along the axes of variation are shown. **Figure 76** and **Figure 77** illustrate the first three PCs explaining most of the variation of the LD-pruned ALB SNP dataset with plots of PC1 vs. PC2 and PC1 vs. PC3. Differently, **Figure 78** and **Figure 79** demonstrate PC4 (6.897%) and PC5 (6.518%), which combined explain 13.415% of the variation. Different to the PCA shown in 5.4.1 from the first evaluation of the dataset, here in **Figure 76** the Swiss and Italian clusters on the left are separated. Conspicuously, the separated four clusters on the left of the scatter plot (~0.0 on x-axis) contain specimens from different Pop-IDs and stick to these assignments in all other scatter plots between PC1 up to PC5 (**Figure 77**, **Figure 78**, **Figure 79**). One of the separated groups contain all specimens from the Swiss locations Marly and Bruensried, another one all specimens from Corbetta and Vittuone, Italy. Another clearly separated group there contains Italian specimens from Vaie, Trescore-Balneario, Cuneo and the two Chinese specimen from Harbin and Tongliao. The two specimens from Ostra (Italy) were separated with the single specimen from Fermo (Italy), while Jinan (China) stands for itself nearby. In the upper right part of the plot, two clusters from Magdeburg stand out, which also stayed separated in all the following scatter plots. On the bottom right side of the plot Feldkirchen forms a cluster with all its specimens. None of the previously described populations assigned to a group showed specimens somewhere else, except Cuneo. In the coordinates aggregated around zero-point, all other populations were represented, but barely separated. Slightly on the side are specimens from Neubiberg (D) and the single specimen from Neukirchen am Inn (D).

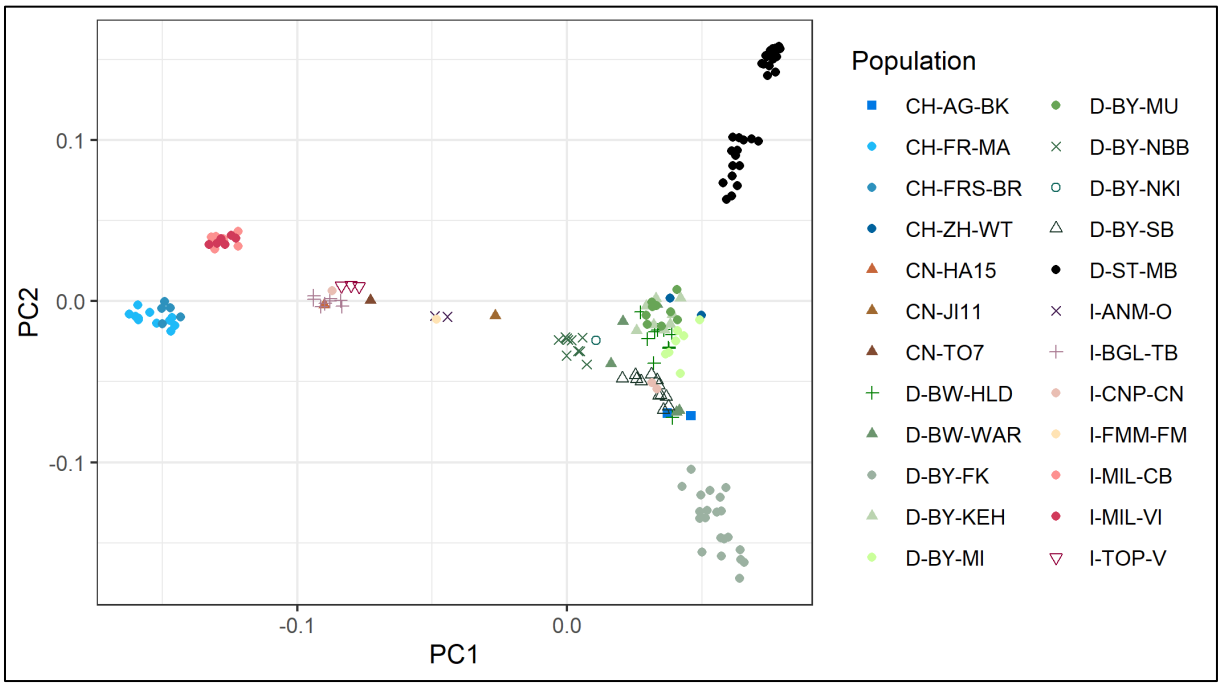


Figure 76. Scatter plot of principal components PC1 (17.513%) and PC2 (11.002%) on the LD-pruned subset 170Indvf8g50m50 of genomic ALB SNP data.

In **Figure 77** it is observable that PC3 separated all specimens from Schoenebach (including the internal controls). The specimens from Neubiberg together with Neukirchen am Inn are more clearly separated with PC3 than with PC2. With PC3 two specimens from Cuneo, two from Weil am Rhein (D), two from Berikon (CH) and one specimen from Hildrizhausen (D) are separated out together. For the Pop-ID D-BW-WAR the other samples are scattered and not grouped together. The remaining Hildrizhausen specimens are in another group with different Pop-IDs (D-BY-KEH, D-BY-MI, D-BY-MU, CH-ZH-WT) and does not separate by geographic location.

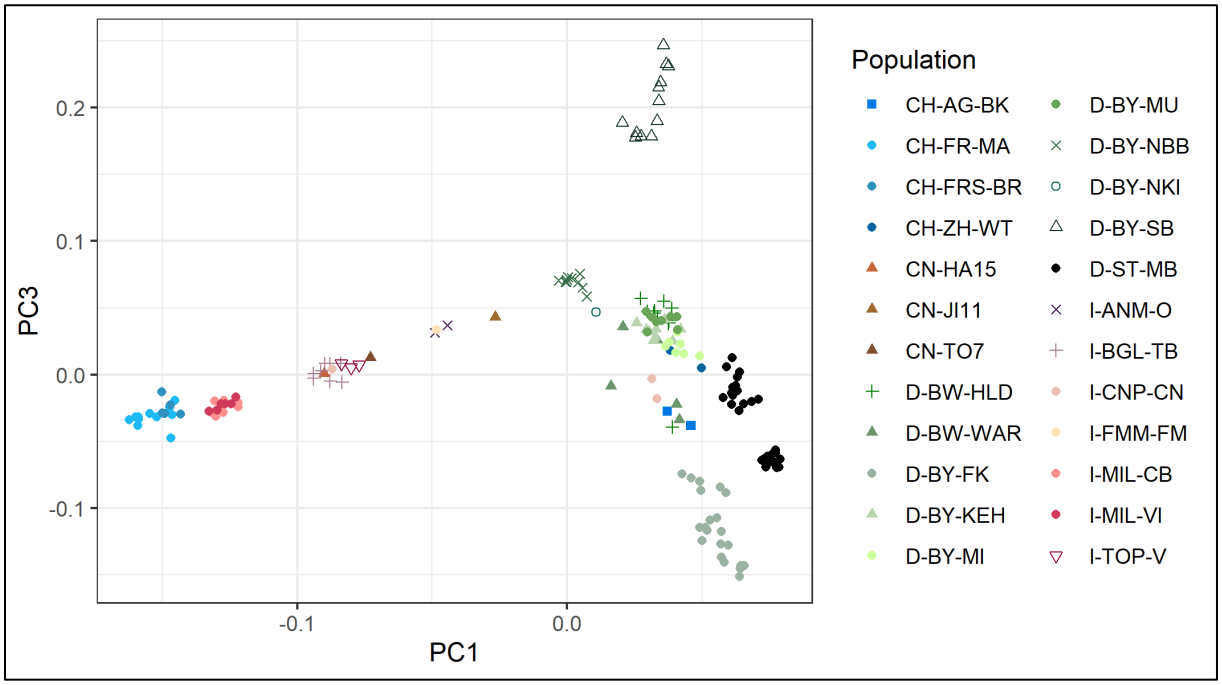


Figure 77. Scatter plot of principal components PC1 (17.513%) and PC3 (9.284%) on the LD-pruned subset 170Indvf8g50m50 of genomic ALB SNP data.

When plotting PC1 with PC4 (**Figure 78**), the locations Neubiberg and Neukirchen am Inn became separated. The previously not separated group remained not clearly divided, but the specimens from D-BW-HLD (except one specimen), D-BY-MI, D-BY-MU, D-BY-KEH, CH-AG-BK, CH-ZH-WT and I-CNP-CN are quite close to each other, respectively. Only the specimens from D-BW-WAR still displayed scattered coordinates.

With the plot of PC1 vs. PC5 (**Figure 79**) no more findings could be perceived, except that the specimens from D-BY-MI slightly stand out from the undivided group around the zero-point.

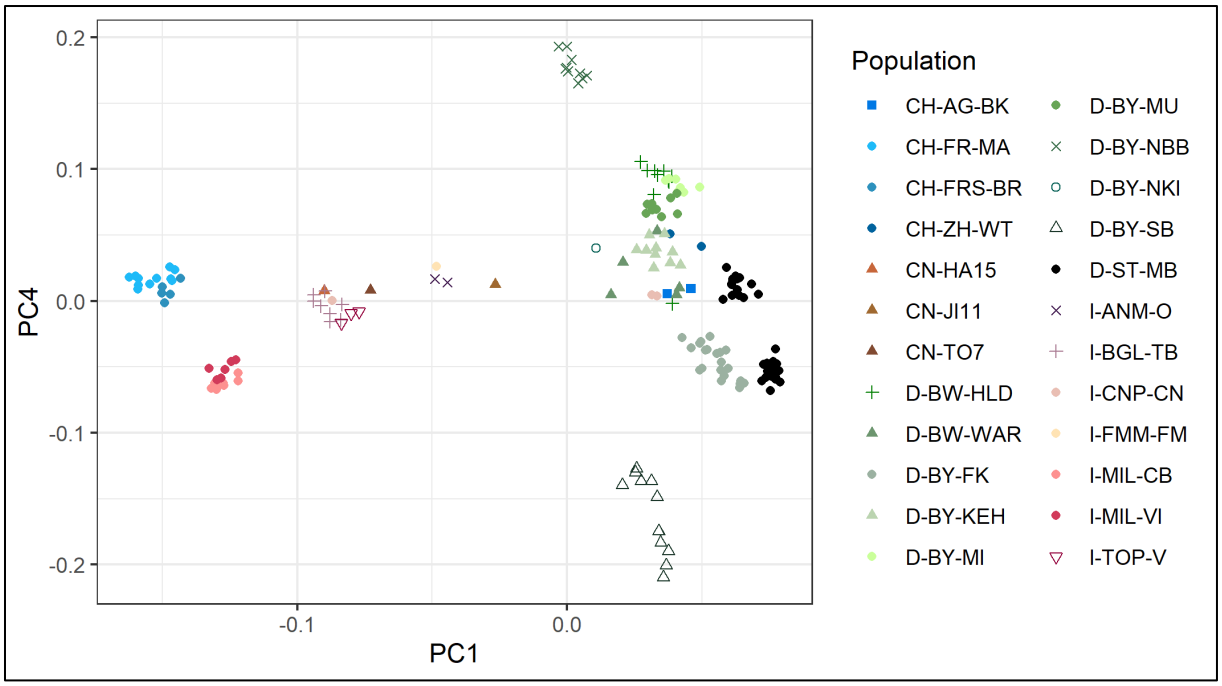


Figure 78. Scatter plot of principal components PC1 (17.513%) and PC4 (6.897%) on the LD-pruned subset 170Indvf8g50m50 of genomic ALB SNP data.

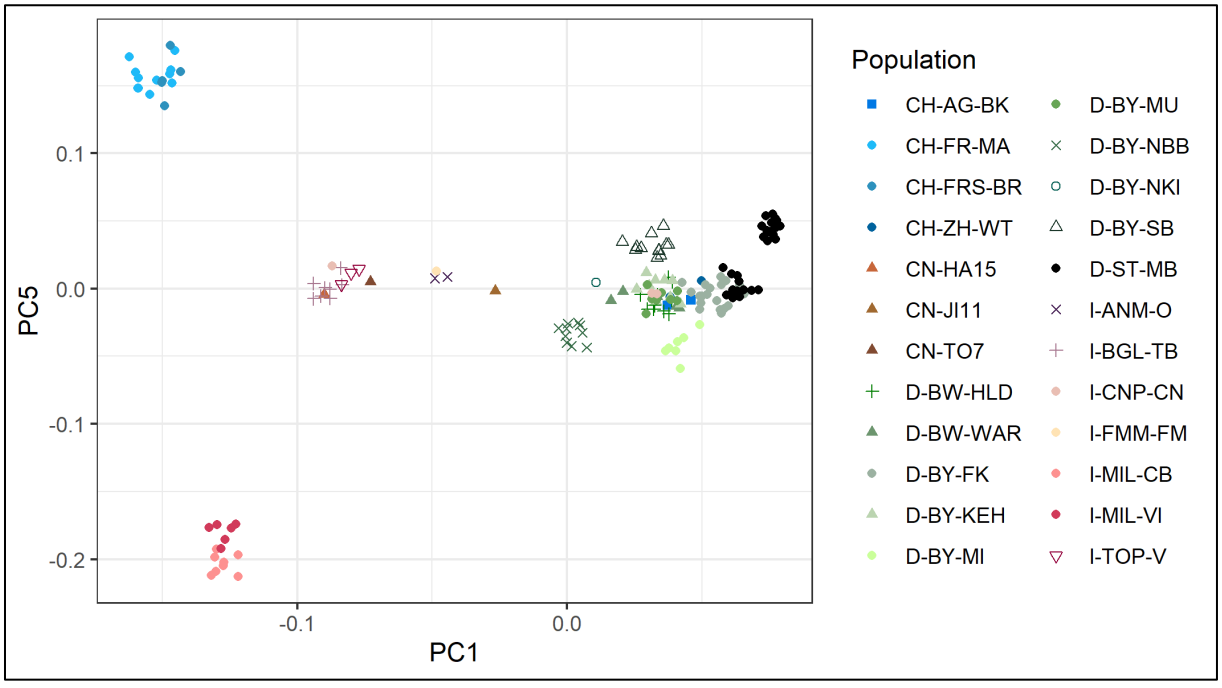


Figure 79. Scatter plot of principal components PC1 (17.513%) and PC5 (6.518%) on the LD-pruned subset 170Indvf8g50m50 of genomic ALB SNP data.

5.4.3 Ancestry of European invasive ALB populations

To determine the individual kinship, an Admixture analysis was performed with a dataset of 7810 independent SNPs of the LD-pruned subset 170Indvf8g50m50. While conducting the Admixture analyses using the clustering software ADMIXTURE v. 1.3.0 the cross-validation procedure was used to help choose the right value for K, which exhibits a low cross-validation error compared to other K values. In **Figure 80** it stands out, that the lowest cross-validation error was estimated for K = 9. The 22 potential subpopulations (K) were tested since there were 21 European locations and samples from China involved in this analysis.

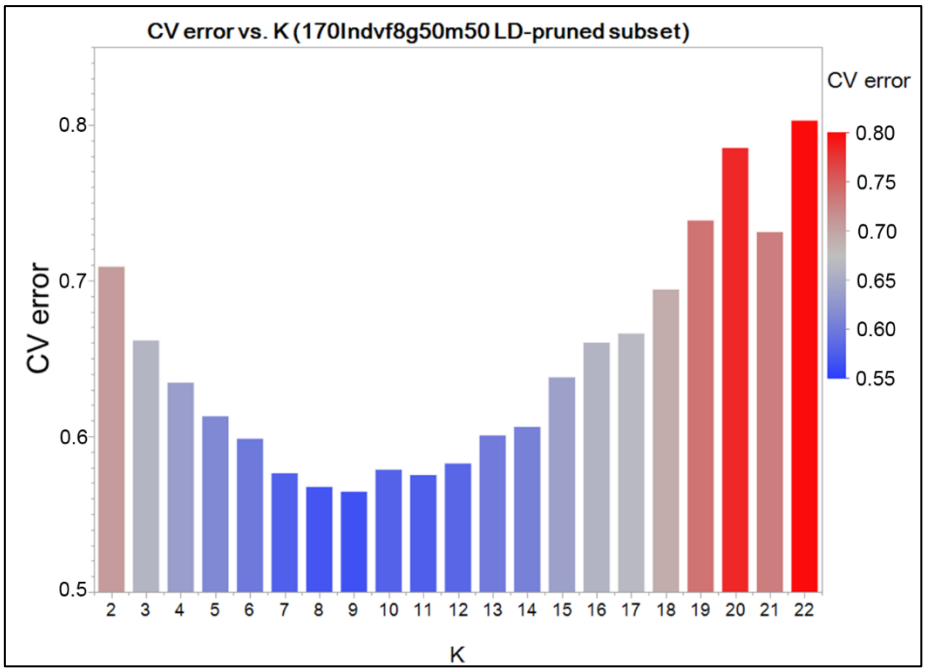


Figure 80. Cross-validation errors to choose the right K-value on the ancestry in the LD-pruned 170Indvf8g50m50 ALB SNP data subset conducted in ADMIXTURE v. 1.3.0.

The results of nine putative subpopulations (k = 9) is depicted in **Figure 81**, while **Figure 82** displays K = 2-11 from the same dataset. It is clearly observable, that most of the individuals within a collection site are assigned to the same subpopulation. In the case of the sites with close proximity to each other, like Marly and Bruensried, as well as Corbetta and Vittuone, these are homogeneous as well. For Murnau (D-BY-MU) and Miesbach (D-BY-MI) this was the case as well for K = 9, despite the fact that they are geographically about 60 km apart from each other. With K = 7 and K = 8 subpopulations Miesbach and Murnau showed different subpopulation patterns, as well as with K = 10 and K = 11 (**Figure 82**). Interestingly, the specimens from Magdeburg (D-ST-MB) are clearly separated in two subpopulations and one mixed one from specimens collected in 2015. Other locations showed a mixture of subpopulations represented in different compositions as well: Berikon (CH-AG-BK), Winterthur (CH-ZH-WT), all three Chinese specimens, one single specimen from Hildrizhausen (D-BW-

HLD), Weil am Rhein (D-BW-WAR), Kelheim (D-BY-KEH), Neukirchen am Inn (D-BY-NKI), Magdeburg-2015 (D-ST-MB), Ostra (I-ANM-O), Trescore-Balneario (I-BGL-TB), Cuneo (I-CNP-CN), Fermo (I-FMM-FM) and Vai (I-TOP-V). Smaller portions of another subpopulation represented additional to a main one were found in Feldkirchen (D-BY-FK), Miesbach (D-BY-MI) and in some Magdeburg-2019 specimens (D-ST-MB). The mixed populations from Switzerland and Germany are similarly composed, except the mixed ones from Magdeburg, which showed a composure which occurred only there. The mixed populations from Italy are also composed similarly and each one contains all a large subpopulation part which can also be found in Marly and Bruensried in Switzerland.

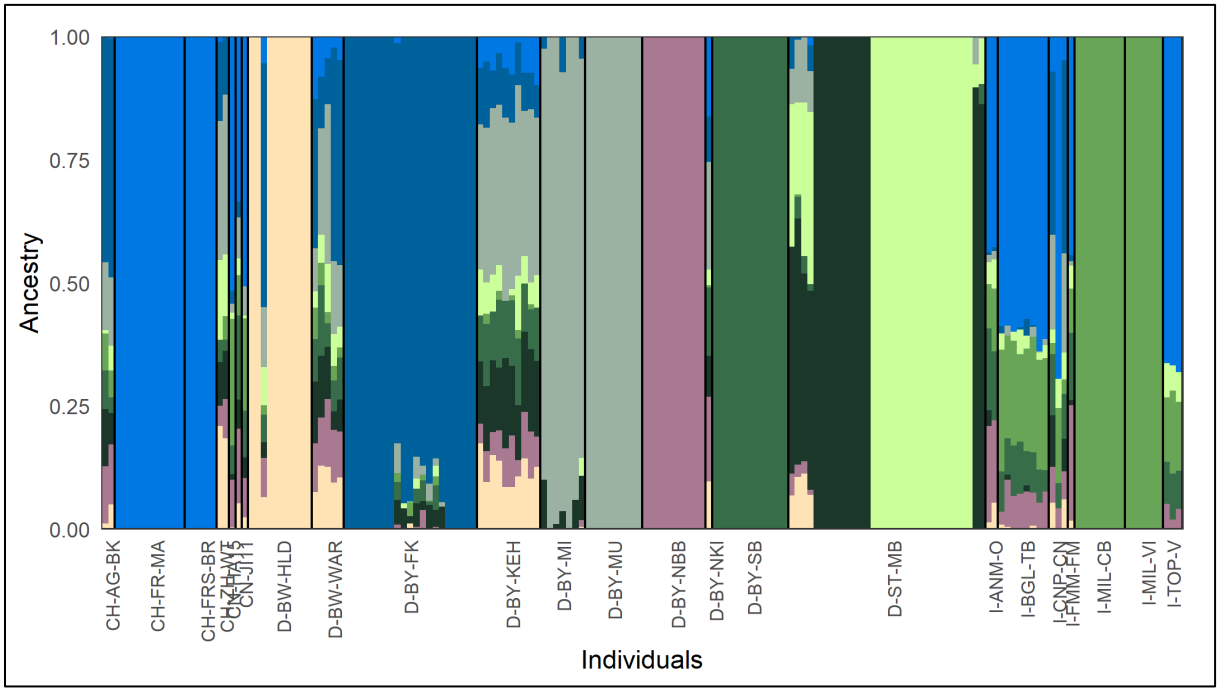


Figure 81. Analysis of Ancestry of 170 ALB samples from Switzerland, China, Germany, and Italy conducted in ADMIXTURE v. 1.3.0; K = 9; 7810 SNPs.
The LD-pruned genomic ALB SNP subset 170Indvf8g50m50 was analysed here. Vertical bars represent each specimen, while coloured segments represent the proportion of ancestry to the different subpopulations.

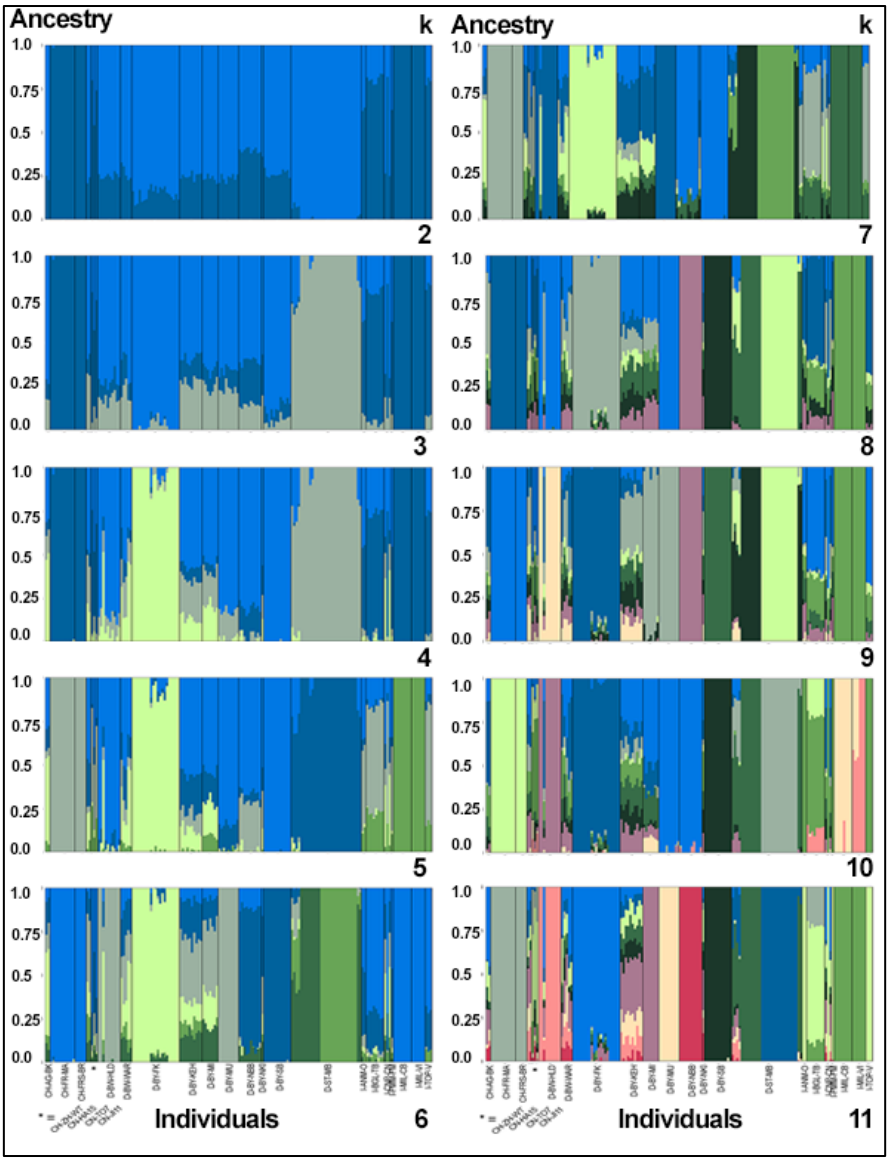


Figure 82. Analysis of Ancestry of 170 ALB samples from Switzerland, China, Germany and Italy conducted in ADMIXTURE v. 1.3.0; sequential K = 2 to K = 9; 7810 SNPs.
The LD-pruned ALB SNP subset 170Indvf8g50m50 was analysed here.

To check if the ancestry is affected by filtering on the allele frequency, another subset was generated where only sites with a minor allele frequency of at least 0.1 are included, which means the rare SNPs were sorted out. The LD-pruned subset of 170Indvf10g50m50 included 6064 SNPs to determine the individual kinship. In this case also a cross-validation procedure was conducted along the way, resulting also in K = 9 as the right K-value to choose (**Figure 83**). The only striking difference in the ancestry from the LD-pruned 170Indvf10g50m50 subset in comparison to the previous one, is visible for the specimens from Miesbach (D-BY-MI). In this subset, Miesbach specimens do not share the subpopulation with Murnau (D-BY-MU), but they show the same composition as the specimens from Kelheim (D-BY-KEH). The only cluster found for Murnau is also represented in a portion of "mixed" Kelheim and Miesbach.

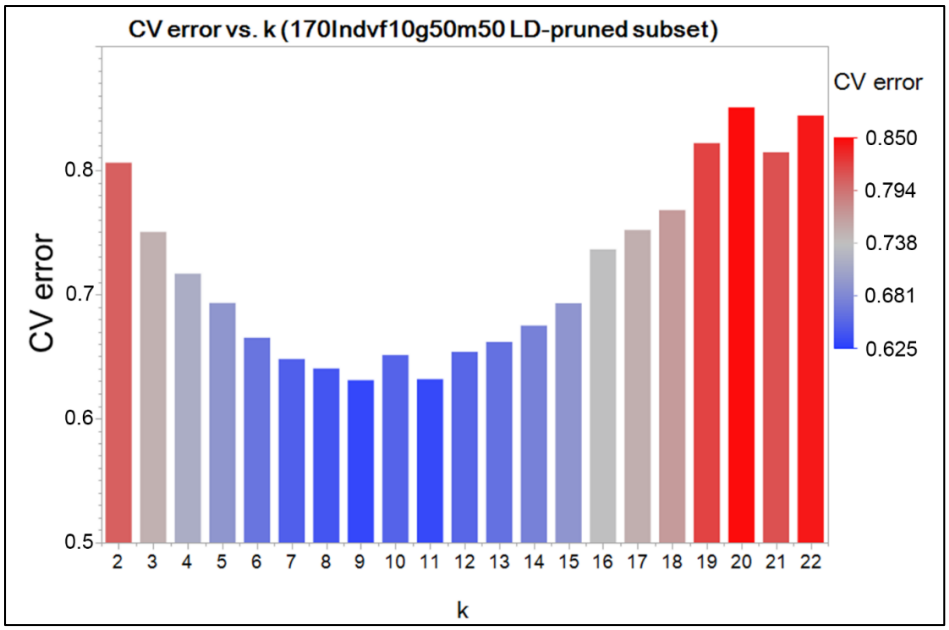


Figure 83. Cross-validation errors to choose the right K-value on the ancestry in the LD-pruned 170Indvf10g50m50 ALB SNP data subset conducted in ADMIXTURE v. 1.3.0.

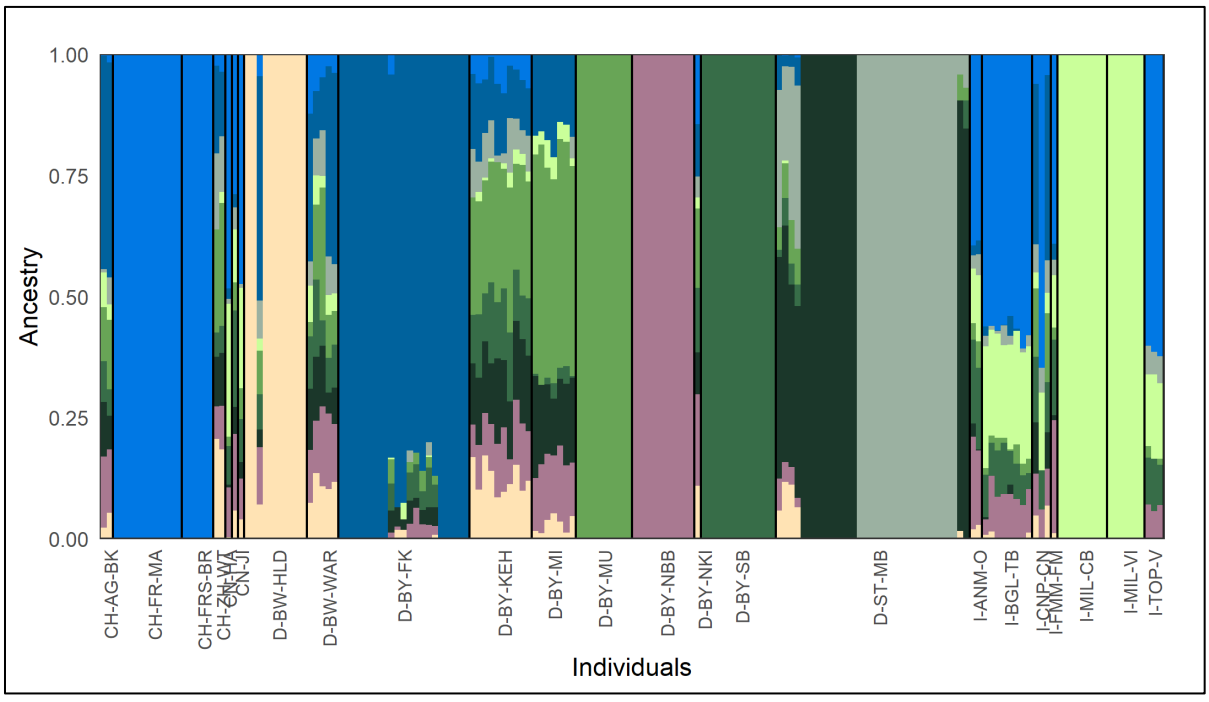


Figure 84. Analysis of Ancestry of 170 ALB samples from Switzerland, China, Germany, and Italy conducted in ADMIXTURE v. 1.3.0; K = 9; 6064 SNPs.
The LD-pruned genomic ALB SNP subset 170Indvf10g50m50 was analysed here.

In Appendix, IV Chapter C, **Figure 93** and **Figure 94**, the cross-validation errors and the Admixture plot of the LD-pruned 178Indvg50m50 subset are depicted for K=10 and 7673 SNPs. This plot shows the same results as **Figure 81**, but validates that the D-BY-18-110 specimen found in a pheromone trap is assigned to the same subpopulation as the other

D-BY-SB specimens from 2015. With the NeighborNet phylogenetic network (**Figure 85**) drawn with Splitstree v. 4.17.1 it is possible to show even more effectively evolutionary relationships due to the crosslinks. This is strongly emphasized, for instance, by the separation in three groups in Magdeburg (D-ST-MB) which all show the same origin with cross-links at a p-distance of approx. 0.1 nucleotide substitutions per site. Similar separation, but with shorter branch length, can be observed for Feldkirchen (D-BY-FK) in two different groups, which both share crosslinks at the same origin at a longer distance together with single specimens from Hildrizhausen (D-BW-HLD), Weil am Rhein (D-BW-WAR), Berikon (CH-AG-BK) and Cuneo (I-CNP-CN). Especially some single specimens are also very striking with their multiple occurrences at different positions in the phylogenetic network like from Hildrizhausen, Weil am Rhein and Cuneo. The very clear separation of most of the Italian findings together with the findings of the Swiss locations Marly (CH-FR-MA) and Bruensried (CH-FRS-BR) are also standing out here. They all share cross-links in the directions of the root together with the three Chinese specimens, but all of them show a clear separation at certain points with a distance of a little bit less than 0.1 nucleotide substitutions per site. The three specimens from China are all standing alone in this network, but are in general genetically closer to subpopulations found in Italy and closest to the findings from Trescore-Balneario (I-BGL-TB). The neighbouring sites Marly and Bruensried, as well as Corbetta and Vittuone again group together in this network, like already shown in the Admixture plots, but for both pairs, a tiny separation in two groups is also apparent. Also, the two specimens from Ostra (I-ANM-O) and Fermo (I-FMM-FM) share the same branch with some cross-links. One specimen from Hildrizhausen and two from Weil am Rhein have different branches, but share cross-links at the beginning. All other specimens clustered together with the other specimens from their own collection site with tight cross-links. An exception on this are the specimens from Kelheim (D-BY-KEH) where the specimens have an own branch respectively that roots together with all the others, about 0.1 branch length apart. Similar observation was made for Trescore-Balneario, but with less genetic divergence compared to Kelheim.

In general, all the subpopulations from Germany, Winterthur (CH-ZH-WT), Berikon (CH-AG-BK) and 2 specimens from Cuneo root in the middle ~0.1-0.2 branch length apart from each other. The branch between this main root toward the middle of the most Italian, Chinese and the two Swiss locations is also at a distance of approx. 0.05-0.1 p-distance. Thus, Marly and Bruensried are the most genetic divergent sites compared to the root of the German and other Swiss subpopulations with approx. 0.2 nucleotide substitutions per site.

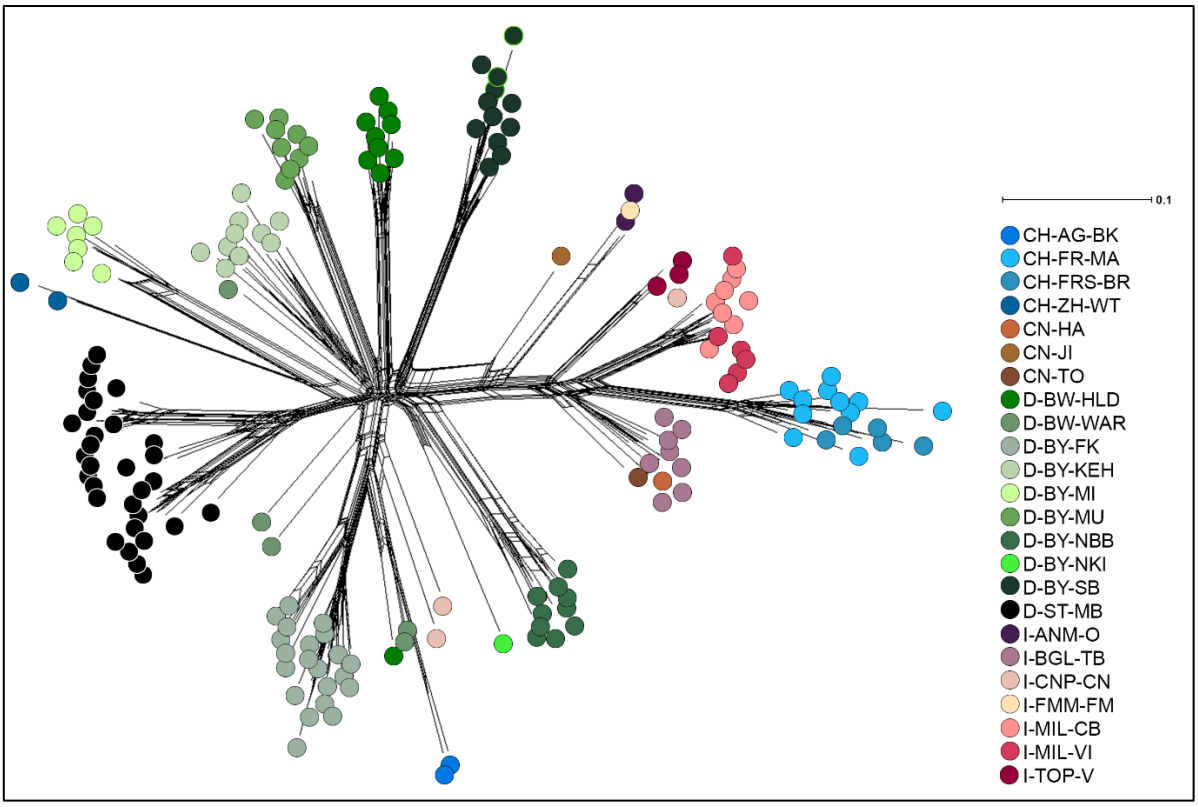


Figure 85. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on genomic ALB SNP data subset 170f8g50m50 from Switzerland, China, Germany and Italy. The ɣ-distributed (+G 3.73) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 588 splits; Fit = 96.515; 170f8g50m50, 170 taxa and 32432 SNPs are shown. The three light green framed specimens from D-BY-SB represent the sample D-BY-SB-047 and the two internal controls.

In Appendix, IV Chapter C, **Figure 95** the same phylogenetic network is illustrated with the results of the validation with 1000 runs of bootstrap method. Also, in the appendix in **Figure 96** the phylogenetic network for the 178Indvg50m50 subset can be found, validating that D-BY-SB-110 specimen shares the same branch than all others from Schoenebach. For all three sampled countries its own network was drawn always together with the three Chinese samples, which were validated by 1000 bootstrapping runs. The networks for Switzerland (**Figure 86**), Italy (**Figure 87**) and Germany (**Figure 88**) confirm the observations from the network with all specimen, but show in more detail the connections of the specimens to each other, especially in more detail for locations like Kelheim, Trescore-Balneario as well as the neighbouring sites Marly/Bruensried and Vittuone/Corbetta.

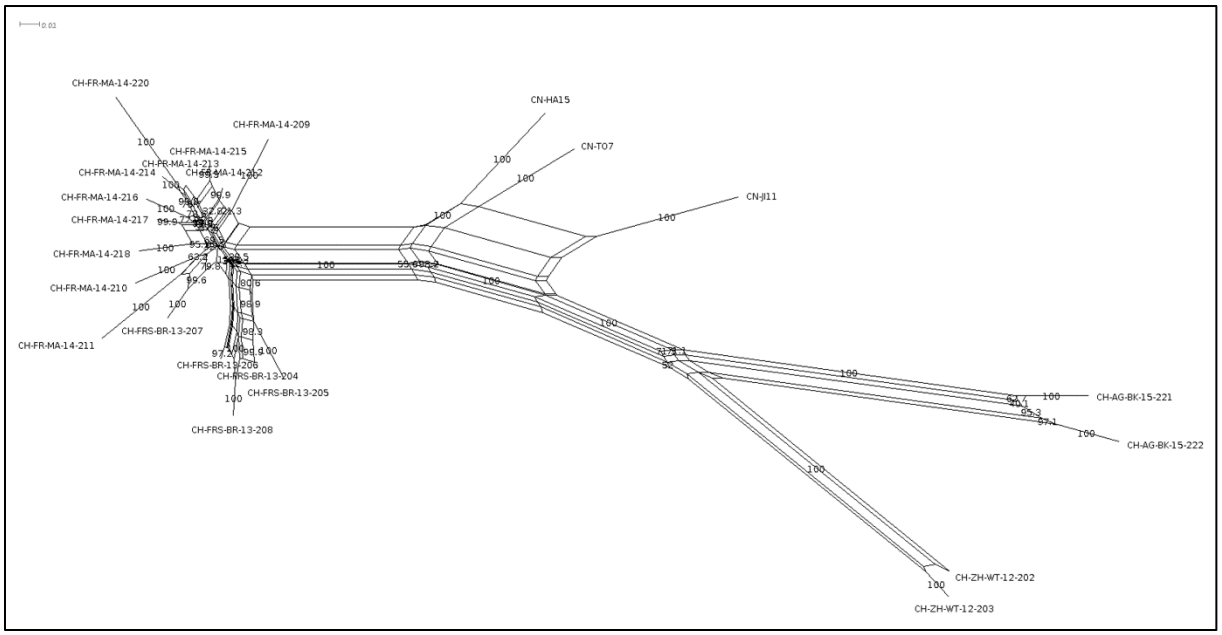


Figure 86. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on genomic ALB SNP data subset 170f8g50m50 from China and Switzerland.
The ɣ-distributed (+G 3.73) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 588 splits; Fit = 96.515; 1000 runs Bootstrap; 23 (of 170) taxa (Switzerland, China) and 32432 SNPs are shown.

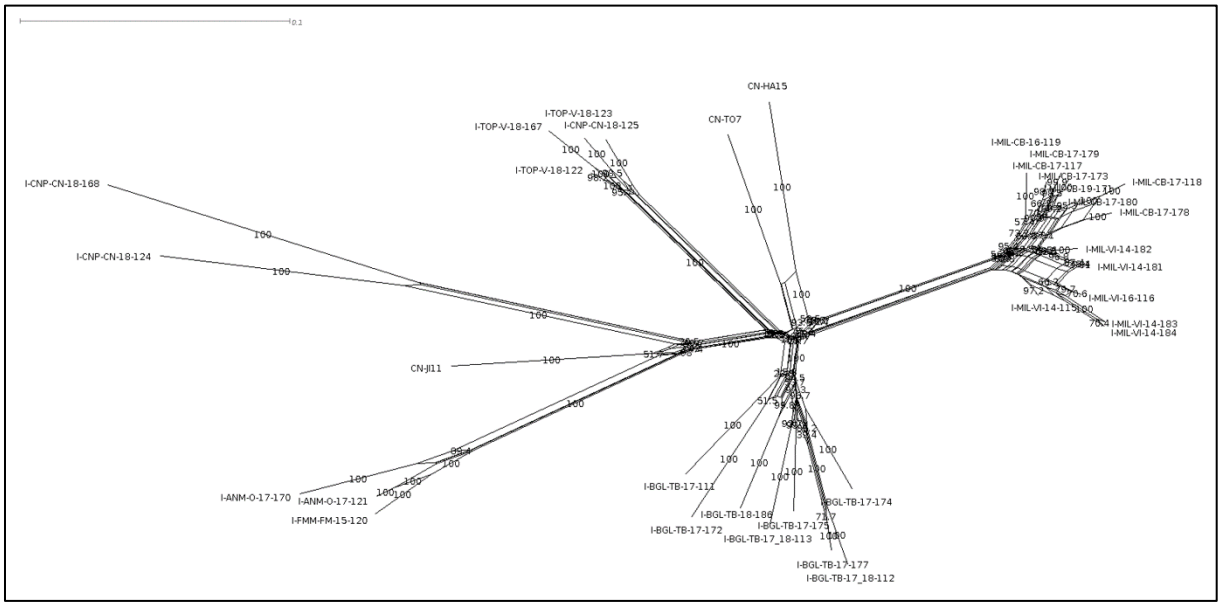


Figure 87. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on genomic ALB SNP data subset 170f8g50m50 from China and Italy.
The ɣ-distributed (+G 3.73) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 588 splits; Fit = 96.515; 1000 runs Bootstrap; 34 (of 170) taxa (Italy, China) and 32432 SNPs are shown.

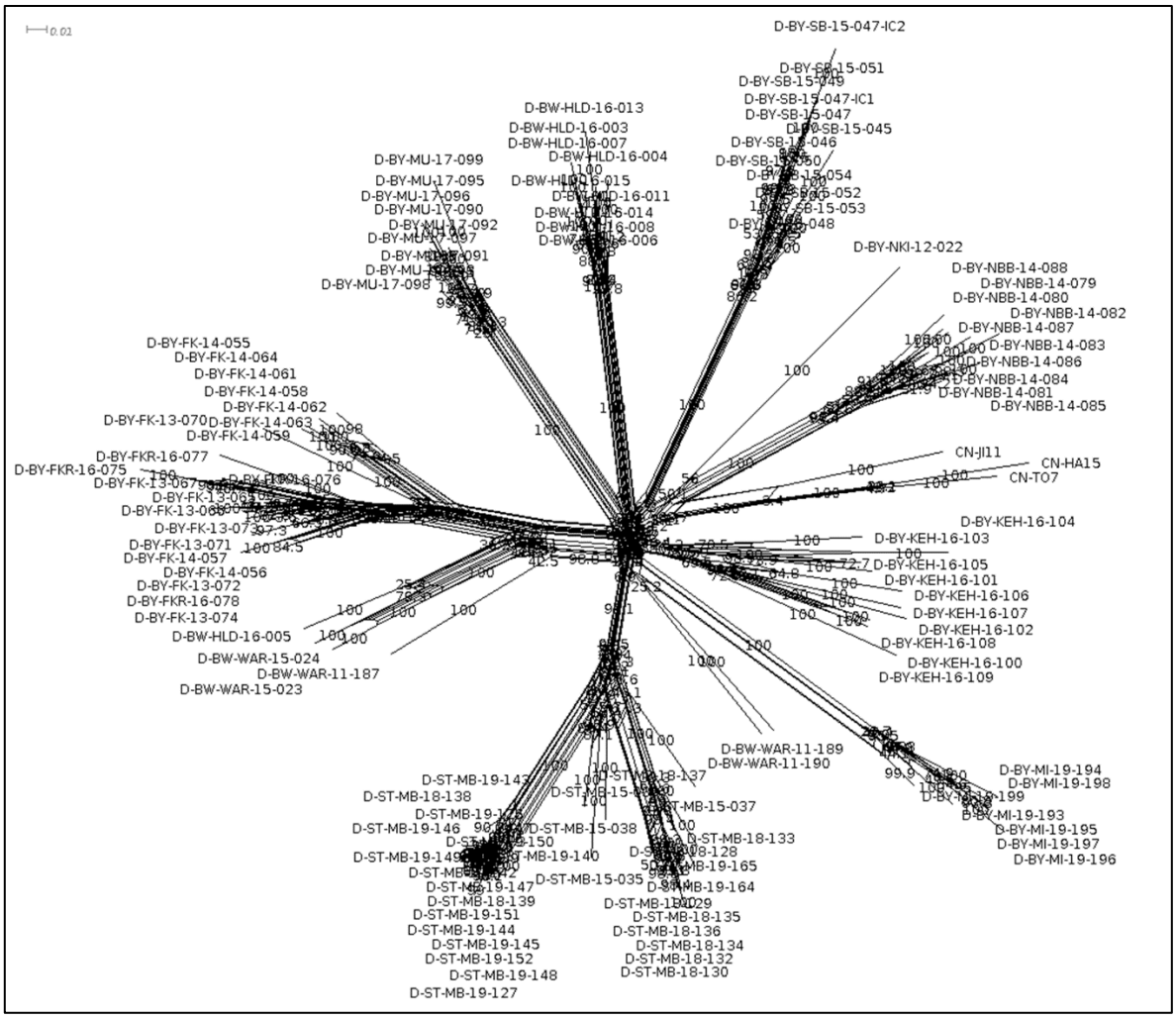


Figure 88. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on genomic ALB SNP data subset 170f8g50m50 from China and Germany.
The ɣ-distributed (+G 3.73) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 588 splits; Fit = 96.515; 1000 runs Bootstrap; 119 (of 170) taxa (Germany, China) and 32432 SNPs are shown.

6 Discussion

6.1 Asian long-horned beetle sampling and DNA sequencing

Sample size and quality imbalance

Since many of the past infestations in Germany, Switzerland and Italy were already eradicated or under strict eradication (3.2, p. 26), for many of the collection sites no sufficient sampling size could be aquired. Some samples of past outbreaks were either not in storage anymore or already depleted from previous studies. This led to a huge imbalance in sampling size of the different collection sites (**Figure 12**, **Figure 13**). Also, the many different sampling sources and practice backgrounds led to an unstandardized assemblance of differently stored ALB samples. In case of sudden new detected infestations, it is obvious, that a quick specimen identification via morphological and DNA-barcoding methods is the next step of action to ensure appropriate quick management responses (Boykin et al. 2012; Darling and Blum 2007; Wu et al. 2017). By this means, the storing conditions over longer periods of times were sometimes unsuitable for DNA based methods in general or for DNA-based approaches beyond DNA-barcoding. With the rising abilities in genomic methods and tools, the integration of further genetic analyses which require excellent DNA quality should be considered in future phytosanitary praxis, since knowledge on this level can facilitate more sustainable, less time and cost consuming courses of action (Chown et al. 2015; Roe et al. 2019). Hence, DNA-based methods beside morphology-based reference sampling is the most suitable practice for all research approaches (e.g. Wu et al. 2017).

Limitations of sequencing by sampling bias

The sampling bias led to a loss of SNPs in GBS approach by different genome coverages and thereby higher overall missingness. Owing to the reduced DNA quality, 15 samples were sorted out from the SNP dataset (170f8g50m50 and 170f10g50m50 subsets) and the remaining ones still struggled from high missingness because of poor DNA quality. In the case of single PCR-product sequencing of COI and NUMTs, some specimens remained without PCR product owed to the reduced DNA quality, but in case of NUMTs it could not be assured if the lack of a PCR product was caused by DNA quality or the absence of the NUMT sequence. Among all excluded samples, the most were dried and pinned adult beetles or otherwise dried and fragmented specimens with either a high portion or solely degraded DNA. With this background, *de novo* GBS would have been impossible due to the sampling quality. Additionally, the sample size bias reduced the possibility of analyses, e.g. some sites had to be exluded in pairwise comparisons of genetic distance. In contrast to the other two approaches, with GBS most specimens and populations could be represented. This is a good sign for the utilization of GBS on environmental or non-optimal specimens and resulted in

general in very high-quality SNPs (phred scores). To evaluate the batch influences by the three different GBS runs, some general quality and population genetic outputs were examined on raw data in relation to the three runs and collection sites. The observations of the plotted outputs of the raw data and different missingness levels on the dataset with 170 individuals (170f8g50m50) together led to the assumption of no or neglectable batch effects, since the results were merely affected by the collection site but not from the library. All observations hint to 170f8g50m50 as the best filter combination to prepare the GBS data for analyses.

Limitations of sequencing by mismatching primers

Beside the issues of lower DNA quality, the amplification of PCR products also revealed some interferences in the current practice. The mitochondrial COI DNA-marker amplified with the universal arthropods primers of Folmer et al. (1994) does not seem to be a decent DNA-marker for ALB, when considering the results of the Primer BLAST (**Table 10**) as well as the prescence of multiple NUMTs (5.3.2, **Figure 46**). The Primer-BLAST revealed an alleged NUMT region which might be coamplified with a similar product length (702 bp). Hence, the unawareness of NUMTs which are not avoidable in the classical strategies as figured by Song et al. (2008) might have contaminated previous studies using COI and thereby also the databases.

6.2 Chapter A discussion: Prospects and limits of COI-barcoding

Chapter A

"Prospects and limits of COI-barcoding practice to characterize the biological invasions of Asian long-horned beetle, *Anoplophora glabripennis*"

As mtCOI of *A. glabripennis* was often object of research and plant health practice in specimen identification, population genetic and phylogenetic studies (3.2), the comparability with many other herein unsampled invasive and native populations was taken advantage from. Additionally, mtDNA COI-marker had enabled also valuable insights into other IAS before (e.g. Scheffer and Grissell 2003), is still commonly used in genetic characterizations of invasive insects (e.g. Fernández et al. 2022; Doorenweerd et al. 2020) and among the most widely used markers for species identification and phylogenies (Mandal et al. 2014).

COI haplotypes and heteroplasmy in the European invasive range

The eight haplotypes of COI found in Europe and the three sites in China were all identified as ALB using BLAST. Most haplotypes were merely few mutations apart from each other which was also the observation in previous ALB studies that used COI barcoding (Carter et al. 2010; Javal et al. 2019a; Tsykun et al. 2019). When examining the haplotype projections, for most locations there is no specific structure linked except COI_Type2 (Magdeburg), COI_Type4 (Feldkirchen), and locations with immediate vicinity in Europe like COI_Type5 (Marly,

Bruensried) and COI_Type6 (Vittuone, Corbetta). In Feldkirchen (Germany), 18 out of 20 specimens exhibited an ambiguous nucleotide. Differently to other studies (Javal et al. 2019a; Lee et al. 2020), both haplotypes were used and checked by translating them into amino acid sequences (Song et al. 2008). Since both did not show any frame shifts, as all other haplotypes as well, the rare presence of heteroplamy was assumed. Several publications have reported the occurrence of heteroplasmy in natural populations, as for instance in *Drosophila melanogaster* (Nunes et al. 2013). It is possible that heteroplasmy is less reported than its natural occurrence, since PCR based methods might favor the amplification of the mtDNA version with higher frequency. This could be solved by HTS of the mitochondrial genomes from this population, since thereby also low frequency haplotypes could be detected and the efforts compared to cloning are minimized. (Ladoukakis and Zouros 2017) Due to the higher frequency detected in the dataset, maternal inherited heteroplasmy seems most likely because the results proposed here suggest stable heritability. Even the double peaks in the trace data during direct PCR product sequencing were the same height. But, in general the causes for the heteroplasmy in Feldkirchen cannot be confirmed with the current dataset. A possible susceptibility to NUMT co-amplification will be discussed in 6.5.

Genetic structure of COI in the European invasive range

For all mtCOI-658 bp sequences, exceptionally high molecular variance between the subpopulations (>90%) were measured, while the variance was lowest within the populations (**Table 12**). The remaining genetic variance of ~10% within subpopulations is not that unusual, especially not when considering the possibility of multiple introductions from different sites of origin to one location (comparable with Javal et al. 2019a), like probably in the few populations that hold two haplotypes. Hence, mating between different demes within the countries (F_{SC} >0.9) is extremly unlikely or gets lost among the high differentiation between the other demes (F_{ST} >0.9). The genetic exchange between the European countries (F_{CT} ~0.3) is also unlikely and suggestes different introduction sources in the most cases. These results point to a pattern of multiple independent introductions within the countries and no or merely a few occurrences of secondary dispersal events as also reported in former European ALB studies (Javal et al. 2019a; Tsykun et al. 2019). The European COI dataset shows population structure but no geographic pattern, especially the shared common haplotypes diminish the possibility to infer. But overall, pairwise differentiation between the populations with significant F_{ST}-values were very high (>0.8). The divergent maternal lineages indicate probably independent introduction events of the COI-haplotypes COI_Type2, Type3, Type4, Type5, and Type6. But when taking the results of Carter et al. (2010) and Javal et al. (2019a) into account, the introduction source cannot be determined solely from the maternal lineage of COI, since the native population only show a blurred geographic pattern, often with multiple COI-haplotypes per site. Hence, the 1-2 haplotypes per location in Europe are solely a sign for bottlenecks when compared with

Chinese multiple haplotypes per site. This is strongly affected by the genetic background which resulted from the admixture within China as a result of the 'Great green wall' campaign (Wang et al. 2010) as also previously suggested (Carter et al. 2009c; Javal et al. 2019a; Javal et al. 2019b). Perhaps the population stratification visible in the European dataset is more a reflection of a haplotype bottleneck of the random sample introduced from the whole spectrum of haplotypes within one population in the native range as also suggested by Carter et al. (2010). Thus, it cannot be reliably assured with COI sequences alone if the multiple maternal lineages originate from different source populations.

When using estimates of the mitochondrial mutation rates in insects as proposed in **Box 5**, the divergences (**Table 13**) and branch lengths (**Figure 20**, **Figure 21**) can be approximated in divergence time. The highest divergence times were caculated between COI_Type1 and Type7 (136500-200200 years), Type8 (91500-134200 years) and Type6 (184500-270600 years) and the lowest between Type1 to 2, 4 and 5 (22500-33000 years). Since the different maternal lineages are genetically distant in evolutionary time scales, it indicates a lower chance for the usage in population genetic approaches of the invasion history of ALB, considering the decade long admixture events which will have destroyed in a lot of cases the possibility to assign haplotypes to specific regions, especially the main haplotypes (Carter et al. 2009c; Javal et al. 2019a).

Box 5. Molecular clock for the divergence time of 0.001 substitutions per site (mtDNA)

$$E[d] = 2\mu t \rightarrow t = \frac{E[d]}{2\mu} \text{ (Nielsen and Slatkin 2013)}$$

$$ta = \frac{0.001}{2\mu a} = 14705.88235 \approx 15000 \text{ years}$$

$$tb = \frac{0.001}{2\mu b} = 21739.13043 \approx 22000 \text{ years}$$

d = number of nucleotide differences; t = time in years; μ = mutation rate; 3.54% divergence per My ≙ μa = $0.034 \cdot 10^{-6}$ estimate of Papadopoulou et al. (2010) for COI (tenebrionid beetles); 2.3% divergence per My ≙ μb = $0.023 \cdot 10^{-6}$ standard estimate for mtDNA insects. (Papadopoulou et al. 2010)

Genetic structure of the global COI dataset

To bring the collected new haplotype data in a global context, the dataset was compared with the whole global available data, inhabiting also other invasive areas from North America, Europe, South Korea and, even more important, the native ones from China and Korea mined from BOLD and Genbank. Despite the very short length, especially the sequences of Javal et al. (2019a) were kept to have comparable data from other herein not samples European infestation sites. Some of the haploypes of the full-length COI-barcoding marker lost some segregating sites and were thereby represented by only one of the 33 short COI-marker haplotypes. Even more data is lost by COI_h02, which additionally had several ambiguities

and masked six parsimony informative sites. In addition, in some cases insufficient concomitant data of collection sites and geographic coordinates led to information loss, so ten out of 33 haplotypes were not usable as population genetic marker. The distinct and rare haplotype COI_Type6 in COI-658 bp marker is represented in COI_h03 as one of the main detected haplotypes on the global dataset, found in the USA, China, South Korea, but in the reduced size. The highest occurrence was reported for Northeast China in Xingcheng, Lishu, Tongliao, Chengde and Yanji, supporting the statement above that the occurrence cannot be estimated without a representative dataset. In North America and Europe, almost the whole native haplotype spectrum was covered, the only difference is the haplotype diversity per site because of bottlenecked populations. The invasive populations within Europe and North Amercia are strongly diminished by genetic bottlenecks during the invasion process (this study; Carter et al. 2010; Javal et al. 2019a; Tsykun et al. 2019), so the high molecular variance (>70%<<80%) is derived by the native populations. The multiple haplotypes (>2) per site in China, in contrast to the North American and European sites (1-2) have contributed to this as well. The genetic differentiation of the global COI-360 bp dataset is high (~0.2 F_{ST} and F_{SC}), but lower than with the full-length COI-658 bp European dataset (>0.9), also suggesting a contribution of the reduced DNA marker size on the less measured variation. The very low molecular variance of ~5% and differentiation among the continents (~0.05 F_{CT}) indicate the broad spectrum of haplotypes distributed by the global invasion. However, the native and invasive ranges are unevenly sampled, especially because some Korean COI sequences were assumed to be native by Javal et al. (2019a) but proven to be invasive by Lee et al. (2020).

COI-barcoding for specimen identification of ALB

Beside other issues, in regards to the challenges accompanied with morphological identification (Wu et al. 2017), the COI barcoding marker was further checked for its capability for specimen identification of ALB. Within the *Anoplophora* genus for not morphological trained personnel many species are difficult to be distinguished either as larva or even imagines (Lingafelter and Hoebeke 2002), which can be assisted by COI-barcoding to counteract misdiagnoses (Hebert et al. 2003). In most cases *Anoplophora* spp. could be distinguished by COI-sequences and ALB is forming a well-supported own group, as also reported by Ohbayashi and Ogawa (2009). But there were several inconsistencies for the species *A. chinensis* (Forster), *A. davidis* (Fairmaire), *A. macularia* (Thomson) and *A. horsfieldii* (Hope), from which the first three are only ~3% divergent to each. Regarding the phylogenies of *A. chinensis*, *A. davidis* and *A. macularia*, it seems like there were three misplaced specimens. According to Lingafelter and Hoebeke (2002), these three species are morphologically similar and share geographically occurrences. While *A. chinensis* occurs in China, Japan, Korea, Taiwan and the Philippines, *A. davidis* is found in China and the Philippines and *A. macularia* in China, Japan, and Taiwan. Hence, morphological

mischaracterizations can be assumed. The sequence assigned to *A. horsfieldii* next to the *A. glabripennis* cluster is also most likely a misdiagnosis, since *A. horsfieldii* is usually not similar to *A. glabripennis*. The characteristic bold yellow spots (Lingafelter and Hoebeke 2002) might have led to a mischaracterization of a yellow spotted *A. glabripennis* (Motschulsky) form *nobilis* specimen. *A. glabripennis* form *nobilis* was neither oberved in the Lombardia region (COI_Type6; Corbetta, Vittuone), nor anywhere else in the Italian territory (Zugno, pers. comm.). Hence, the low divergence to COI_Type6 might be coincidential and could still be mischaracterization of a yellow spotted *A. glabripennis.* Maybe due to the many different morphotypes of *Anoplophora* spp., the morphological identifications were not always reliable. The complications stemming from using databases for COI sequences of *Anoplophora* spp. apparently highlights that none of the quick identification methods (molecular, morphological) can be solely used without the other. Inconsistent length standards influenced the utility of some COI sequences. The results of the phylogeny with *Anoplophora* spp. sequences confirmed the reliable identification on species level of ALB sequences, even when the are strongly diverged. Hence, the availablility of matching reference barcodes needs to be increased with specimens which were all morphological classified by trained personnel from the plant health offices in advance and cross-checked using BLAST as well. This is really important, because most problems and confusions endusers of public databases encounter are insufficient taxonomic coverage, gaps and misidentifications (Wu et al. 2017).

Prospects and limits of COI-barcoding for ALB

Despite the many advantageous features of COI-barcoding, this study also spotted limitations in the usage as population genetic DNA marker for ALB, produced by biological and technical factors. COI sequences of ALB are limited in the inference, as mentioned before, due to mixed genetic background of ALB populations. Additionally, the interpretations by comparing them with previous studies is limited by the databases themselves. Missing information, short lengths, and ambiguities can vastly reduce the reliablility of the comparision. Some information about collection sites or geographical coordinates were difficult or impossible to mine from the databases, which additionally reduced the chance to assign haplotypes to source populations, as well as decelerated the assignment speed. Furthermore, the susceptibility of such databases for misidentified specimens, as well as the suceptibility of NUMTs co-amplification in the PCR also reduce the reliablility. Especially the mischaracterizations found within the *Anoplophora* spp. emphasize how important an integrative approach with close cooperation between experts in taxonomy and in molecular approaches are, as well as a flawless documentation. Therefore, standards for data submission on public databases should be aligned to avoid misinformation in the future. Especially databases for COI barcoding (BOLD) should reduce the sequences to one standardized length and compulsory deposition of sample information including collection sites.

Nevertheless, COI-barcoding still is rightfully widely used to characterize invasive insects, since the easy and quick specimen identifcation is paying off for the phytosanitary practice and the results of this study confirmed its ability to distinguish ALB from other *Anoplophora* spp. The reproducibility is quite simple in performing PCRs and simple Sanger-sequencing. Even with imperfect sequence analysis software available, easy tools available as freeware are able to do a sufficient job, even for less experienced personnel. Only the Primer-BLAST results and suceptibility to NUMTs co-amplification demand for the common usage of species-specific primers, such as the specific primers proposed in EPPO (2016), which should be tested for specificity in advance. The unique and not shared haplotypes in the invasive range could at least indicate an independent introduction e.g. Magdeburg, Marly-Bruensried, Corbetta-Vittuone. With an appropriate sampling size, sample quality and provided information, COI barcoding can even contribute to knowledge about genetic structure and history assessment in low resolution, as for instance shown by Lee et al. (2020). The utilization quality could be improved enormously for this kind of analyses, if databases were fed with more sequences in larger sizes and sufficient sample information. In combination with other DNA-markers, even the usage for population genetic inferences for ALB could be improved for quick assessments (Dupuis et al. 2012). But future perspectives for the inference of invasion sources and dispersal routes lie more in HTS sequencing approaches, like mitogenome sequencing, Amplicon-sequencing or genome-wide SNP data.

6.3 Chapter B discussion: Discovery and genetic diversity of NUMTs

Chapter B

"Mitochondrial insertions in the genome of the invasive long-horned beetle *Anoplophora glabripennis* – genetic diversity of NUMTs and their impact on DNA barcoding"

The discovery of NUMTs in the genome of *A. glabripennis* is reported for the first time in this study. Some previous studies which used COI sequences, tested their sequences for pseudogenes (Javal et al. 2019a; Kim et al. 2019; Wu et al. 2017) by checking for frameshifts and ambiguities, while Lee et al. (2020) re-sequenced ambiguous sequences and removed them if still unclear without naming to look for NUMTs or pseudogenes. No tests on mitochondrial pseudogenes were reported in the studies of Carter et al. (2009a; 2009b; 2010). There were 209 pseudogenes described within the first draft genome of ALB from McKenna et al. (2016), but without further specification if these stem from mitochondrial or nuclear origin. The only pseudogene on a herein studied scaffold (sc1356) resembles a Glycoside hydrolase family 1 protein 27, which is thereby no NUMT sequence. Hence, none of the studied NUMTs or alleged NUMTs revealed here have been reported anywhere else.

Detection of mitochondrial insertions in ALB's nuclear genome

Since the occurrence of NUMTs can strongly compromise population genetic studies with their similarity to the mtDNA sequences (Haran et al. 2015; Song et al. 2008), the vulnerability of COI barcoding by overestimating mitochondrial variation was evaluated. During sequencing of a 1271 bp COI-II segment, indels, stop codons and extreme high nucleotide differences gave a hint to the presence of NUMTs. The phylogeny of the cloned sequences of ALB-COI-II (**Figure 38**) imply that the sequences captured were neither mitochondrial, nor some of the sequenced NUMTs, since all were very distant and without any cross-links. The phylogenies (**Figure 37**, **Figure 38**) further imply that some of these copies could have happened a long time ago and underwent multiple post-insertion processes like duplications. These co-amplified sequences must derive from parts of the ALB genome that are still not sequenced or considered in the newest assembly so far. Since mitochondrial sequences are integreated into the genome at double-strand break sites (Blanchard and Schmidt 1996; Kleine et al. 2009) there is a high likelihood that these kinds of unstable regions were also difficult to sequence. The screening of ALB's genome with BLAST revealed several putative NUMTs which were either continous or rearranged, from which three continous ones were sequenced.

Genetic characterization and divergence of three NUMTs of ALB

All three sequenced NUMTs ALBsc158-NUMT1, ALBsc1709-NUMT2 and ALBsc1709-NUMT3 could successfully be confirmed to be real NUMTs and no assembly artefacts in ALB's WGS by PCR product ampifications with one of the primers on a scaffold-fragment and the other one on the mitochondrial-like sequence. NUMT1 and NUMT3 had several difficulties in amplification and sequencing like multiple banding or ambiguities. Therefore, long-winded process adjustments were necessary, which were not always successful in the end. Surprisingly, the divergences were very different in the fragment parts for NUMT1 and NUMT3, even contradicting in regards to the ALBsc158-NUMT1(int). The divergence of ALBsc158-NUMT1(int) sequences to the mitochondrial sequences were so low, that another duplication which was preferantially amplified was possible. These are signs of multiple NUMTs disturbing the amplification of other NUMTs, which is also supported by the BLAST results from the NUMT haplotypes (**Table 31**-**Table 33**).

The equation depicted in **Box 6** was used to approximate the time of insertion of the mitochondrial sequence into the nuclear genome. The mutation rates between $5 \cdot 10^{-9}$ and $4 \cdot 10^{-9}$ are common to most invertebrate genomes. Due to different divergence times between the single fragments of NUMT1 and NUMT3, the highest divergence was used to roughly estimate the divergence time, assuming these NUMT fragments are linked to each other. The longest divergence time between the mitochondrial sequence and the NUMT sequence was measured for NUMT2 (~5.4-9 My), followed by NUMT3 (~4.2-7 My) and NUMT1 (~3.2-5.3 My).

Box 6. Molecular clock for 6.295% mean divergence of ALBsc158-NUMT1 (gDNA)

$$E[d] = 2\mu t \rightarrow t = \frac{E[d]}{2\mu}$$ (Nielsen and Slatkin 2013)

$$ta = \frac{0.06295}{2\mu a} \approx 6.3$$ million generations | : 2 → ta ≈ 3.2 My

$$tb = \frac{0.06295}{2\mu b} \approx 10.5$$ million generations | : 2 → tb ≈ 5.3 My

d = number of nucleotide differences; t = time in generations (two years); μ = mutation rate; $\mu a = 5 \cdot 10^{-9}$; $\mu b = 3 \cdot 10^{-9}$ (Lynch 2010; Rashid et al. 2022)

e.g. *Drosophila melanogaster* $2.8 \cdot 10^{-9}$ (Keightley et al. 2014); *Bombus terrestris* $3.6 \cdot 10^{-9}$ (Liu et al. 2017)

Even though all NUMTs had an interesting potential as 'molecular fossil' representing millions of years of evolution, there are numerous reasons against their usage as phylogenetic marker. First, the difficulties in PCR product amplification and sequencing are unsuitable for the usage as marker (NUMT1, NUMT3). Second, it is difficult to assess, if the missing PCR products of some specimens were caused by poor DNA quality or by a real absence or shortness of the NUMT itself as a result of an unstable genome region. Beside the obvious co-ampflification of other NUMTs (e.g. on ALBsc1709, **Table 10**) or maybe mitochondrial sequence fragments (e.g. NUMT1(int)), there was no segregating geographic pattern observable for any of the three sequenced NUMTs, or even completely diverse ones in the different fragments (NUMT1, NUMT3). All three NUMTs showed signs of multiple copies in the genome, such as BLAST hits on the same or other scaffolds with high %identities, or high nucleotide differences to the NUMT-reference. The multiple copies also made it difficult to distinguish which one of the copies was sequenced. Furthermore, for ALBsc1709-NUMT3 some specimens possessed very short microsatellite-like structures like $(GTT)_n$ and $(CAA)_n$ with imperfect matching tandem repeats when sequencing the cloned front-end sequences of NUMT3. Probably this can be explained by heterozygous individuals for this locus with a prescense/abscense variation for NUMT3, which might have replaced a former microsatellite due to double-strand breaks (Blanchard and Schmidt 1996) or the primers matched another unknown genome region with this microsatellite which is occationally co-amplified in low frequences. All these observations fit to the insertion age calculated for them. The divergence times of these NUMTs to the mitochondrial counterpart agree with the observation that older NUMTs exhibit vast numbers of duplications (Hazkani-Covo et al. 2003).

Impact of NUMTs on DNA barcoding

Since no NUMT has been revealed or reported from ALB's genome until now, contaminations in databases with NUMTs which occurred during COI-barcoding or other mtDNA markers stayed most likely undetected as well. Therefore, the references of the sequenced NUMTs as well as a subset of the clusters A and B amplified by ALB-COI-II primers were blasted against the whole nucleotide collection and representative genomes of *A. glabripennis* on NCBI.

Strikingly, the megaBLAST of six specimens with cluster A and B sequences respectively, revealed possible contaminations by BLAST hits of >99% identity with cluster B sequences. The sequences were KF737825.1, EU914794.1, EU914771.1, EU914759.1, EU914698.1 and EU914665.1. The first is an unpubished sequence, while the others were all from Carter et al. (2009c). Primer-BLAST of the primers of Carter et al. (2009c), which were also used by Lee et al. (2020), showed suceptibility to NUMTs too. All these sequences were included in the global perspective on mtCOI-360 bp (5.2.2), represented in the main haplotypes COI_h01 and h02. This is an alarming signal, since they were not noticeable as possible NUMT sequence, because parts of the NUMTs might still resemble most likely real mitochondrial sequences.

Implications by the presence of NUMTs in ALB's genome

This study showed evidence for the first time that ALB has a lot of NUMTs in its genome, which should be considered in future population genetic, evolutionary, and barcoding studies. Thereby, the detected presence of NUMTs can be seen as a chance to improve identification and phylogenetic approaches on ALB in two different ways, either in avoidance or in utilization.

First of all, the results presented here propose that it is time to adjust the usage of COI barcoding in population genetic analyses concerning Asian long-horned beetle, despite all the valuable information gained (6.2, Carter et al. 2010; Javal et al. 2019a; Kim et al. 2019; Lee et al. 2020). The high genetic diversity within mtCOI in ALB is challenged by the abundance of several NUMTs and possibly high amount of undetectable NUMTs. In regards that e.g. ALBsc158-NUMT1 did not show any frame shifts despite beeing allegedly 3.2-5.3 My old, the other with ALB-COI-II primers co-amplified ones must be even older than that (e.g. humans oldest NUMTs 58 My, Bensasson et al. 2003). Contaminations of the databases with young NUMTs that did not lead to frame shifts and/or sudden aborts so far, could already be part and cannot be disproven easily. Thus, it is important to always use other markers beside COI in easy diagnosis approaches, hence the approach of Rizzo et al. (2020) with several mitochondrial markers, and Taddei et al. (2021) using additional 18S DNA-marker, are quite promising in this regard. A performance with cDNA instead of gDNA would erase the problem of NUMT sequences with frame shifts but might not be sufficient for other NUMTs without shifts. Though easy adjustments on the standard practice could help to counteract the adverse effects of NUMTs co-amplification, e.g. the usage of muscle tissue to enhance mtDNA, new long-range mtDNA markers, occasional whole genome sequencing of single individuals, usage of cDNA and species-specific primers (Song et al. 2008). Whole mitogenome sequencing of many individuals might be only feasible for research purpose only. When considering the Primer-BLAST results of the COI marker with Folmer et al. (1994) primers LCO1490/HCO2198 or the Carter et al. (2009c) primers 4S/4A, the susceptibility of co-amplification with NUMTs is apparently probable. The disability to fully recover databases from putative NUMT sequence

contaminations supports the statement of Dupuis et al. (2012), which states one marker is not enough to discriminate closely related species or as in this study, to discriminate geographically distinct populations. Hence, together with avoidance strategies of unintentional co-amplification of NUMTs, long-range PCR amplifications of several regions of the mitchondrial genome could be sequenced by efficient high-throughput multiplexed Amplicon-sequencing, for instance by covering the whole mitogenome (e.g Heumann-Kiesler et al. 2021). Regions like e.g ND2-COI-COII and ND4-ND6-CYTB in full-length could be applied as well, from which the first might be still suceptible for NUMTs because of genome insertions over longer stretches of mitochondrial-like sequences (**Figure 46**). The doubled haplotype of specimens from Feldkirchen might be a very recent insertion into the nuclear genome too. Using whole mitogenome sequencing it could be solved whether it is a NUMT sequence or a real heteroplasmy.

Furthermore, follow up research could help with finding optimal NUMTs that might be used as neutral segregating markers in population genetic approaches, since they can combine the evolutionary history of mitochondria and nuclear genomes at once (Bensasson et al. 2003; Hazkani-Covo et al. 2010). However, such a marker has yet to be found, since the three sequenced ones were too disturbed by co-amplification of other NUMTs themselves. Therefore, rearranged and long NUMTs could be a good option, because due to the rearrangement they are more unique and include passages of nuclear DNA between the NUMTs. This might give some insights in ancient population structure beside the current blurred genetic background of Chinese ALB populations. Not all of them seem to be sequenced yet, so the number of under-reported NUMTs cannot be identified so far and should be further addressed in future genome assemblies, because very recent NUMTs might be mistaken for mtDNA contamination and therefore sorted out (Bensasson et al. 2003). The time of insertion of the NUMTs could also be dated by comparing NUMTs of closely related species (Hazkani-Covo et al. 2010). Therefore, the closely related species *A. chinensis*, which is also invasive in Europe (EPPO 2021c), should be screened for NUMTs.

6.4 Chapter C discussion: Complex introduction of ALB in Europe

Chapter C

"Complex European invasion history of *Anoplophora glabripennis* – new insights in its population genomic differentiation using Genotype-by-sequencing"

The nGBS method from LGC Genomics was used to discover genomic SNPs for population genomic insights revealing kinship or population structure between the different infestation sites to infer introduction patterns into Germany, Switzerland and Italy. The dispersal pattern of *A. glabripennis* into and within Europe is proven to be highly complex and is shaped by

multiple independent introductions and some scattered cases of human mediated secondary dispersal, as also indicated by previous studies using COI and microsatellites only (Javal et al. 2019b; Tsykun et al. 2019).

Evidence for genetic bottlenecks within European infestation sites

The presumably mostly single introduction events into Europe are concomitant with the evidence for population structure, resulting in a severe reduction of genetic diversity. The LD decay with distance can also give a rough insight into the effective population size. In small bottlenecked populations, higher background levels of LD and slow LD decays are common, while lower LD measures and faster LD decays are typical for large and outbred populations. (Rogers 2014; Sved et al. 2013) Beside the population size and population structure, LD is also influenced by genetic drift, which is enhanced in small bottlenecked populations. The European ALB SNP dataset showed a moderate LD with very slow decay of LD (**Figure 68**), which is very likely the result of small and bottlenecked founder populations within Europe. The very low nucleotide diversity within each of the sampled European populations (overall π, **Figure 69**, 0-0.00005 in 10 kb), as a measure of the degree of polymorphisms, supports the assumption of a small effective population size with a prior severe genetic bottleneck in most European infestation sites. The small founding populations within Europe with fewer genetic diversity apparently also suffer from high inbreeding as pictured in **Figure 71**. High inbreeding coefficients >0.25, as measured for this dataset, are often found in populations where mating with parents and siblings occurs. The herein observed Wahlund effect (reduction of heterozygosity) follows the assumption of small population sizes, inbreeding and population structure. There was also a conspiciously high amount of very rare SNPs, which is likewise typical for a small random sample introduced by source populations. This phenomenon is called founder effect, when the frequencies of originally rare alleles is increased by genetic drift (Dlugosch and Parker 2008). The results of the Admixture analysis confirm this observation, displaying almost all infestation sites with a homogeneous genetic structure within the site.

Low secondary dispersal and high population stucture in Europe

The extremely strong differentiation among all demes (>0.4 F_{ST}) suggests no mating and mixture between most of the infestation sites within Europe (**Table 19**). Furthermore, the differentiation and variation among the demes within a country (~0.35 F_{SC}) suggest that introduction sources within the countries are likely not shared in most cases, hence the levels of secondary spread are indicated to be low. On the other hand, the high but comparatively lower differentiation among the countries indicates, that within Europe a huge spectrum of genetic backgrounds from the source populations is represented, from which some came from shared introduction sources or introductions with less differentiated source populations. The

pairwise F_{ST}-values between the European subpopulations validated the measured population structure in more detail. Moderate differentiation was only measured between the neighbouring locations Marly-Bruensried and between populations that showed high proportion of subpopulations in the Admixture analysis (**Figure 81**). This is either a sign of a human mediated secondary dispersal, the same source of introduction, or some other shared proportion of genetic ancestry. Surprisingly Corbetta and Vittuone showed population structure within the whole 170f8g50m50 dataset (F_{ST}=0.260), while after the removal of dependent SNPs they shared the same structure (**Figure 81**), hence these subpopulations have high variance in the SNPs in LD, stemming more from genetic drift in the bottlenecked founder populalions than from genetic structure. This can also be explained by the two different clusters of this location pair visible in **Figure 86**. Some of the sporadic measured F_{ST}-values per bins >0.5 up to ~0.9 in all pairwise comparisons could be further invastigated to find possible target regions which are most segratagting between the different populations.

Complex introduction history of ALB in Europe

The Admixture analysis models the proportions of the genome derived from different source populations (Alexander et al. 2009) and the ancestry (**Figure 81**) revealed for most individuals the same subpopulation(s) within their collection site, which is a sign of one introduction and genetic drift. Hence, even the collection sites with several subpopulation proportions were in the most cases homogeneous within the sites. The complementation with a phylogenetic network analysis based on 32432 SNPs on the 170f8g50m50 subset confirmed this observation. Beside a few exceptions, most specimens were clustered together with their own collection site and are tightly cross-linked, reflecting the strong kinship within the collection sites, most likely due to a small amount of founding individuals, as observed before in Europe (Javal et al. 2019b). Evidence of secondary translocation and high population structure within the European infestation sites of ALB, could also be confirmed. Apparently, most of the individuals within a collection site were directly genetically relatable and differentiated from other infestestion sites. This is most likely the result of non-random mating within Europe and can also be a sign of severe genetic bottlenecks. So, in populations with only one dominating subpopulation proportion this could have also been the only introduced subset from a source population and got fixed in the new invasive population by genetic drift and/or preferred mating of some individuals (Bock et al. 2015; Dlugosch and Parker 2008; Hill 2007; Nei et al. 1975). Different to the majority, the phylogeny of the specimens found in Weil am Rhein, Cuneo and Hildrizhausen assume either multiple introductions from different sources, or recurrent introductions of a genetically diverse source.

In the case of the neighbouring sites Marly and Bruensried, as well as Corbetta and Vittuone, it can be assumed that there was secondary transfer between those sites, as it was already

reported by Tsykun et al. (2019) from Marly to Bruensried and Furiani in Corsica (Javal et al. 2019b). Also, Fermo and Ostra, which are 90 km apart, indicate either one introduction and secondary dispersal or the same source. Possibly, this could also be the case between Murnau (D-BY-MU) and Miesbach (D-BY-MI), which are 60 km apart. On closer inspection of several potential subpopulation compositions, a secondary dispersal seems comparatively unlikely between Murnau and Miesbach, since the supopulation proportions with K=7-8 and K=10-11 differ, whereas in the cases of Marly-Bruensried and Corbetta-Vittuone the clustering was stable in different K. The subset 170f10g50m50 based on more common alleles showed a very similar ancestry with 6064 independent SNPs (K=9), **Figure 84**, hence the proven population structure was not overestimated by rare alleles. Although, interestingly Miesbach showed mixed supopulation proportions similar to Kehlheim, which can be deviated from the admixed background in China, or introduction sources within close proximity. This means Miesbach and Kehlheim may have shared ancestries, since both root somewhere in the middle of the phylogenetic network (**Figure 85**).

The highest differentiation was measured between Marly/Bruensried and all the other collection sites in Europe, as well as between Corbetta/Vittuone to all other populations, which is following the phylogenetic connections where these two location-pairs showed the highest divergence to all other populations as well (**Figure 85**). The serial increased K on the Admixture analysis (**Figure 82**) underline the topologies drawn by the PCA and phylogeny showing an ancestrial split from most Italian and Swiss specimens from the German, few Swiss and few Italian specimens. This observation suggests presumably two different geographic larger source regions within the native range, which were not admixed with each other. According to the recent study of Cui et al. 2022 (after submission of this dissertation in March 2022) this might reflect the clear distinct northern and southern groups, or the clear separation of most populations from South Korea and China. Each ot the two groups are highly differentiated between the collection sites but have a shared proportion of genetic ancestry. As also depicted with the PCA, in the phylogenetic network also a "main cluster" is visible rooting in the middle, including all German, some Swiss and Italian specimens. Assuming South Korea as source for the Marly/Bruensried infestation as suggested by Javal et al. (2019b), it can be speculated but not proven that most Italian and Swiss specimens from this study originate from the area of Northeastern China (to the east of gulf of Bohai) and South Korea, while the rest originated from Western and central China. An analysis of the specimens from this study with the SNP panel obtained by Cui et al. 2022 could solve this open question.

The Admixture analysis showed mixed populations which differ between Italy/China and Germany/Switzerland in their subpopulation composition. Maybe recurrent introductions from a strongly admixed introduction source, or from multiple different introduction sources and admixture in Europe resulted in this mixed populations. For Cuneo and Weil am Rhein, multiple

introductions from different sources are probable, since single specimens are differently located in the phylogeny. Otherwise, particularly the fan-shaped structure in the phylogeny of Kelheim and Trescore-Balneario suggests either multiple introductions and subsequent mixture in Europe or consecutive introductions keeping the propagule pressure high and maintaining genetic diversity of one introduction source. Additionally, there is also the possibility of a dataset bias resulting from a related unsampled population which is closely related to several other subpopulations. Hence, the ancestry might be historically uncorrect since the algorithm always fits the data into a pattern of admixture proportions (Lawson et al. 2018), which would not be possible to be proven due to the shortage of specimens from native populations. In Magdeburg/Feldkirchen, the phylogeny revealed a separation into three/two groups. These special cases could also result from one admixed population from which specimens dispersed in different city parts and were thereby bottlenecked. Two (FK)/three (MB) introduction sources that are regional in close vicinity are also a possibility and could still have emerged from one introduction from regional export hubs. Several introductions from different regions would be least parsimonious. The more likely first scenario is only possible with several generations prior the first detections (FK 2012; MB 2014), which is not unusual for IAS (Crooks 2005).

The introduction pattern of ALB into Europe is very complex, but without a representative sample from native or invasive source populations, the full introduction route cannot be deviated. The three sampled Chinese locations Harbin, Jinan and Tongliao do not seem related to the invastigated European populations, but as single individuals, which were most likely not bottlenecked in contrast to the populations in Europe, this is not comparable. In most cases the introductions into Europe were presumably independent, strictly separated and due to the strong detected bottlenecks and founder effects by genetic drift not connected with any other ALB population in Europe. The connections visible in the phylogeny are many substitutions per site in the past, which can be either ancestral splits or connections of former distinct groups due to the admixture in the native range. The founder effect could have been diminished by recurrent introductions with higher propagule pressure (Dlugosch and Parker 2008; Kolbe et al. 2004), which might be the case in populations like e.g. Magdeburg or Feldkirchen. Multiple introduction events could have helped to increase genetic variation (Kolbe et al. 2004), especially in the case of multiple introductions coming from different source populations, which happened in collection sites like Cuneo, Weil am Rhein and Hildrizhausen. This could have also been the case for Kelheim and Trescore-Balneario, which can not be assured without genomic data from source populations. The insights into the introduction patterns into Europe could be improved in comparison to COI and microsatellites. Even for secondary dispersals the genetic drift and probably no further mating in cases of Marly-Bruensried and Corbetta-Vittuone was observable. Additionally, the female single finding D-

BY-SB-18-110 specimen could be directly related to earlier findings in Schoenebach from 2015 by using the 178f8g50m50 subset (**Figure 94**, **Figure 96**). Based on this result the quarantine measures were extended to 31th December 2022 according to the EU guideline by the responsible federal state plant protection organization Bayrische Landesanstalt für Landwirtschaft, LfL (2020).

Future perspectives of Genotype-by-sequencing in invasion science of ALB

For a lot of collection sites, the sample number was too low to make any suggestions on the genetic composition within the infestation site, especially for cases were only one specimen was left for analysis, which might not represent the other individuals from the same site. Hence, these populations were not considered in some analyses. At least it was possible to gain reliable individual kinship and phylogenetic information for all 168 (176) specimens. However, for a serious tracing back of the introduction sources, data on the genomic variation within the natural distribution area and other invasive areas would have been necessary, to be capable to distiguish native populations in advance. For this purpose, especially to detect admixed populations, an appropriate sampling size is mandatory (Roe et al. 2019). The lack of representative native potential source populations made it also impossible to apply ABC methods to reveal the complex evolutionary scenarios of the invasion history in Europe (Cornuet et al. 2014; Pudlo et al. 2016). The collection sites in all three sampled countries with several proportions of source subpopulations gave important insights into the genetic background and might be a confirmation on the by admixture shaped populations within China, proposed by Carter et al. (2009c) based on microsatellites. Considering the complex demographic background in the native areas, it would have been very difficult to guarantee if, when admixture is confirmed, it is originating from the native range or by admixture in the invasive range by multiple introduction sources. Also Cui et al. 2022 described contemporary movement between some regions in the native range, although they found no evidence for large-scale admixture. In addition to the lack of samples from alleged source populations, a representive number of admixed samples would be needed as well, and herein only Trescore-Balneareo (n=8) and Kehlheim (n=9) of the mixed populations would have an approximately decent sample amount. This should be further investigated with F-statistics to detect the admixture events (F3-statistics) and admixture proportions (F4-statistics) (Patterson et al. 2012). Nevertheless, the informative value in the determination of population structure, kinship and admixture is comparatively much higher with the high-resolution genome wide SNP markers in comparison to low resolution standard genetic markers used before.

6.5 Conclusions

In general, for this study many infestation sites within Europe were sampled for population genetic analyses, which was never analyzed in this extent so far. Thereby new knowledge

about the invasion history within Germany and Italy was gained and more details with high resolution genomic SNP markers, was adding up to the results from Tsykun et al. (2019) in Switzerland. In summary, when comparing the results of the COI-Barcoding and GBS approaches, the enormous increase of resolution is observable (**Figure 90**, **Figure 89**). In **Figure 90**, the simplified haplotypes of COI with 17 SNPs are shown and compared to the 7810 SNPs from the Admixture analysis. In constrast to COI-barcoding, the GBS method is more complex and most likely not easy to conduct in all diagnostic laboratories, although it is considerable easy and cheap when compared with other HTS methods like WGS resequencing approaches. Nevertheless, with up to 32000 SNPs the inference is much more precise (**Figure 89**). With this dataset, population structure among most collection sites and no structure between at least two pairs of neighbouring locations were discovered with both approaches. The usage of GBS resulted in a much higher resolution and thereby a better inference of the kinship, the introduction pattern and translocations within Europe in general. Thus, with mtCOI alone there would have not been a clear separation in several independent subpopulations compared to the COI-Type1 haplotype (**Figure 90**, **Figure 89**). The separation in three groups within the Magdeburg collection sites or Feldkirchen in two groups would also not have been detected with COI-Barcoding. Carter et al. (2009b) also described single haplotypes in mtDNA, while more polymorphisms could be discovered with the nuclear marker on the same specimens in Canada. Hence, microsatellites were an improvement over COI DNA barcoding, but substantial enhancement over microsatellites was achieved by GBS, with which even not much differentiated populations could be distinguished, e.g. Marly-Bruensried (**Figure 86**), Corbetta-Vittuone (**Figure 87**), Feldkirchen (**Figure 88**).

The resolution could be enhanced compared to previous studies and offers a valuable and highly reliable pattern of genetic relationships between German, Italian and Swiss infestation sites, but astonishingly, the inference of the European SNP dataset came to comparable conclusions as former studies using COI and microsatellites before (Carter et al. 2010; Javal et al. 2019a; Javal et al. 2019b; Kim et al. 2019; Lee et al. 2020; Tsykun et al. 2019). All the previous studies emphasized the complexity of the invasive pathways of *A. glabripennis,* which are shaped by a highly indistinct genetic background from the native range and multiple independent, maybe consecutive introductions combined with secondary spread events in the invasive range. Most likely, these conditions, influenced by human-mediated translocations, fostered the invasion success of ALB in North-America and Europe, since the limited genetic diversity did not attenuate the invasiveness (Carter et al. 2010; Javal et al. 2019b; Tsykun et al. 2019). Some of the mixed populations in the European dataset from this study are in accordance to the strongly admixed genetic background from Chinese ALBs as suggested by the previous studies. Carter et al. (2009b) and Javal et al. (2019b) proposed the genetic background from China to be shaped by complex migrations of *A. glabripennis* caused by

humans and mixture from different sources during the reforestation program of the Chinese government, starting in the 1960s (Haack et al. 2010; Wang et al. 2010). Although, Cui et al. 2022 suggest that recent admixture in the native range occurred less than previously expected. However, in this study probable admixture in the invasive range was reported for the first time (e.g. Kelheim, Trescore-Balneario), which either happened in the source populations prior the invasion, or within the respective European infestation site, or is at risk to happen there (e.g. Cuneo) as a result of multiple introduction sources. Additionally, the impression of multiple independent introductions into Europe is supported by the already proven extremely high genetic differentiation between most European populations with mtCOI-658 bp DNA marker and genomic SNP dataset, as well as by the measured founder effects. Under the assumption of an admixed genetic background in China, only the detection of independent introduction events into Europe is possible. Independent introduction sources cannot be deviated, because due to genetic bottlenecks accompanied by genetic drift, equal introduction sources may look different in the invasive range when other random samples were introduced. The high genetic structure within Europe are further evidence that not much mating is possible between the different populations, even when they are only slightly spatially distant, e.g. in Bavaria. ALB is mostly dispersed in large spatial scales across countries and continents by human transport (containers, wood packing material, transport of wooden clippings, etc.), rather than among locations of one invaded country. However, exceptions like Marly/Bruensried and Corbetta/Vittuone argue for either human-mediated or natural range expansion within a small spatial scale, which was also described in previous studies (Javal et al. 2019b; Lee et al. 2020; Tsykun et al. 2019).

Despite the evidence of scarce genetic diversity within most infestation sites due to population bottlenecks, *A. glabripennis* could establish in many infestation sites, at least for a while in case of already eradicated sites. The sites with recurrent or multiple introductions have most likely not much loss of genetic diversity. Anthropogenically induced adaptation to invade (AIAI) is a very likely scenario why ALB was not much affected by hostile effects from genetic bottlenecks (Hufbauer et al. 2012). The natural, not invasive populations in China/South Korea from rural or forest areas are improbable introduction sources because of reduced connectibility. Transportation hubs, which accelerate spread (Morel-Journel et al. 2019), are in proximate distance to (sub)urban infestation sites that are most likely themselves invasive (Carter et al. 2009c; Javal et al. 2019a; Lee et al. 2020). In Europe, ALB does not face adaptive challenges because the climate conditions and host trees are comparable to introduction sources from (sub)urban areas in Asia e.g. as Facon et al. 2006 described how matching environments can promote invasion success. Hence there is no genetic paradox in Europe. Within China and Korea, the presumably invasive populations are also not genetically paradox, since admixture events could have created new genetic variation and therefore may have

strengthened the populations (Carter et al. 2009c; Lavergne and Molofsky 2007; Rius and Darling 2014).

The multitude of different quality and storage levels of the received ALB samples from the federal plant health offices disclose the necessity of a reformation for advanced future practical approaches. Even though keeping specimens pinned and dried is very important for the training in morphological identification, as well as exhibition for other educational purpose, it should not be the prefered practice for sampling, because integrative research is more desirable to face the challenge of quarantine pests. Good quality studies which can also assess histories of invasions should be involved in practice when standard methods evolve over time, not only specimen identification. Storing of specimens at -20 to -80°C and with appropriate sample amounts per collection site (at least 10) that are soley dedicated for DNA-based approaches, along with keeping morphological references should be the norm. For genome sequencing approaches, immediate shock freezing at the collection site would be most suitable. As proposed by Roe et al. (2019) and Blackburn et al. (2020), intenational frameworks and cooperation enganging NPPOs for sample collection and research institutions (e.g. Universities) for respective analyses are very promising to counteract the global spread of IAS. To merge all collected knowledge about biological invasions between different research institutions around the world, methods, standards and data quality need to be aligned (Blackburn et al. 2020; Latombe et al. 2017). To ensure high quality genomic DNA, the sample preservation and trapping protocols need to be appropriate and effective to be feasible within the plant health practice to reduce management costs and efforts. Some adaptions are advised, based on this study, to enhance the identification methods of ALB. COI DNA-barcoding markers are possibly susceptible to NUMT co-amplification although all translations of the COI haplotypes in NCBI, BOLD and collected in this study had no frame shifts when translated to amino acid sequences (658 bp, 360 bp). Nevertheless, younger NUMTs would still be invisible, as for instance NUMT1 on ALBsc158 did not have any frame shift as well. Hence, the use of species-specific primers, full length of 658 bp COI to compare with databases and checks for multibanding, ambiguities, frame shifts and stop codons are highly recommended. The combination with another DNA-marker, randomly picked mitogenome sequencing or a complete replacement with long-range Amplicon sequencing would refine a quick first genetic characterization. Further advice for phytosanitary measures based on the results of this study are special precautions when genetic analyses reveal multiple introduction sources, because this populations might be more invasive (Dlugosch and Parker 2008; Kolbe et al. 2004; Uller and Leimu 2011) Hence, the strict restrictions in transport of wooden products in infested areas to avoid translocations in adjacent areas that already exist should be implemented very rigorously. Given that ALB usually flies only short distances, along with high distances amongst most collection sites, the results of this study are quite encouraging for the

regional plant health offices, since they allow suggestions of no or only few human-mediated secondary translocations on medium or long distances.

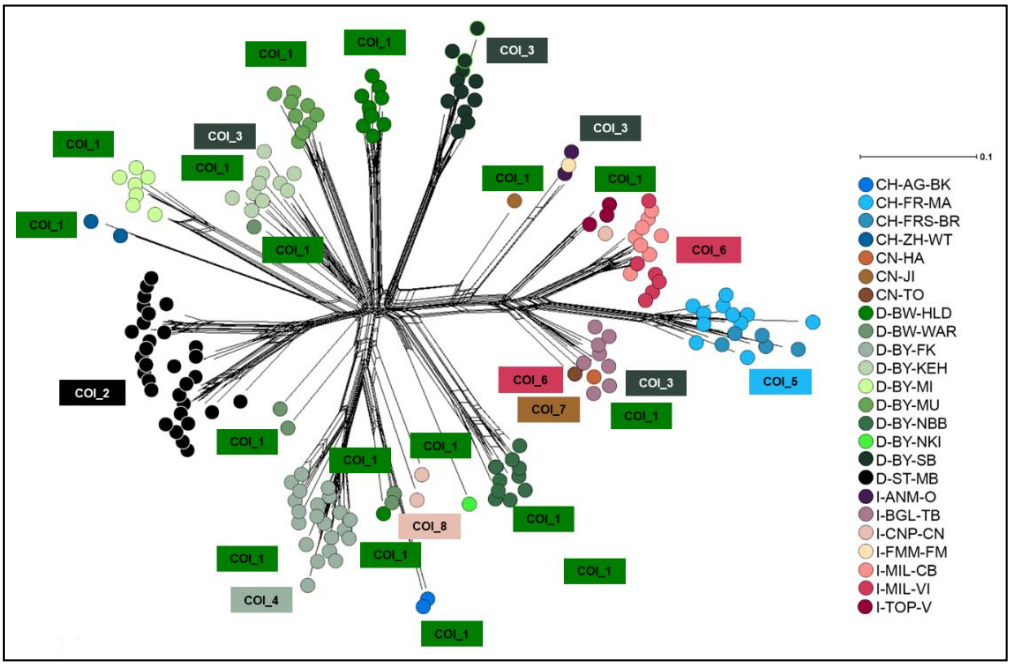


Figure 89. Comparison of polymorphic sites detected with COI and GBS approach. For COI 17 informative polymorphisms were found, whereas for the phylogenetic network 32432 SNPs were used. Some COI-sequenced specimens were not considered in the network (e.g. I-FMM-FM-15-169, COI_Type1) and some GBS-sequenced specimens were not sequenced on mtCOI (e.g. D-BY-NKI-12-022).

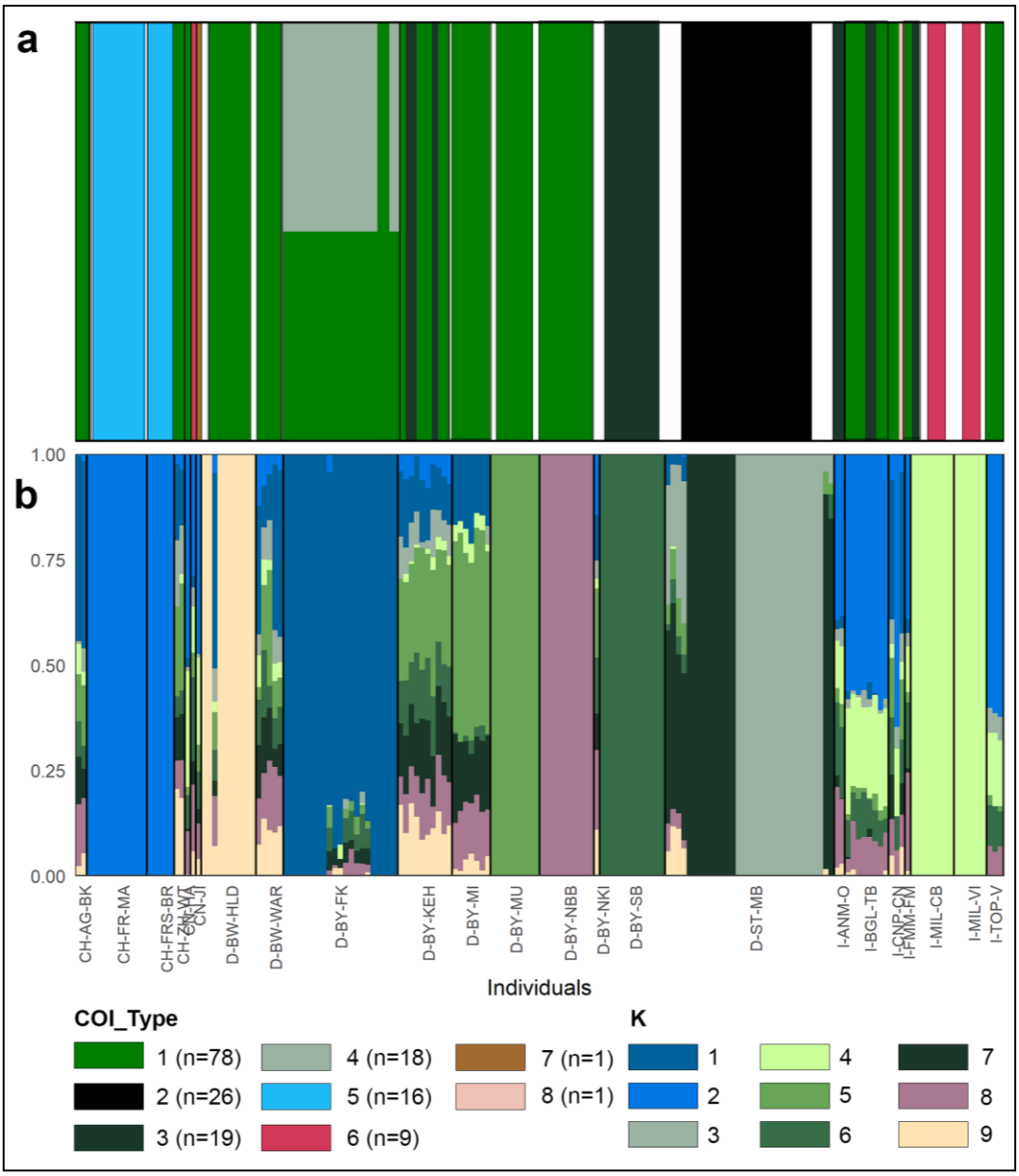


Figure 90. Comparison of polymorphic sites detected with **a)** COI and **b)** GBS approach. For COI 17 informative polymorphisms were found, whereas for the ADMIXTURE analyses 7810 SNPs were used. The gaps in **a)** are due to less COI-sequenced individuals than with GBS approach.

6.6 Outlook

Based on the results of the intraspecific characterization of *A. glabripennis* within 22 invasive populations in Germany, Italy and Switzerland, many detailed insights on small and bigger scales were gained by COI, NUMTs and genomic SNPs, but also a lot of open-ended questions were added on, featuring many future research opportunities to follow.

The detailed insights gained into the introduction patterns within the European countries Germany, Italy and Switzerland are an unprecedented model of how new state-of-the art methods can help to minimize the threat posed by IAS. By adding substantial sequence and

variant knowledge on *A. glabripennis* in the invaded habitats of Europe, the knowledge about processes involved in a successful establishment are more comprehensive to develop efficient tools for control measures against invasive species. To achieve this goal, all information has to be shared and communicated with the international plant health and invasion biology research community (Blackburn et al. 2020; Roe et al. 2019). With accumulative acquaintance on a global scale, risk assessment can be fed with information on introduction patterns and secondary dispersal risks (Mastin et al. 2020; Roe et al. 2019). For instance, the data of this study supposes that secondary dispersal within Europe over long-distance is not very likely, only within small scales. Prevention measures can profit from these insights. The invasion success is directly influenced by the quantity of human activities (Eritja et al. 2017; Simberloff 2009) and the results of this study emphasize how eminently important avoiding of recurrent and especially multiple origin introduction events are to diminish the invasion success. Hence, with further knowledge of introduction sources, introduction hubs could be identified and should be of major concern in the future (Bilodeau et al. 2019; Morel-Journel et al. 2019). Sequencing technologies will become much cheaper in the future and databases will be filled with more information (González et al. 2018; Roe et al. 2019). Pioneer studies like the study presented here can strongly enhance phytosanitary measures and early responses in invasion science. Therefore, the results of this study can serve as a guideline on how to reprocess past invasion events as well as on how to prepare for upcoming invasions.

A direct follow up approach on the data received of invasive European ALB specimens would certainly be the comparison with other invasive and native populations. For an improved global perspective on invasion history and dispersal patterns, a larger and more balanced sampling size including native areas and other invasion areas (Cristescu 2015; Roe et al. 2019) would be the best approach to conduct screening for genome wide SNPs with Genotype-by-sequencing of ALB. This would naturally include an extension of international collaborations to get a representative sample of the whole natural distribution. As already criticized in previous population genetic investigations of *A. glabripennis* before, the results from this study also emphasize the need for more representative samples from native areas which are not so much affected by extense turbulences of genetic backgrounds throughout China by human activities. Real natural populations from rural and also forest areas were never studied, since the sampling included mostly specimens from urban infestations, which might be invasive inside China themselves (Carter et al. 2009c; Carter et al. 2010; Javal et al. 2019b; Lee et al. 2020). Even with high-resolution methods like GBS, a clear picture of the introduction history would be strongly restricted by admixted source populations without the named data extension, as clearly observable with the European invasive SNP dataset of this study. Genome-wide SNPs from such populations would be critical to understand the ancestral genetic structure from Chinese source populations. Furthermore, after investigating the genome-wide SNPs on a

global *A. glabripennis*, multiple variable loci which reflect geographic population structure could be determined and used for a multilocus amplicon sequencing approach (e.g. *Anopheles* genus, Makunin et al. 2022) to assign specimens. Based on the results of Cui et al. 2022, a panel for amplicon sequencing was developed on *A. glabripennis* GBS data from North America and Asia by the Canadian bioSAFE project (https://www.biosafegenomics.com/) and it was subsequently tested with the European samples from this study if it is also feasible with unknown data (GBS and Amplicon-seq data unpublished, in prep).

Since *A. chinensis* is also invasive in Europe (EPPO 2021c), the inclusion of this species into large-scale GBS approaches in the overlapping native and invasive areas would be another important future research. Thereby, common introduction patterns and import routes could be further specified to unmask weak points, power of certain measures and to improve global efforts to maintain plant health around the world. However, especially overlapping distribution in the native range of these two species, the comparison of genome-wide SNPs would help to identify introgression, which might strongly alter the invasiveness when certain essential traits are enhanced by that (Wang and Keena 2021). The follow-up project of this study, PHID-Coleo II, will address this issue by investigating the intraspecific differentation of the A2 list (present in EPPO region) quarantine pests *Aromia bungii* and *Anoplophora chinensis* (EPPO 2021c) which will be sampled from native and invasive areas.

References

Aguilera A (2002) The connection between transcription and genomic instability. The EMBO Journal 21:195–201

Al Rawi S, Louvet-Vallée S, Djeddi A, Sachse M, Culetto E, Hajjar C, Boyd L, Legouis R, Galy V (2011) Postfertilization autophagy of sperm organelles prevents paternal nitochondrial DNA transmission. Science 334:1141–1144. https://doi.org/10.1126/science.1210333

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. https://doi.org/10.1101/gr.094052.109

Allen E, Noseworthy M, Ormsby M (2017) Phytosanitary measures to reduce the movement of forest pests with the international trade of wood products. Biol Invasions 19:3365–3376. https://doi.org/10.1007/s10530-017-1515-0

Allendorf FW, Lundquist LL (2003) Introduction: Population Biology, Evolution, and Control of Invasive Species. Conservation Biology 17:24–30. https://doi.org/10.1046/j.1523-1739.2003.02365.x

Armbruster P, Bradshaw WE, Holzapfel CM (1998) Effects of Postglacial Range Expansion on Allozyme and Quantitative Genetic Variation of the Pitcher-Plant Mosquito, Wyeomyia smithii. Evolution 52:1697–1704. https://doi.org/10.1111/j.1558-5646.1998.tb02249.x

Arvidsson S, Fartmann B, Winkler, Silke, Zimmermann, Wolfgang (2016) Efficient high-throughput SNP discovery and genotyping using normalised Genotyping-by-Sequencing (nGBS). LGC genomics, Berlin, Germany: Technical note AN-161104.01

Asner GP, Vitousek PM (2005) Remote analysis of biological invasion and biogeochemical change. Proc Natl Acad Sci U S A 102:4383–4386. https://doi.org/10.1073/pnas.0500823102

Avise JC (2000) Phylogeography: The history and formation of species. Harvard University Press, Cambridge Mass.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729–744. https://doi.org/10.1046/j.1365-294x.2003.02063.x

Barcode of life database (2021) Bold Systems v4. https://www.boldsystems.org/. Accessed 30 March 2021

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends Ecol Evol 23:38–44. https://doi.org/10.1016/j.tree.2007.09.008

Batista PD, Janes JK, Boone CK, Murray BW, Sperling FAH (2016) Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (Dendroctonus ponderosae). Ecol Evol 6:6292–6300. https://doi.org/10.1002/ece3.2367

Bayrische Landesanstalt für Landwirtschaft, LfL (2020) Asiatischer Laubholzbockkäfer – Verlängerung der Allgemeinverfügung der Quarantänezone Ziemetshausen-Schönebach bis Jahresende 2022 notwendig; Pressemitteilung 18.12.2020, Ziemetshausen. https://www.lfl.bayern.de/verschiedenes/presse/pms/2020/262041/index.php. Accessed 3 December 2021

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–2035. https://doi.org/10.1093/genetics/162.4.2025

Behura SK (2007) Analysis of nuclear copies of mitochondrial sequences in honeybee (Apis mellifera) genome. Mol Biol Evol 24:1492–1505. https://doi.org/10.1093/molbev/msm068

Bellard C, Cassey P, Blackburn TM (2016) Alien species as a driver of recent extinctions. Biol Lett 12:20150623. https://doi.org/10.1098/rsbl.2015.0623

Bensasson D, Zhang DX, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol Biol Evol 17:406–415. https://doi.org/10.1093/oxfordjournals.molbev.a026320

Bensasson D, Petrov DA, Zhang D-X, Hartl DL, Hewitt GM (2001a) Genomic Gigantism: DNA Loss Is Slow in Mountain Grasshoppers. Mol Biol Evol 18:246–253

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001b) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16:314–321. https://doi.org/10.1016/S0169-5347(01)02151-6

Bensasson D, Feldman MW, Petrov DA (2003) Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol 57:343–354. https://doi.org/10.1007/s00239-003-2485-7

Bernt M, Braband A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol 69:328–338. https://doi.org/10.1016/j.ympev.2012.10.020

Bertelsmeier C, Keller L (2018) Bridgehead Effects and Role of Adaptive Evolution in Invasive Populations. Trends Ecol Evol 33:527–534. https://doi.org/10.1016/j.tree.2018.04.014

Bertelsmeier C, Ollier S (2021) Bridgehead effects distort global flows of alien species. Divers Distrib 27:2180–2189. https://doi.org/10.1111/ddi.13388

Bertelsmeier C, Ollier S, Liebhold AM, Brockerhoff EG, Ward D, Keller L (2018) Recurrent bridgehead effects accelerate global alien ant spread. Proc Natl Acad Sci U S A 115:5486–5491. https://doi.org/10.1073/pnas.1801990115

Berthier K, Chapuis M-P, Moosavi SM, Tohidi-Esfahani D, Sword GA (2011) Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. Systematic Entomology 36:285–299. https://doi.org/10.1111/j.1365-3113.2010.00561.x

Bilodeau P, Roe AD, Bilodeau G, Blackburn GS, Cui M, Cusson M, Doucet D, Griess VC, Lafond VMA, Nilausen C, Paradis G, Porth I, Prunier J, Srivastava V, Stewart D, Torson AS, Tremblay E, Uzunovic A, Yemshanov D, Hamelin RC (2019) Biosurveillance of forest insects: part II—adoption of genomic tools by end user communities and barriers to integration. J Pest Sci 92:71–82. https://doi.org/10.1007/s10340-018-1001-1

bioSAFE project BioSurveillance of Alien Forest Enemies. https://www.biosafegenomics.com/. Accessed 15 March 2022

Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. Trends Ecol Evol 26:333–339. https://doi.org/10.1016/j.tree.2011.03.023

Blackburn TM, Lockwood JL, Cassey P (2015) The influence of numbers on invasion success. Mol Ecol 24:1942–1953. https://doi.org/10.1111/mec.13075

Blackburn GS, Bilodeau P, Cooke T, Cui M, Cusson M, Hamelin RC, Keena MA, Picq S, Roe AD, Shi J, Wu Y, 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. https://doi.org/10.1093/aesa/saz072

Blanchard JL, Schmidt GW (1996) Mitochondrial DNA Migration Events in Yeast and Humans: Integration bY a Common End-joining Mechanism and Alternative Perspectives on Nucleotide Substitution Patterns. Mol Biol Evol 13:537–548

BLAST (2021) BLAST: Basic Local Alignment Search Tool. https://blast.ncbi.nlm.nih.gov/Blast.cgi. Accessed 27 August 2021

Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney KD, Rieseberg LH (2015) What we still don't know about invasion genetics. Mol Ecol 24:2277–2297. https://doi.org/10.1111/mec.13032

Bourlat SJ, Nielsen C, Economou AD, Telford MJ (2008) Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol Phylogenet Evol 49:23–31. https://doi.org/10.1016/j.ympev.2008.07.008

Boykin LM, Armstrong KF, Kubatko L, Barro P de (2012) Species delimitation and global biosecurity. Evol Bioinform Online 8:1–37. https://doi.org/10.4137/EBO.S8532

Brown WM, Goerge M, JR., Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76:1967–1971

Budde KB, Nielsen LR, Ravn HP, Kjær ED (2016) The Natural Evolutionary Potential of Tree Populations to Cope with Newly Introduced Pests and Pathogens—Lessons Learned From Forest Health Catastrophes in Recent Decades. Curr Forestry Rep 2:18–29. https://doi.org/10.1007/s40725-016-0029-9

Bullock JM, Bonte D, Pufal G, Da Silva Carvalho C, Chapman DS, García C, García D, Matthysen E, Delgado MM (2018) Human-Mediated Dispersal and the Rewiring of Spatial Networks. Trends Ecol Evol 33:958–970. https://doi.org/10.1016/j.tree.2018.09.008

BWA-MEM version 0.7.12 (Burrows-Wheeler Aligner). http://bio-bwa.sourceforge.net/. Accessed 6 December 2021

CABI (2021) Anoplophora glabripennis (Asian longhorned beetle) datasheet. https://www.cabi.org/isc/datasheet/5557. Accessed 12 January 2022

Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). Mitochondrion 11:246–254. https://doi.org/10.1016/j.mito.2010.10.004

Campbell CL, Thorsness PE (1998) Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. J Cell Sci 111 (Pt 16):2455–2464. https://doi.org/10.1242/jcs.111.16.2455

Campbell EO, Brunet BMT, Dupuis JR, Sperling FAH (2018) Would an RRS by any other name sound as RAD ? Methods Ecol Evol 9:1920–1927. https://doi.org/10.1111/2041-210X.13038

Cao L-J, Wei S-J, Hoffmann AA, Wen J-B, Chen M (2016) Rapid genetic structuring of populations of the invasive fall webworm in relation to spatial expansion and control campaigns. Diversity Distrib. 22:1276–1287. https://doi.org/10.1111/ddi.12486

Carter M, Casa AM, Zeid M, Mitchell SE, Kresovich S (2009a) Isolation and characterization of microsatellite loci for the Asian longhorned beetle, Anoplophora glabripennis. Mol Ecol Resour 9:925–928. https://doi.org/10.1111/j.1755-0998.2008.02486.x

Carter M, Smith M, Harrison R (2010) Genetic analyses of the Asian longhorned beetle (Coleoptera, Cerambycidae, Anoplophora glabripennis), in North America, Europe and Asia. Biol Invasions 12:1165–1182. https://doi.org/10.1007/s10530-009-9538-9

Carter ME, Smith MT, Turgeon JJ, Harrison RG (2009b) Analysis of genetic diversity in an invasive population of Asian long-horned beetles in Ontario, Canada. Can Entomol 141:582–594. https://doi.org/10.4039/n09-026

Carter ME, Smith MT, Harrison RG (2009c) Patterns of Genetic Variation Among Populations of the Asian Longhorned Beetle (Coleoptera: Cerambycidae) in China and Korea. an 102:895–905. https://doi.org/10.1603/008.102.0516

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. Mol Ecol 22:2864–2883. https://doi.org/10.1111/mec.12330

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8

Chown SL, Hodgins KA, Griffin PC, Oakeshott JG, Byrne M, Hoffmann AA (2015) Biological invasions, climate change and genomics. Evol Appl 8:23–46. https://doi.org/10.1111/eva.12234

Ciampitti M, Cavagna B (2014) Public awareness: a useful tool for the early detection and a successful eradication of the longhorned beetles Anoplophora chinensis and A. glabripennis. EPPO Bull 44:248–250. https://doi.org/10.1111/epp.12116

Clifton EH, Gardescu S, Behle RW, Hajek AE (2020) Optimizing Application Rates of Metarhizium brunneum (Hypocreales: Clavicipitaceae) Microsclerotia for Infecting the Invasive Asian Longhorned Beetle (Coleoptera: Cerambycidae). J Econ Entomol 113:2650–2656. https://doi.org/10.1093/jee/toaa222

Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. Mol Ecol 24:1999–2017. https://doi.org/10.1111/mec.13162

Collins RA, Armstrong KF, Meier R, Yi Y, Brown SDJ, Cruickshank RH, Keeling S, Johnston C (2012) Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. PLoS One 7:e28381. https://doi.org/10.1371/journal.pone.0028381

Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. Mol Ecol Resour 13:969–975. https://doi.org/10.1111/1755-0998.12046

Connallon T, Sgrò CM (2018) In search of a general theory of species' range evolution. PLoS Biol 16:e2006735. https://doi.org/10.1371/journal.pbio.2006735

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719. https://doi.org/10.1093/bioinformatics/btn514

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, 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 30:1187–1189. https://doi.org/10.1093/bioinformatics/btt763

Cox AJ, Hebert PDN (2001) Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. Mol Ecol 10:371–386

Coyle DR, Trotter RT, Bean MS, Pfister SE (2021) First Recorded Asian Longhorned Beetle (Coleoptera: Cerambycidae) Infestation in the Southern United States. Journal of Integrated Pest Management 12. https://doi.org/10.1093/jipm/pmab007

Crease TJ (1999) The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera: Crustacea). Gene 233:89–99

Cristescu ME (2015) Genetic reconstructions of invasion history. Mol Ecol 24:2212–2225. https://doi.org/10.1111/mec.13117

Cristiano MP, Fernandes-Salomão TM, Yotoko KSC (2012) Nuclear mitochondrial DNA: an Achilles' heel of molecular systematics, phylogenetics, and phylogeographic studies of stingless bees. Apidologie 43:527–538. https://doi.org/10.1007/s13592-012-0122-4

Crooks JA (2005) Lag times and exotic species: The ecology and management of biological invasions in slow-motion1. Écoscience 12:316–329. https://doi.org/10.2980/i1195-6860-12-3-316.1

Crozier RH, Crozier YC (1993) The Mitochondrial Genome of the Honeybee Apis mellifera: Complete Sequence and Genome Organization. Genetics 133:97–117

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

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darling JA, Blum MJ (2007) DNA-based methods for monitoring invasive species: a review and prospectus. Biol Invasions 9:751–765. https://doi.org/10.1007/s10530-006-9079-4

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. https://doi.org/10.1038/nrg3012

Davidson AM, Jennions M, Nicotra AB (2011) Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. Ecol Lett 14:419–431. https://doi.org/10.1111/j.1461-0248.2011.01596.x

Dayama G, Emery SB, Kidd JM, Mills RE (2014) The genomic landscape of polymorphic human nuclear mitochondrial insertions. Nucleic Acids Res 42:12640–12649. https://doi.org/10.1093/nar/gku1038

Dhandapani RK, Gurusamy D, Duan JJ, Palli SR (2020) RNAi for management of Asian long-horned beetle, Anoplophora glabripennis: identification of target genes. J Pest Sci 93:823–832. https://doi.org/10.1007/s10340-020-01197-8

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431–449. https://doi.org/10.1111/j.1365-294X.2007.03538.x

Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. Mol Ecol 24:2095–2111. https://doi.org/10.1111/mec.13183

Doorenweerd C, San Jose M, Barr N, Leblanc L, Rubinoff D (2020) Highly variable COI haplotype diversity between three species of invasive pest fruit fly reflects remarkably incongruent demographic histories. Sci Rep 10:6887. https://doi.org/10.1038/s41598-020-63973-x

Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. Proc Natl Acad Sci U S A 109:E2183-91. https://doi.org/10.1073/pnas.1209329109

Dowle EJ, Bracewell RR, Pfrender ME, Mock KE, Bentz BJ, Ragland GJ (2017) Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (Dendroctonus ponderosae). Mol Ecol 26:6071–6084. https://doi.org/10.1111/mec.14342

Dupuis JR, Roe AD, Sperling FAH (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Mol Ecol 21:4422–4436. https://doi.org/10.1111/j.1365-294X.2012.05642.x

Early R, Bradley BA, Dukes JS, Lawler JJ, Olden JD, Blumenthal DM, Gonzalez P, Grosholz ED, Ibañez I, Miller LP, Sorte CJB, Tatem AJ (2016) Global threats from invasive alien species in the twenty-first century and national response capacities. Nat Commun 7:12485. https://doi.org/10.1038/ncomms12485

EFSA, Baker R, Gilioli G, Behring C, Candiani D, Gogin A, Kaluski T, Kinkar M, Mosbach-Schulz O, Neri, Franco, Maria, Preti S, Rosace, Maria, Chiara, Siligato R, Stancanelli G, Tramontini S (2019a) Anoplophora glabripennis: Pest Report and Datasheet to support ranking of EU candidate, priority pests. European Food Safety Authority (EFSA). https://doi.org/10.5281/zenodo.2786189

EFSA, Hoppe B, Schrader G, Kinkar M, Vos S (2019b) Pest survey card on Anoplophora glabripennis - European Food Safety Authority (EFSA). EFSA Supporting Publications 16. https://doi.org/10.2903/sp.efsa.2019.EN-1750

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (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

Elshire RJ, Glaubitz JC, Sun Q, Harriman James V., Buckler ES, Mitchell SE (2014) Genotyping by sequencing (GBS) method overview. https://biohpc.cornell.edu/lab/doc/gbs_method_overview1.pdf. Accessed 24 February 2022

EPPO (2013) PM 9/15 (1) Anoplophora glabripennis: procedures for official control. EPPO Bull 43:510–517. https://doi.org/10.1111/epp.12064

EPPO (2016) PM 7/129 (1) DNA barcoding as an identification tool for a number of regulated pests. EPPO Bull 46:501–537. https://doi.org/10.1111/epp.12344

EPPO (2021a) Anoplophora glabripennis. EPPO datasheets on pests recommended for regulation. Available online. https://gd.eppo.int; [accessed on 15-10-2021]. https://gd.eppo.int/taxon/ANOLGL/datasheet. Accessed 15 October 2021

EPPO (2021b) EPPO Global Database - Reportings Anoplophora glabripennis (ANOLGL). https://gd.eppo.int/taxon/ANOLGL/reporting. Accessed 23 December 2021

EPPO (2021c) EPPO Standard PM 1/2(8) - EPPO A1 and A2 lists of pests recommended for regulations as quarantine pests

Eritja R, Palmer JRB, Roiz D, Sanpera-Calbet I, Bartumeus F (2017) Direct Evidence of Adult Aedes albopictus Dispersal by Car. Sci Rep 7:14399. https://doi.org/10.1038/s41598-017-12652-5

Estoup A, Baird SJE, Ray N, Currat M, Cornuet J-M, Santos F, Beaumont MA, Excoffier L (2010) Combining genetic, historical and geographical data to reconstruct the dynamics of bioinvasions: application to the cane toad Bufo marinus. Mol Ecol Resour 10:886–901. https://doi.org/10.1111/j.1755-0998.2010.02882.x

Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B (2016) Is There a Genetic Paradox of Biological Invasion? Annu. Rev. Ecol. Evol. Syst. 47:51–72. https://doi.org/10.1146/annurev-ecolsys-121415-032116

EU (2015) Commission implementing decision (EU) 2015/ 893 - of 9 June 2015 - as regards measures to prevent the introduction into and the spread within the Union of Anoplophora glabripennis (Motschulsky) - (notified under document C(2015) 3772). Official Journal of the European Union:1–13

EU (2019) Commission Implementing Regulation (EU) 2019/ of 28 November 2019 establishing uniform conditions for the implementation of Regulation (EU) 2016/2031 of the European Parliament and the Council, as regards protective measures against pests of plants, and repealing Commission Regulation (EC) No 690/2008 and amending

Commission Implementing Regulation (EU) 2018/2019. Official Journal of the European Union:1-279

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Eyre D, Barbrook J (2021) The eradication of Asian longhorned beetle at Paddock Wood, UK. CABI Agric Biosci 2. https://doi.org/10.1186/s43170-021-00034-x

Eyre D, Haack RA (2017) Invasive Cerambycid pests and biosecurity measures. Chapter 13. Invasive Cerambycid Pests and Biosecurity Measures

Eyre D, Macarthur R, Haack RA, Lu Y, Krehan H (2018) Variation in Inspection Efficacy by Member States of Wood Packaging Material Entering the European Union. J Econ Entomol 111:707–715. https://doi.org/10.1093/jee/tox357

Faccoli M, Favaro R, Smith MT, Wu J (2015) Life history of the Asian longhorn beetle Anoplophora glabripennis (Coleoptera Cerambycidae) in southern Europe. Agr Forest Entomol 17:188–196. https://doi.org/10.1111/afe.12096

Faccoli M, Favaro R (2016) Host preference and host colonization of the Asian long-horned beetle, Anoplophora glabripennis (Coleoptera Cerambycidae), in Southern Europe. Bull Entomol Res 106:359–367. https://doi.org/10.1017/S0007485315001157

Facon B, Pointier J-P, Glaubrecht M, Poux C, Jarne P, David P (2003) A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. Mol Ecol 12:3027–3039. https://doi.org/10.1046/j.1365-294x.2003.01972.x

Facon B, Genton BJ, Shykoff J, Jarne P, Estoup A, David P (2006) A general eco-evolutionary framework for understanding bioinvasions. Trends Ecol Evol 21:130–135. https://doi.org/10.1016/j.tree.2005.10.012

Facon B, Hufbauer RA, Tayeh A, Loiseau A, Lombaert E, Vitalis R, Guillemaud T, Lundgren JG, Estoup A (2011) Inbreeding depression is purged in the invasive insect Harmonia axyridis. Curr Biol 21:424–427. https://doi.org/10.1016/j.cub.2011.01.068

Fang J, Qian L, Xu M, Yang X, Wang B, An Y (2016) The complete nucleotide sequence of the mitochondrial genome of the Asian longhorn beetle, Anoplophora glabripennis (Coleoptera: Cerambycidae). Mitochondrial DNA A DNA Mapp Seq Anal 27:3299–3300. https://doi.org/10.3109/19401736.2015.1015012

FAO (2019) Food and Agriculture Organization (FAO) of the United Nations - ISPM 15. Regulation of wood packaging material in international trade

FAO (2021) Food and Agriculture Organization (FAO) of the United Nations - ISPM 5. Glossary of phytosanitary terms. International Plant Protection Convention (IPPC)

Favaro R, Wichmann L, Ravn HP, Faccoli M (2015) Spatial spread and infestation risk assessment in the Asian longhorned beetle, Anoplophora glabripennis. Entomol Exp Appl 155:95–101. https://doi.org/10.1111/eea.12292

Fernández DC, VanLaerhoven SL, Rodríguez-Leyva E, Zhang YM, Labbé R (2022) Population Structure and Genetic Diversity of the Pepper Weevil (Coleoptera: Curculionidae) Using the COI Barcoding Region. J Insect Sci 22. https://doi.org/10.1093/jisesa/ieac012

Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about biological invasions? Biol Invasions 14:245–253. https://doi.org/10.1007/s10530-011-0064-1

Folmer O, Black M, Hoeh W, Lutz, R. Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3:294–299. https://doi.org/10.1371/journal.pone.0013102

Fontanesi F (2015) Mitochondria: Structure and Role in Respiration. In: eLS. John Wiley & Sons, Ltd: Chichester.:1–13. https://doi.org/10.1002/9780470015902.a0001380.pub2

Forster B, Wermelinger B (2012) First records and reproductions of the Asian longhorned beetle Anoplophora glabripennis (Motschulsky) (Coleoptera, Cerambycidae) in Switzerland. Journal of the Swiss Entomological Society 85:267–275

Freebayes v1.0.2-16 (Bayesian haplotype-based genetic polymorphism discovery and genotyping). https://github.com/freebayes/freebayes#readme. Accessed 6 December 2021

Garnas JR, Auger-Rozenberg M-A, Roques A, Bertelsmeier C, Wingfield MJ, Saccaggi DL, Roy HE, Slippers B (2016) Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. Biol Invasions 18:935–952. https://doi.org/10.1007/s10530-016-1082-9

Gippet JM, Liebhold AM, Fenn-Moltu G, Bertelsmeier C (2019) Human-mediated dispersal in insects. Curr Opin Insect Sci 35:96–102. https://doi.org/10.1016/j.cois.2019.07.005

González VL, Devine AM, Trizna M, Mulcahy DG, Barker KB, Coddington JA (2018) Open access genomic resources for terrestrial arthropods. Curr Opin Insect Sci 25:91–98. https://doi.org/10.1016/j.cois.2017.12.003

Gordo I, Charlesworth B (2000) The Degeneration of Asexual Haploid Populations and the Speed of Muller's Ratchet. Genetics 154:1379–1387

Gould JR, Aflague B, Murphy TC, McCartin L, Elkinton JS, Rim K, Duan JJ (2018) Collecting Nontarget Wood-Boring Insects for Host-Specificity Testing of Natural Enemies of Cerambycids: A Case Study of Dastarcus helophoroides (Coleoptera: Bothrideridae), a

Parasitoid of the Asian Longhorned Beetle (Coleoptera: Cerambycidae). Environ Entomol 47:1440–1450. https://doi.org/10.1093/ee/nvy121

Grapputo A, Boman S, Lindström L, Lyytinen A, Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. Mol Ecol 14:4207–4219. https://doi.org/10.1111/j.1365-294X.2005.02740.x

Greiner S, Sobanski J, Bock R (2015) Why are most organelle genomes transmitted maternally? Bioessays 37:80–94. https://doi.org/10.1002/bies.201400110

Haack RA (2006) Exotic bark- and wood-boring Coleoptera in the United States: recent establishments and interceptions. Can. J. For. Res. 36:269–288. https://doi.org/10.1139/x05-249

Haack RA, Cavey JF, Hoebeke ER, Law KR (1996) Anoplophora glabripennis: A new tree-infesting exotic cerambycid invades New York. Michigan Entomological Society 41:1–3

Haack RA, Law KR, Mastro VC, Ossenbruggen SH, Raimo BJ (1997) New York's Battle with the Asian Long-Horned Beetle. Journal of Forestry 95

Haack RA, Bauer LS, Gao R, McCarthy JJ, Miller DL, Petrice TR, Poland TM (2006) Anoplophora glabripennis within-tree distribution, seasonal development, and host suitability in China and Chicago. The Great Lakes Entomologist 39

Haack RA, Hérard F, Sun J, Turgeon JJ (2010) Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. Annu Rev Entomol 55:521–546. https://doi.org/10.1146/annurev-ento-112408-085427

Haack RA, Britton KO, Brockerhoff EG, Cavey JF, Garrett LJ, Kimberley M, Lowenstein F, Nuding A, Olson LJ, Turner J, Vasilaky KN (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:e96611. https://doi.org/10.1371/journal.pone.0096611

Hänel C, Chown SL (1998) The impact of a small, alien invertebrate on a sub-Antarctic terrestrial ecosystem: Limnophyes minimus (Diptera, Chironomidae) at Marion Island. Polar Biology 20:99-106. https://doi.org/10.1007/s003000050282

Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G (2015) Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. J Zoolog Syst Evol Res 53:109–115. https://doi.org/10.1111/jzs.12095

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174. https://doi.org/10.1007/BF02101694

Hastings IM (1992) Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genet Res 59:215–225. https://doi.org/10.1017/s0016672300030500

Hazkani-Covo E (2009) Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol Biol Evol 26:2175–2179. https://doi.org/10.1093/molbev/msp131

Hazkani-Covo E, Graur D (2007) A comparative analysis of numt evolution in human and chimpanzee. Mol Biol Evol 24:13–18. https://doi.org/10.1093/molbev/msl149

Hazkani-Covo E, Sorek R, Graur D (2003) Evolutionary Dynamics of Large Numts in the Human Genome: Rarity of Independent Insertions and Abundance of Post-Insertion Duplications. J Mol Evol 56:169–174

Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet 6:e1000834. https://doi.org/10.1371/journal.pgen.1000834

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270:313–321. https://doi.org/10.1098/rspb.2002.2218

Hérard F, Ciampitti M, Maspero M, Krehan H, Benker U, Boegel C, Schrage R, Bouhot-Delduc L, Bialooki P (2006) Anoplophora species in Europe: infestations and management processes. OEPP/EPPO Bulletin 36:470–474. https://doi.org/10.1111/j.1365-2338.2006.01046.x

Hérard F, Maspero M, Ramualde N, Jucker C, Colombo M, Ciampitti M, Cavagna B (2009) Anoplophora glabripennis infestation (col.: cerambycidae) in Italy. OEPP/EPPO Bulletin 39:146–152

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352. https://doi.org/10.1534/genetics.104.036947

Heumann-Kiesler C, Sommerfeld V, Iffland H, Bennewitz J, Rodehutscord M, Hasselmann M (2021) Insights into the Mitochondrial and Nuclear Genome Diversity of Two High Yielding Strains of Laying Hens. Animals (Basel) 11. https://doi.org/10.3390/ani11030825

Hill WG (2007) Impact of selection on effective population size: a commentary on 'Inbreeding in artificial selection programmes' by Alan Robertson. Genet Res 89:273–274. https://doi.org/10.1017/S0016672308009464

Hoyer-Tomiczek U, Sauseng G, Hoch G (2016) Scent detection dogs for the Asian longhorn beetle, Anoplophora glabripennis. EPPO Bull 46:148–155. https://doi.org/10.1111/epp.12282

Hu J, Angeli S, Schuetz S, Luo Y, Hajek AE (2009) Ecology and management of exotic and endemic Asian longhorned beetle Anoplophora glabripennis. Agricultural and Forest Entomology 11:359–375. https://doi.org/10.1111/j.1461-9563.2009.00443.x

Hufbauer RA, Facon B, Ravigné V, Turgeon J, Foucaud J, Lee CE, Rey O, Estoup A (2012) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. Evol Appl 5:89–101. https://doi.org/10.1111/j.1752-4571.2011.00211.x

Hulme PE, Bacher S, Kenis M, Klotz S, Kühn I, Minchin D, Nentwig W, Olenin S, Panov V, Pergl J, Pyšek P, Roques A, Sol D, Solarz W, Vilà M (2008) Grasping at the routes of biological invasions: a framework for integrating pathways into policy. Journal of Applied Ecology 45:403–414. https://doi.org/10.1111/j.1365-2664.2007.01442.x

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

Hulme PE (2015) Invasion pathways at a crossroad: policy and research challenges for managing alien species introductions. J Appl Ecol 52:1418–1424. https://doi.org/10.1111/1365-2664.12470

Hulme PE (2021) Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. One Earth 4:666–679. https://doi.org/10.1016/j.oneear.2021.04.015

Hurst LD (1995) Selfish genetic elements and their role in evolution: the evolution of sex and some of what that entails. Philos Trans R Soc Lond B Biol Sci 349:321–332. https://doi.org/10.1098/rstb.1995.0120

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267. https://doi.org/10.1093/molbev/msj030

Ibáñez I, Diez JM, Miller LP, Olden JD, Sorte CJB, Blumenthal DM, Bradley BA, D'Antonio CM, Dukes JS, Early RI, Grosholz ED, Lawler JJ (2014) Integrated assessment of biological invasions. Ecological Applications 24:25–37. https://doi.org/10.1890/13-0776.1

Javal M (2017) Invasion du capricorne asiatique Anoplophora glabripennis: génétique, traits d'histoire de vie et écologie. Dissertation University of Orléans

Javal M, Roux G, Roques A, Sauvard D (2018) Asian Long-horned Beetle dispersal potential estimated in computer-linked flight mills. J Appl Entomol 142:282–286. https://doi.org/10.1111/jen.12408

Javal M, Roques A, Haran J, Hérard F, Keena M, Roux G (2019a) Complex invasion history of the Asian long-horned beetle: fifteen years after first detection in Europe. J Pest Sci 92:173–187. https://doi.org/10.1007/s10340-017-0917-1

Javal M, Lombaert E, Tsykun T, Courtin C, Kerdelhué C, Prospero S, Roques A, Roux G (2019b) Deciphering the worldwide invasion of the Asian long-horned beetle: A recurrent invasion process from the native area together with a bridgehead effect. Mol Ecol 28:951–967. https://doi.org/10.1111/mec.15030

Jordal BH, Kambestad M (2014) DNA barcoding of bark and ambrosia beetles reveals excessive NUMTs and consistent east-west divergence across Palearctic forests. Mol Ecol Resour 14:7–17. https://doi.org/10.1111/1755-0998.12150

José Cerca (2021a) Guideline for LD pruning. https://github.com/jcerca/population_genomics/tree/master/LD_prunning_HowTo. Accessed 6 December 2021

José Cerca (2021b) R script for plotting VCFtool outputs. https://github.com/jcerca/population_genomics/tree/master/SNPcalling_GATK. Accessed 6 December 2021

Keena MA (2002) Anoplophora glabripennis (Coleoptera: Cerambycidae) Fecundity and Longevity Under Laboratory Conditions: Comparison of Populations from New York and Illinois on Acer saccharum. Environ Entomol 31:490–498. https://doi.org/10.1603/0046-225X-31.3.490

Keena MA (2005) Pourable Artificial Diet for Rearing *Anoplophora glabripennis* (Coleoptera: Cerambycidae) and Methods to Optimize Larval Survival and Synchronize Development. an 98:536–547. https://doi.org/10.1603/0013-8746(2005)098[0536:PADFRA]2.0.CO;2

Keena MA (2006) Effects of Temperature on Anoplophora glabripennis (Coleoptera: Cerambycidae) Adult Survival, Reproduction, and Egg Hatch. Environ Entomol 35:912–921. https://doi.org/10.1603/0046-225X-35.4.91

Keightley PD, Ness RW, Halligan DL, Haddrill PR (2014) Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics 196:313–320. https://doi.org/10.1534/genetics.113.158758

Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. Ecol Lett 11:852–866. https://doi.org/10.1111/j.1461-0248.2008.01188.x

Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. Journal of Evolutionary Biology 23:1720–1731. https://doi.org/10.1111/j.1420-9101.2010.02037.x

Kelly DW, Muirhead JR, Heath DD, MacIsaac HJ (2006) Contrasting patterns in genetic diversity following multiple invasions of fresh and brackish waters. Mol Ecol 15:3641–3653. https://doi.org/10.1111/j.1365-294X.2006.03012.x

Kethidi DR, Roden DB, Ladd TR, Krell PJ, Retnakaran A, Feng Q (2003) Development of SCAR markers for the DNA-based detection of the Asian long-horned beetle,

Anoplophora glabripennis (Motschulsky). Arch Insect Biochem Physiol 52:193–204. https://doi.org/10.1002/arch.10082

Kim M-S, Kim KY, Lee W, Hong K-J (2019) Reporting the possibility of invasive populations of Anoplophora spp. in Korea. Journal of Asia-Pacific Biodiversity 12:211–216. https://doi.org/10.1016/j.japb.2019.01.007

Kjærandsen J (2022) Current State of DNA Barcoding of Sciaroidea (Diptera)—Highlighting the Need to Build the Reference Library. Insects 13:147. https://doi.org/10.3390/insects13020147

Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 60:115–138. https://doi.org/10.1146/annurev.arplant.043008.092119

Knopp T, Cano JM, Crochet P-A, Merilä J (2007) Contrasting Levels of Variation in Neutral and Quantitative Genetic Loci on Island Populations of Moor Frogs (Rana arvalis). Conserv Genet 8:45–56. https://doi.org/10.1007/s10592-006-9147-4

Kolbe JJ, Glor RE, Rodríguez Schettino L, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. Nature 431:177–181. https://doi.org/10.1038/nature02807

Koutroumpa FA, Lieutier F, Roux-Morabito G (2009) Incorporation of mitochondrial fragments in the nuclear genome (Numts) of the longhorned beetle Monochamus galloprovincialis (Coleoptera, Cerambycidae). Journal of Zoological Systematics and Evolutionary Research 47:141–148. https://doi.org/10.1111/j.1439-0469.2008.00492.x

Krieger MJB, Ross KG (2002) Identification of a Major Gene Regulating Complex Social Behavior. Science 295. https://doi.org/10.1126/science.1065247

Kulhanek SA, Ricciardi A, Leung B (2011) Is invasion history a useful tool for predicting the impacts of the world's worst aquatic invasive species? Ecological Applications 21:189–202. https://doi.org/10.1890/09-1452.1

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Ladoukakis ED, Zouros E (2017) Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. J Biol Res (Thessalon) 24:2. https://doi.org/10.1186/s40709-017-0060-4

Lampert A, Liebhold AM (2021) Combining multiple tactics over time for cost-effective eradication of invading insect populations. Ecol Lett 24:279–287. https://doi.org/10.1111/ele.13640

Latombe G, Pyšek P, Jeschke JM, Blackburn TM, Bacher S, Capinha C, Costello MJ, Fernández M, Gregory RD, Hobern D, Hui C, Jetz W, Kumschick S, McGrannachan C,

Pergl J, Roy HE, Scalera R, Squires ZE, Wilson JR, Winter M, Genovesi P, McGeoch MA (2017) A vision for global monitoring of biological invasions. Biological Conservation 213:295–308. https://doi.org/10.1016/j.biocon.2016.06.013

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. The Proceedings of the National Academy of Sciences 104:3883–3888. https://doi.org/10.1073/pnas.0607324104

Lawson DJ, van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun 9:3258. https://doi.org/10.1038/s41467-018-05257-7

Leal I, Allen E, Humble L, Sela S, Uzunovic A (2010) Phytosanitary risks associated with the global movement of forest products: A commodity-based approach. Information report / Pacific Forestry Centre, BC-X-419. Natural Resources Canada Canadian Forest Service Pacific Forestry Centre, Victoria, BC

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

Lee CE, Gelembiuk GW (2008) Evolutionary origins of invasive populations. Evol Appl 1:427–448. https://doi.org/10.1111/j.1752-4571.2008.00039.x

Leigh JW, Bryant D (2015) popart : full-feature software for haplotype network construction. Methods Ecol Evol 6:1110–1116. https://doi.org/10.1111/2041-210X.12410

Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet 21:655–663. https://doi.org/10.1016/j.tig.2005.09.004

Lesieur V, Lombaert E, Guillemaud T, Courtial B, Strong W, Roques A, Auger-Rozenberg M-A (2019) The rapid spread of Leptoglossus occidentalis in Europe: a bridgehead invasion. J Pest Sci 92:189–200. https://doi.org/10.1007/s10340-018-0993-x

Leung B, Lodge DM, Finnoff D, Shogren JF, Lewis MA, Lamberti G (2002) An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proc Biol Sci 269:2407–2413. https://doi.org/10.1098/rspb.2002.2179

Liebhold AM, Tobin PC (2008) Population ecology of insect invasions and their management. Annu Rev Entomol 53:387–408. https://doi.org/10.1146/annurev.ento.52.110405.091401

Liebl AL, Schrey AW, Richards CL, Martin LB (2013) Patterns of DNA methylation throughout a range expansion of an introduced songbird. Integr Comp Biol 53:351–358. https://doi.org/10.1093/icb/ict007

Lingafelter SW, Hoebeke ER (eds) (2002) Revision of the Genus Anoplophora (Coleoptera: Cerambycidae)

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S (2017) Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. Mol Biol Evol 34:119–130. https://doi.org/10.1093/molbev/msw226

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:e9743. https://doi.org/10.1371/journal.pone.0009743

Loomans A, Wessels-Berk BF, Copini P, Mentink N, Hoop MB de, Hartog W den (2013) Import-inspections, surveys, detection and eradication of the longhorn beetles Anoplophora chinensis and A. glabripennis in The Netherlands. Journal of Entomological and Acarological Research 45:1

Lopez MLD, Lin Y-Y, Sato M, Hsieh C-H, Shiah F-K, Machida RJ (2022) Using metatranscriptomics to estimate the diversity and composition of zooplankton communities. Mol Ecol Resour 22:638–652. https://doi.org/10.1111/1755-0998.13506

Lopez VM, Hoddle MS, Francese JA, Lance DR, Ray AM (2017) Assessing Flight Potential of the Invasive Asian Longhorned Beetle (Coleoptera: Cerambycidae) With Computerized Flight Mills. J Econ Entomol 110:1070–1077. https://doi.org/10.1093/jee/tox046

Lowe S, Browne M, Boudjelas S, Poorter M de (2000) 100 of the world's worst invasive alien species: a selection from the global invasive species database. Aliens 12:12 pp.

LTZ (2013) Befallsanzeichen des Asiatischen Laubholzbocks (Anoplophora glabripennis) an Verpackungsholz. Landwirtschaftliches Technologiezentrum Augustenberg (LTZ)

LTZ (2016) Asiatischer Laubholzbockkäfer, Anoplophora glabripennis. Landwirtschaftliches Technologiezentrum Augustenberg (LTZ), Hinweise zur Pflanzengesundheit - Quarantäne Schadorganismen

Lynch M (2010) Evolution of the mutation rate. Trends Genet 26:345–352. https://doi.org/10.1016/j.tig.2010.05.003

Ma H, Xu H, O'Farrell PH (2014) Transmission of mitochondrial mutations and action of purifying selection in Drosophila melanogaster. Nat Genet 46:393–397. https://doi.org/10.1038/ng.2919

Makunin A, Korlević P, Park N, Goodwin S, Waterhouse RM, Wyschetzki K von, Jacob CG, Davies R, Kwiatkowski D, St Laurent B, Ayala D, Lawniczak MKN (2022) A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium

presence across the entire Anopheles genus. Mol Ecol Resour 22:28–44. https://doi.org/10.1111/1755-0998.13436

Malik AN, Czajka A, Cunningham P (2016) Accurate quantification of mouse mitochondrial DNA without co-amplification of nuclear mitochondrial insertion sequences. Mitochondrion 29:59–64. https://doi.org/10.1016/j.mito.2016.05.003

Mandal S de, Chhakchhuak L, Gurusubramanian G, Kumar NS (2014) Mitochondrial markers for identification and phylogenetic studies in insects – A Review. DNA Barcodes 2. https://doi.org/10.2478/dna-2014-0001

Mankin RW, Smith MT, Tropp JM, Atkinson EB, Jong DY (2008) Detection of Anoplophora glabripennis (Coleoptera: Cerambycidae) Larvae in Different Host Trees and Tissues by Automated Analyses of Sound-Impulse Frequency and Temporal Patterns. J Econ Entomol 101:838–849. https://doi.org/10.1093/jee/101.3.838

Mason CJ, Scully ED, Geib SM, Hoover K (2016) Contrasting diets reveal metabolic plasticity in the tree-killing beetle, Anoplophora glabripennis (Cerambycidae: Lamiinae). Sci Rep 6:33813. https://doi.org/10.1038/srep33813

Mason CJ, Campbell AM, Scully ED, Hoover K (2019) Bacterial and Fungal Midgut Community Dynamics and Transfer Between Mother and Brood in the Asian Longhorned Beetle (Anoplophora glabripennis), an Invasive Xylophage. Microb Ecol 77:230–242. https://doi.org/10.1007/s00248-018-1205-1

Mastin AJ, Gottwald TR, van den Bosch F, Cunniffe NJ, Parnell S (2020) Optimising risk-based surveillance for early detection of invasive plant pathogens. PLoS Biol 18:e3000863. https://doi.org/10.1371/journal.pbio.3000863

McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn S-J, Arsala D, Benoit JB, Blackmon H, Bledsoe T, Bowsher JH, Busch A, Calla B, Chao H, Childers AK, Childers C, Clarke DJ, Cohen L, Demuth JP, Dinh H, Doddapaneni H, Dolan A, Duan JJ, Dugan S, Friedrich M, Glastad KM, Goodisman MAD, Haddad S, Han Y, Hughes DST, Ioannidis P, Johnston JS, Jones JW, Kuhn LA, Lance DR, Lee C-Y, Lee SL, Lin H, Lynch JA, Moczek AP, Murali SC, Muzny DM, Nelson DR, Palli SR, Panfilio KA, Pers D, Poelchau MF, Quan H, Qu J, Ray AM, Rinehart JP, Robertson HM, Roehrdanz R, Rosendale AJ, Shin S, Silva C, Torson AS, Jentzsch IMV, Werren JH, Worley KC, Yocum G, Zdobnov EM, Gibbs RA, 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 Biol 17:227. https://doi.org/10.1186/s13059-016-1088-8

Meurisse N, Rassati D, Hurley BP, Brockerhoff EG, Haack RA (2019) Common pathways by which non-native forest insects move internationally and domestically. J Pest Sci 92:13–27. https://doi.org/10.1007/s10340-018-0990-0

Migliorini D, Ghelardini L, Tondini E, Luchi N, Santini A (2015) The potential of symptomless potted plants for carrying invasive soilborne plant pathogens. Diversity Distrib. 21:1218–1229. https://doi.org/10.1111/ddi.12347

Mitchell RF, Hall LP, Reagel PF, McKenna DD, Baker TC, Hildebrand JG (2017) Odorant receptors and antennal lobe morphology offer a new approach to understanding olfaction in the Asian longhorned beetle. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 203:99–109. https://doi.org/10.1007/s00359-016-1138-4

Morel-Journel T, Assa CR, Mailleret L, Vercken E (2019) Its all about connections: hubs and invasion in habitat networks. Ecol Lett 22:313–321. https://doi.org/10.1111/ele.13192

Moulton MJ, Song H, Whiting MF (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). Mol Ecol Resour 10:615–627. https://doi.org/10.1111/j.1755-0998.2009.02823.x

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The Human Genome Project Reveals a Continuous Transfer of Large Mitochondrial Fragments to the Nucleus. Mol Biol Evol 18:1833–1837. https://doi.org/10.31436/ijes.v4i2.129.g78

Moussa Z, Cocquempot C (2017) Anoplophora glabripennis Motschulsky, 1854, a new introduced pest that could threat hardwood trees in Lebanon (Coleoptera, Cerambycidae). Bulletin de la Société entomologique de France 122:501–508

Multiple Primer Analyzer (2021) Multiple Primer Analyzer, Thermo Fisher Scientific - DE. https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html. Accessed 30 March 2021

NCBI (2021) National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/. Accessed 30 March 2021

NCBI Agla 2.0 (Anoplophora glabripennis genome assembly Agla 2.0). https://www.ncbi.nlm.nih.gov/assembly/GCF_000390285.2. Accessed 6 December 2021

Nehme ME, Trotter RT, Keena MA, McFarland C, Coop J, Hull-Sanders HM, Meng P, Moraes CM de, Mescher MC, Hoover K (2014) Development and evaluation of a trapping system for Anoplophora glabripennis (Coleoptera: Cerambycidae) in the United States. Environ Entomol 43:1034–1044. https://doi.org/10.1603/EN14049

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press

Nei M, Maruyama T, Chakraborty R (1975) The Bottleneck Effect and Genetic Variability in Populations. Evolution 29:1–10

Neiman M, Taylor DR (2009) The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci 276:1201–1209. https://doi.org/10.1098/rspb.2008.1758

Nielsen R, Slatkin M (2013) An Introduction to Population Genetics: Theory and Applications. Sinauer Associates, Inc. Publishers Sunderland, Massachusetts U.S.A.

Nunes MDS, Dolezal M, Schlötterer C (2013) Extensive paternal mtDNA leakage in natural populations of Drosophila melanogaster. Mol Ecol 22:2106–2117. https://doi.org/10.1111/mec.12256

O'Dowd DJ, Green PT, Lake PS (2003) Invasional 'meltdown' on an oceanic island. Ecol Letters 6:812–817. https://doi.org/10.1046/j.1461-0248.2003.00512.x

Ohbayashi N, Ogawa J (2009) Phylogenetic Analysis of the Lamiine Genus Anoplophora and its Relatives (Coleoptera, Cerambycidae) Based on the Mitochondrial COI Gene. Spec Bull Jpn Soc Coleopterol, Tokyo 7:309–324

Oligo Calculator (2021) OligoCalc: Oligonucleotide Properties Calculator. http://biotools.nubic.northwestern.edu/OligoCalc.html. Accessed 30 March 2021

Ortiz EM, Cook DE, Seemann T (2021) Python script to convert a collection of SNPs in VCF format into a PHYLIP, FASTA, NEXUS, or binary NEXUS file for phylogenetic analysis. https://github.com/edgardomortiz/vcf2phylip/blob/master/vcf2phylip.py. Accessed 6 December 2021

Pajović I, Petrić D, Bellini R, Pajović L, Quarrie S (2017) First record of Anoplophora glabripennis (Coleoptera: Cerambycidae) in Montenegro. Baltic Forestry 23:706-710

Pamilo P, Viljakainen L, Vihavainen A (2007) Exceptionally high density of NUMTs in the honeybee genome. Mol Biol Evol 24:1340–1346. https://doi.org/10.1093/molbev/msm055

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Mol Biol Evol 27:1659–1672. https://doi.org/10.1093/molbev/msq051

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. Genetics 192:1065–1093. https://doi.org/10.1534/genetics.112.145037

Pejchar L, Mooney HA (2009) Invasive species, ecosystem services and human well-being. Trends Ecol Evol 24:497–504. https://doi.org/10.1016/j.tree.2009.03.016

Pennacchio F, Peverieri GS, Jucker C, Allegro G, Roversi PF (2012) A key for the identification of larvae of Anoplophora chinensis, Anoplophora glabripennis and Psacothea hilaris (Coleoptera Cerambycidae Lamiinae) in Europe. Redia Journal of Zoology 95:57–65

Pfrender ME, Spitze K, Hicks J, Morgan K, Latta L, Lynch M (2000) Lack of concordance between genetic diversity estimates at the molecular and quantitative-trait levels. Conservation Genetics 1:263–369. https://doi.org/10.1023/A:1011518125488

Physalia courses Home. https://www.physalia-courses.org/. Accessed 27 March 2022

Picq S, Keena M, Havill N, Stewart D, Pouliot E, Boyle B, Levesque RC, Hamelin RC, 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. Evol Appl 11:325–339. https://doi.org/10.1111/eva.12559

Pochon X, Bott NJ, Smith KF, Wood SA (2013) Evaluating Detection Limits of Next-Generation Sequencing for the Surveillance and Monitoring of International Marine Pests. PLoS One 8:e73935. https://doi.org/10.1371/journal.pone.0073935

Poland TM, Haack RA, Petrice TR (1998) Chicago joins New York in battle with the Asian longhorned beetle. NEWSLETTER of the MICHIGAN ENTOMOLOGICAL SOCIETY 43:15–17

Primer-BLAST (2021) Primer-BLAST - A tool for finding specific primers. https://www.ncbi.nlm.nih.gov/tools/primer-blast/. Accessed 30 March 2021

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1093/genetics/155.2.945

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

Purcell SM, Chang CC (2021) PLINK 1.9. https://www.cog-genomics.org/plink/. Accessed 30 March 2021

Puzey J, Vallejo-Marín M (2014) Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (Mimulus guttatus). Mol Ecol 23:4472–4485. https://doi.org/10.1111/mec.12875

Pyšek P, Richardson DM (2010) Invasive Species, Environmental Change and Management, and Health. Annu. Rev. Environ. Resour. 35:25–55. https://doi.org/10.1146/annurev-environ-033009-095548

Pyšek P, Hulme PE, Simberloff D, Bacher S, Blackburn TM, Carlton JT, Dawson W, Essl F, Foxcroft LC, Genovesi P, Jeschke JM, Kühn I, Liebhold AM, Mandrak NE, Meyerson LA, Pauchard A, Pergl J, Roy HE, Seebens H, van Kleunen M, Vilà M, Wingfield MJ, Richardson DM (2020) Scientists' warning on invasive alien species. Biol Rev Camb Philos Soc 95:1511–1534. https://doi.org/10.1111/brv.12627

R Core Team (2021) R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org. Accessed 30 March 2021

Rashid I, Campos M, Collier T, Crepeau M, Weakley A, Gripkey H, Lee Y, Schmidt H, Lanzaro GC (2022) Spontaneous mutation rate estimates for the principal malaria

vectors Anopheles coluzzii and Anopheles stephensi. Sci Rep 12:226. https://doi.org/10.1038/s41598-021-03943-z

Rašić G, Filipović I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. BMC Genomics 15:275

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes 7:355–364. https://doi.org/10.1111/j.1471-8286.2006.01678.x

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103. https://doi.org/10.1554/0014-3820(2001)055[1095:HCCAMA]2.0.CO;2

Regione Lombardia (2022) Scheda informativa. Anoplophora chinensis e A. glabripennis (Tarlo asiatico). https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Imprese/Imprese-agricole/servizio-fitosanitario-regionale/organismi-nocivi/anoplophora-chinensis/anoplophora-chinensis-glabripennis. Accessed 26 January 2022

Ric J, Groot P de, Gasman B, Orr M, Doyle J, Smith MT, Dumouchel L, Scarr T, Turgeon JJ (2006) Detecting signs and symptoms of Asian longhorned beetle injury: Training guide. Natural Resources Canada, Great Lakes Forestry Centre, Sault Ste. Marie, Ontario, Canadian Food Inspection Agency. 118 pp.

Ricchetti M, Fairhead C, Dujon B (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402:96–100. https://doi.org/10.1038/47076

Ricchetti M, Tekaia F, Dujon B (2004) Continued colonization of the human genome by mitochondrial DNA. PLoS Biol 2:E273. https://doi.org/10.1371/journal.pbio.0020273

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. Mol Biol Evol 21:1081–1084. https://doi.org/10.1093/molbev/msh110

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? Trends Ecol Evol 29:233–242. https://doi.org/10.1016/j.tree.2014.02.003

Rizzo D, Taddei A, Da Lio D, Bruscoli T, Cappellini G, Bartolini L, Salemi C, Luchi N, Pennacchio F, Rossi E (2020) Molecular Identification of Anoplophora glabripennis (Coleoptera: Cerambycidae) From Frass by Loop-Mediated Isothermal Amplification. J Econ Entomol 113:2911–2919. https://doi.org/10.1093/jee/toaa206

Robertson A (1961) Inbreeding in artificial selection programmes. Genet Res 2:189–194. https://doi.org/10.1017/S0016672308009452

Rodrigues TB, Dhandapani RK, Duan JJ, Palli SR (2017) RNA interference in the Asian Longhorned Beetle: Identification of Key RNAi Genes and Reference Genes for RT-qPCR. Sci Rep 7:8913. https://doi.org/10.1038/s41598-017-08813-1

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

Rogers AR (2014) How population growth affects linkage disequilibrium. Genetics 197:1329–1341. https://doi.org/10.1534/genetics.114.166454

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. Trends Ecol Evol 18:411–417. https://doi.org/10.1016/S0169-5347(03)00125-3

Roy HE, Adriaens T, Isaac NJB, Kenis M, Onkelinx T, Martin GS, Brown PMJ, Hautier L, Poland R, Roy DB, Comont R, Eschen R, Frost R, Zindel R, van Vlaenderen J, Nedvěd O, Ravn HP, Grégoire J-C, Biseau J-C de, Maes D (2012) Invasive alien predator causes rapid declines of native European ladybirds. Diversity and Distributions 18:717–725. https://doi.org/10.1111/j.1472-4642.2012.00883.x

Roy HE, Brown PMJ, Adriaens T, Berkvens N, Borges I, Clusella-Trullas S, Comont RF, Clercq P de, Eschen R, Estoup A, Evans EW, Facon B, Gardiner MM, Gil A, Grez AA, Guillemaud T, Haelewaters D, Herz A, Honek A, Howe AG, Hui C, Hutchison WD, Kenis M, Koch RL, Kulfan J, Lawson Handley L, Lombaert E, Loomans A, Losey J, Lukashuk AO, Maes D, Magro A, Murray KM, Martin GS, Martinkova Z, Minnaar IA, Nedved O, Orlova-Bienkowskaja MJ, Osawa N, Rabitsch W, Ravn HP, Rondoni G, Rorke SL, Ryndevich SK, Saethre M-G, Sloggett JJ, Soares AO, Stals R, Tinsley MC, Vandereycken A, van Wielink P, Viglášová S, Zach P, Zakharov IA, Zaviezo T, Zhao Z (2016) The harlequin ladybird, Harmonia axyridis: global perspectives on invasion history and ecology. Biol Invasions 18:997–1044. https://doi.org/10.1007/s10530-016-1077-6

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Rugman-Jones PF, Seybold SJ, Graves AD, Stouthamer R (2015) Phylogeography of the walnut twig beetle, Pityophthorus juglandis, the vector of thousand cankers disease in North American walnut trees. PLoS One 10:e0118264. https://doi.org/10.1371/journal.pone.0118264

Rusterholz H-P, Ursenbacher S, Coray A, Weibel U, Baur B (2015) DNA Quantity and Quality in Remnants of Traffic-Killed Specimens of an Endangered Longhorn Beetle: A Comparison of Different Methods. J Insect Sci 15. https://doi.org/10.1093/jisesa/iev099

Saccone C, Giorgi C de, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238:195–209. https://doi.org/10.1016/S0378-1119(99)00270-X

Sato M, Sato K (2011) Degradation of paternal mitochondria by fertilization-triggered autophagy in C. elegans embryos. Science 334:1141–1144. https://doi.org/10.1126/science.1208619

Sax DF, Brown JH (2000) The paradox of invasion. Global Ecology & Biogeography 9:363–371. https://doi.org/10.1046/j.1365-2699.2000.00217.x

Sax DF, Stachowicz JJ, Brown JH, Bruno JF, Dawson MN, Gaines SD, Grosberg RK, Hastings A, Holt RD, Mayfield MM, O'Connor MI, Rice WR (2007) Ecological and evolutionary insights from species invasions. Trends Ecol Evol 22:465–471. https://doi.org/10.1016/j.tree.2007.06.009

Scheffer SJ, Grissell EE (2003) Tracing the geographical origin of Megastigmus transvaalensis (Hymenoptera: Torymidae): an African wasp feeding on a South American plant in North America. Mol Ecol 12:415–421

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. Genetics 156:879–891. https://doi.org/10.1093/genetics/156.2.879

Schrey AW, Coon CAC, Grispo MT, Awad M, Imboma T, McCoy ED, Mushinsky HR, Richards CL, Martin LB (2012) Epigenetic Variation May Compensate for Decreased Genetic Variation with Introductions: A Case Study Using House Sparrows (Passer domesticus) on Two Continents. Genet Res Int 2012:979751. https://doi.org/10.1155/2012/979751

Scully ED, Geib SM, Carlson JE, Tien M, McKenna D, Hoover K (2014) Functional genomics and microbiome profiling of the Asian longhorned beetle (Anoplophora glabripennis) reveal insights into the digestive physiology and nutritional ecology of wood feeding beetles. BMC Genomics 15:1096. https://doi.org/10.1186/1471-2164-15-1096

Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jäger H, Kartesz J, Kenis M, Kreft H, Kühn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schierwater B, Štajerová K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. Nat Commun 8:14435. https://doi.org/10.1038/ncomms14435

Simberloff D (2009) The Role of Propagule Pressure in Biological Invasions. Annu. Rev. Ecol. Evol. Syst. 40:81–102. https://doi.org/10.1146/annurev.ecolsys.110308.120304

Simmons RB, Weller SJ (2001) Utility and evolution of cytochrome b in insects. Mol Phylogenet Evol 20:196–210. https://doi.org/10.1006/mpev.2001.0958

Smith MA, Fisher BL (2009) Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. Front Zool 6:31. https://doi.org/10.1186/1742-9994-6-31

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35. https://doi.org/10.1017/S0016672300014634

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci U S A 105:13486–13491

Sorenson MD, Quinn TW (1998) Numts: A Challenge for Avian Systematics and Population Biology. The Auk 115:214–221. https://doi.org/10.2307/4089130

Stapley J, Santure AW, Dennis SR (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24:2241–2252. https://doi.org/10.1111/mec.13089

Stewart JB, Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet 16:530–542. https://doi.org/10.1038/nrg3966

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. Ecol Evol 7:10974–10986. https://doi.org/10.1002/ece3.3625

Sved JA, Cameron EC, Gilchrist AS (2013) Estimating effective population size from linkage disequilibrium between unlinked loci: theory and application to fruit fly outbreak populations. PLoS One 8:e69078. https://doi.org/10.1371/journal.pone.0069078

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050. https://doi.org/10.1038/srep00074

Taddei A, Becker M, Berger B, Da Lio D, Feltgen S, König S, Hoppe B, Rizzo D (2021) Molecular identification of Anoplophora glabripennis (Coleoptera: Cerambycidae) and detection from frass samples based on real-time quantitative PCR. J Plant Dis Prot 128:1587–1601. https://doi.org/10.1007/s41348-021-00501-7

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol Biol Evol 9:678–687. https://doi.org/10.1093/oxfordjournals.molbev.a040752

Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol Ecol 13:321–335. https://doi.org/10.1046/j.1365-294x.2003.02070.x

Thomas VG, Hanner RH, Borisenko AV (2016) DNA-based identification of invasive alien species in relation to Canadian federal policy and law, and the basis of rapid-response management. Genome 59:1023–1031. https://doi.org/10.1139/gen-2016-0022

Torson AS, Zhang ML, Ong K, Mohammad L, Smith AJ, Doucet D, Roe AD, Sinclair BJ (2021a) Cold tolerance of laboratory-reared Asian longhorned beetles. Comp Biochem Physiol A Mol Integr Physiol 257:110957. https://doi.org/10.1016/j.cbpa.2021.110957

Torson AS, Lei Zhang M, Smith AJ, Mohammad L, Ong K, Doucet D, Roe AD, Sinclair BJ (2021b) Dormancy in laboratory-reared Asian longhorned beetles, Anoplophora glabripennis. J Insect Physiol 130:104179. https://doi.org/10.1016/j.jinsphys.2020.104179

Tourmen Y, Baris O, Dessen P, Jacques C, Malthièry Y, Reynier P (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 80:71–77. https://doi.org/10.1006/geno.2002.6798

Tsykun T, Javal M, Hölling D, Roux G, Prospero S (2019) Fine-scale invasion genetics of the quarantine pest, Anoplophora glabripennis, reconstructed in single outbreaks. Sci Rep 9:19436. https://doi.org/10.1038/s41598-019-55698-3

Turgeon JJ, 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 69:674–678

Uller T, Leimu R (2011) Founder events predict changes in genetic diversity during human-mediated range expansions. Glob. Change Biol. 17:3478–3485. https://doi.org/10.1111/j.1365-2486.2011.02509.x

van der Kuyl AC, Kuiken CL, Dekker JT, Perizonius WR, Goudsmit J (1995) Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: A problem of ancient DNA and molecular phylogenies. J Mol Evol 40:652–657. https://doi.org/10.1007/BF00160513

van Heerwaarden B, Willi Y, Kristensen TN, Hoffmann AA (2008) Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest Drosophila. Genetics 179:2135–2146. https://doi.org/10.1534/genetics.107.082768

VCFtools Manual (2018) VCFtools Manual. https://vcftools.github.io/man_latest.html. Accessed 6 December 2021

Venette R, Morey A (2019) Advances in understanding the ecology of invasive crop insect pests and their impact on IPM. In: Kogan M, Heinrichs EA (eds) Integrated management of insect pests: Current and future developments. Burleigh Dodds Science Publishing, pp 161–190

Venette RC, Hutchison WD (2021) Invasive Insect Species: Global Challenges, Strategies & Opportunities. Front. Insect Sci. 1. https://doi.org/10.3389/finsc.2021.650520

Viljakainen L, Oliveira DCSG, Werren JH, Behura SK (2010) Transfers of mitochondrial DNA to the nuclear genome in the wasp Nasonia vitripennis. Insect Molecular Biology 19 Suppl 1:27–35. https://doi.org/10.1111/j.1365-2583.2009.00932.x

Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA (2017) Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor Appl Genet 130:123–135. https://doi.org/10.1007/s00122-016-2798-8

Walter L, Lee SS (2009) Mitochondria as a Key Determinant of Aging. In: Encyclopedia of Life Sciences (ELS). John Wiley& Sons, Ltd: Chichester. https://doi.org/10.1002/9780470015902.a0020881

Wang B, Mastro V, Gao R (2005) Host range of Anoplophora glabripennis: What we've learned from common-garden experiment data. 16th U.S. Department of Agriculture interagency research forum on gypsy moth GTR-NE-337:89

Wang C, Youle RJ (2009) The role of mitochondria in apoptosis*. Annu Rev Genet 43:95–118. https://doi.org/10.1146/annurev-genet-102108-134850

Wang D, Xiang H, Ning C, Liu H, Liu J-F, Zhao X (2020a) Mitochondrial DNA enrichment reduced NUMT contamination in porcine NGS analyses. Brief Bioinform 21:1368–1377. https://doi.org/10.1093/bib/bbz060

Wang J-X, Liu J, Miao Y-H, Huang D-W, Xiao J-H (2020b) Tracking the Distribution and Burst of Nuclear Mitochondrial DNA Sequences (NUMTs) in Fig Wasp Genomes. Insects 11. https://doi.org/10.3390/insects11100680

Wang X, Keena MA (2021) Hybridization Potential of Two Invasive Asian Longhorn Beetles. Insects 12. https://doi.org/10.3390/insects12121139

Wang X, Aparicio EM, Murphy TC, Duan JJ, Elkinton JS, Gould JR (2019) Assessing the host range of the North American parasitoid Ontsira mellipes: Potential for biological control of Asian longhorned beetle. Biological Control 137:104028. https://doi.org/10.1016/j.biocontrol.2019.104028

Wang X, Aparicio EM, Duan JJ, Gould J, Hoelmer KA (2020c) Optimizing Parasitoid and Host Densities for Efficient Rearing of Ontsira mellipes (Hymenoptera: Braconidae) on Asian Longhorned Beetle (Coleoptera: Cerambycidae). Environ Entomol 49:1041–1048. https://doi.org/10.1093/ee/nvaa086

Wang XM, Zhang CX, Hasi E, Dong ZB (2010) Has the Three Norths Forest Shelterbelt Program solved the desertification and dust storm problems in arid and semiarid China? Journal of Arid Environments 74:13–22. https://doi.org/10.1016/j.jaridenv.2009.08.001

Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the north atlantic intertidal. Evol 55:2455. https://doi.org/10.1554/0014-3820(2001)055[2455:PAHEOT]2.0.CO;2

Wermelinger B, Forser B, Hölling D, Plüss T, Reamy O, Klay A (2015) Invasive Laubholz-Bockkäfer aus Asien. Merkblatt für die Praxis (WSL Birmensdorf) 50

White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. Mol Ecol 17:4925–4942. https://doi.org/10.1111/j.1365-294X.2008.03982.x

Williams DW, Lee H-P, Kim I-K (2004) Distribution and Abundance of Anoplophora glabripennis (Coleoptera: Cerambycidae) in Natural Acer Stands in South Korea. Environ Entomol 33:540–545. https://doi.org/10.1603/0046-225X-33.3.540

Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. Trends Ecol Evol 24:136–144. https://doi.org/10.1016/j.tree.2008.10.007

Wray GA (2013) Genomics and the Evolution of Phenotypic Traits. Annu. Rev. Ecol. Evol. Syst. 44:51–72. https://doi.org/10.1146/annurev-ecolsys-110512-135828

Wu Y, Trepanowski NF, Molongoski JJ, Reagel PF, Lingafelter SW, Nadel H, Myers SW, Ray AM (2017) Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in trade-associated solid wood packaging material using DNA barcoding and morphology. Sci Rep 7:40316. https://doi.org/10.1038/srep40316

Xu J, Fonseca DM, Hamilton GC, Hoelmer KA, Nielsen AL (2014) Tracing the origin of US brown marmorated stink bugs, Halyomorpha halys. Biol Invasions 16:153–166. https://doi.org/10.1007/s10530-013-0510-3

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. https://doi.org/10.1186/1471-2105-13-134

Zardoya R, Meyer A (1996) Phylogenetic Performance of Mitochondrial Protein-Coding Genes in Resolving Relationships Among Vertebrates. Mol Biol Evol 13:933–942

Zhang D-X, Hewitt GM (1996a) Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol Evol 11:247–251. https://doi.org/10.1016/0169-5347(96)10031-8

Zhang A, Oliver JE, Aldrich JR, Wang B, Mastro VC (2002) Stimulatory Beetle Volatiles for the Asian Longhorned Beetle, Anoplophora glabripennis (Motschulsky). Zeitschrift für Naturforschung C 57:553–558. https://doi.org/10.1515/znc-2002-5-626

Zhang A, Oliver JE, Chauhan K, Zhao B, Xia L, Xu Z (2003) Evidence for contact sex recognition pheromone of the Asian longhorned beetle, Anoplophora glabripennis (Coleoptera: Cerambycidae). Naturwissenschaften 90:410–413. https://doi.org/10.1007/s00114-003-0452-1

Zhang DX, Hewitt GM (1996b) Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. Insect Molecular Biology 6:143–150

“Let the past hold on to itself and let the present move forward into the future.”

— Douglas Adams, Mostly Harmless

Anoplophora glabripennis, picture provided by Olaf Zimmermann (LTZ, Germany)

Acknowledgments

ADVISOR AND EXAMINER

First of all, I would like to thank my supervisor Prof. Dr. Martin Hasselmann for his great support, mentoring and trust in my abilities. Especially, I want to thank him for being way more optimistic than I was during the whole time and his many encouragements when I was doubtful. Of course, I also acknowledge gratefully that I was able to continue with this project in his department.

FINANCIAL SUPPORT

I acknowledge the support by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme and Carmen Lübken for supervising the project PHID Coleo.

THE BEETLE PEOPLE: PHID COLEO TEAM

I am thankful for the PHID Coleo Team of Martin Hasselmann, Olaf Zimmermann, Philipp Bauer and Claus Zebitz. Thank you all for being an indispensable part of this project, for productive discussions and brainstromings during our meetings. Special thank goes to Olaf Zimmermann for his amazing advice regarding many concerns and his connections to experts he allowed me to make use of, without whom this project would not have been possible. Of course, I would like to also thank him for many lunches he invited me at meetings and conferences. Many thanks to Philipp Bauer, I am glad about our ‘remote fellowship’, the many calls we had for sharing our ups, downs and beyond.

THE POPGENOM PEOPLE: GROUP MEMBERS

I am deeply thankful for all my current and former group members for many advices, good memories, coffee breaks, ‘notworking’ rounds in the park, christmas markets and so on. I specially want to thank Melanie for her technical support in the lab beside her open ears and advice. Furthermore, thanks and hugs to Melanie, Clara, and Denyse for all the amazing travels and trips we did together, work and not work related. Also, many thanks to Birgit Tewelde, Karin, Lina and Paul for encouragements, great conversations, and advice.

THE ALB PEOPLE: SAMPLE COLLECTION

I appreciate (alphabetical order) Amanda Roe, Andrea Taddei, Beatrice Berger, Beniamino Cavagna, Birgit Gessler, Björn Hoppe, Doris Hölling, Frank Nüsser, Hannes Lemme, Heiko Disch, Hermann Meier, Matthias von Wuthenau, Marion Javal, Matteo Zugno, Matthias Becker, Mingming Cui, Olaf Zimmermann, Philipp Bauer, Simone Prospero, Stephan König, Thomas Schröder and Ullrich Benker for providing information, support and ALB samples for this research project, and/or for giving the permission to use figures of ALB, damages or phytosanitary measures (**Figure 4**-**Figure 7**) and **Figure 3**.

THE COMPUTATIONAL PEOPLE

I appreciate the amazing team of Physalia courses and the organizer Carlo Pecoraro, who taught me everything I needed to know for my analyses. Special thanks to José Cerca, Lewis Spurgin, Martin Taylor, Christoph Hahn and Daniel Pass for their great instructions. I am sure, they saved me from lots of hours banging the head against the keyboard. I would also like to thank Philipp Schlüter who gave me great advice on the analyses, when I needed a new perspective on what I've already done and what still needed to be done. Everything starts somewhere. Thus, additionally many thanks go to Lennart, Julian, Clara and Che Wei who helped me with all my small and bigger coding issues in the very beginning, when I was still new to the command line. Thanks for all the motivation of the co-writers on 'The writing academic', we went through all the ups and downs during months of writing together.

THE BACKBONE: FAMILY AND FRIENDS

Finally, this work would have been impossible without the tremendous support and continuous encouragements from Marcel, the support and believe in me by my parents as well as my sisters Eva and Michaela, and the open ears of my friends Sandra, Tina, Patrick and more.

Curriculum Vitae

PERSONAL DATA

—Address: Schubartstraße 27, 71638 Ludwigsburg, Germany
—Date of birth: 24.02.1988 in Schorndorf, Germany
—Contact: iris.haeuss@gmail.com

LANGUAGES

German (native language), English (full professional proficiency), French (elementary)

OCCUPATION

Scientific staff

—12/2021-present University of Hohenheim, Population Genomics in Livestock
—07/2017-05/2021 University of Hohenheim, Population Genomics in Livestock
—08/2016-12/2016 University of Tübingen, Centre for molecular biology of plants

Lab assistant/Student Research Assistant

—05/2011-07/2016 University of Hohenheim, several positions as assistant in S1 labs and field research in the departments Phytopathology, Nutritional Crop Physiology, as well as Fertilization and Soil Chemistry

EDUCATION

Graduate studies to obtain the Dr. Sc. Agr. in agricultural sciences (EN)

—09/2017-present University of Hohenheim, Population Genomics in Livestock
- "European population genomic differentiation and dispersal pattern of the invasive beetle *Anoplophora glabripennis*"; *in process*

Master of science in agricultural biology (DE)

—10/2012-09/2015 University of Hohenheim, Major: Agricultural Biotechnology; very good (1.5)
- "Validierung der Interaktion zwischen dem Effektorkandidaten EC23 von *Phakopsora pachyrhizi* und seinen vermeintlichen Zielproteinen bei *Glycine max*"; very good (1.0)

Bachelor of science in agricultural biology (DE)

—10/2009-12/2012 University of Hohenheim; good (2.0)
- "Identifizierung von niedrig affinen Harnstoff-Transportern in *Arabidopsis thaliana*"; very good (1.0)

Vocational education as lab assistant in biotechnology (DE)

—09/2007-07/2009 Vocational collage for Biotechnology, Hohenheim; very good (1.3)

General qualification for university entrance (DE)

—09/1998-06/2007 Burg-Gymnasium Schorndorf

ADVANCED TRAINING

Summer school in bioinformatics

—07/2021 Physalia Courses (online), Free University of Berlin, 30 hours (EN)

Introduction to population genomics

—05/2019 Physalia Courses, Free University of Berlin, 40 hours (EN)

Introductory and intermediate course: high performance computing (HPC)

—04/2019 Karlsruhe Institute for technology (KIT), 12 hours (EN)

Usage of the statistics software R

—03/2018 R for Animal breeders, University of Hohenheim, 40 hours (DE)

CONFERENCE ATTENDANCE / CONTRIBUTIONS

—09/2021 Oral presentation (DE), German Plant protection conference (DPST), Online, Germany

—10/2020 Digital Innovation days of the German ministry of Agriculture and Nutrition (BLE)

—08/2019 Poster (EN), Congress of the ESEB, Turku, Finland

—03/2019 Oral presentation (EN), Entomology Congress of DGaaE, Halle an der Saale, Germany

—09/2018 Oral presentation (DE), German Plant protection conference (DPST), Hohenheim, Germany

—01/2018 Oral presentation (DE), Expert meeting Asian long-horned beetle, Hildrizhausen, Germany

TEACHING

Workshop in molecular identification of longicorn beetles (DE)

—10/2019 (2.5 h) Meeting of the entomological diagnostics of German states, LTZ Augustenberg, Karlsruhe, Germany

—09/2018 (2.5 h) German Plant protection conference (DPST), Hohenheim, Germany

Supervision/teaching of students (DE, EN)

—10/2017-06/2018 University of Hohenheim, Population Genomics in Livestock, supervision of a Master student (EN)

—10/2010-02/2013 University of Hohenheim, Institute for Chemistry, Tutoress for Inorganic experimental chemistry (Bachelor level) (DE)

MEMBERSHIPS

Member of the European society of evolutionary biology (ESEB)

Ludwigsburg, 28.03.2022

Place, Date

I. Häußermann

Signature

Affidavit

Declaration in lieu of an oath on independent work

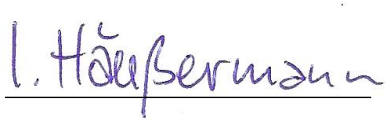
according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences

1. The dissertation submitted on the topic
   "European population genomic differentiation and dispersal pattern of the invasive beetle *Anoplophora glabripennis*"
   is work done independently by me.
2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.
3. I did not use the assistance of a commercial doctoral placement or advising agency.
4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Ludwigsburg, 28.03.2022

Place, Date



Signature

Appendix

I. Asian long-horned beetle sample, PCR product and DNA-marker information

Table 20. ALB sampling information about collection year, developmental stage, storing condition, source and hosts per Pop-ID. – (1/2)

Pop-ID	Year(s)	n	Stage(s)	Storing condition	Source(s)	Host (s)
CH-AG-BK	2015	2	larva	-20°C	WSL (CH)	*Acer sp.*
CH-FR-MA	2014	12	imago (209, 210), larva	-20°C	WSL (CH)	*Aesculus hippocastanu, Acer sp., Acer negundo, Acer campestre, Acer pseudoplatanus*
CH-FRS-BR	2013	5	larva	-20°C	WSL (CH)	*Acer pseudoplatanus*
CH-ZH-WT	2012	2	larva	-20°C	WSL (CH)	*Populus nigra, Betula sp.*
CN-HA	2016	1	N/A	-20°C	Laval University (CAN)	N/A
CN-JI	2016	1	N/A	-20°C	Laval University (CAN)	N/A
CN-TO	2016	1	N/A	-20°C	Laval University (CAN)	N/A
D-BW-HLD	2016	15	imago	RT, 80% ethanol	LTZ (D)	*Acer sp.*
D-BW-KA	2016	1	imago	RT, dried	LTZ (D)	N/A (import control)
D-BW-WAR	2011, 2015	9	imago, pupa (191)	RT, dried	RP Freiburg (D)	N/A
D-BY-FK/ D-BY-FKR	2013, 2014, 2016	27	larva	-20°C, 80% ethanol (027, 030)/-20°C	LWF, LTZ (D) (027, 030) / LWF, LfL (D)	*Aesculus hippocastanum, Acer sp.*
D-BY-KEH	2016	10	larva	-20°C	LWF, LfL (D)	*Acer sp.*
D-BY-MI	2019	7	larva (193), imago	-20°C, 80% ethanol (193)/RT, dried	LfL (D)	*Acer sp.*
D-BY-MU	2017	10	larva	-20°C	LWF, LfL (D)	*Acer sp., Salix sp.*
D-BY-NBB	2013, 2014	15	larva	-20°C, 80% ethanol (026, 031-034)/-20°C	LWF, LTZ (D) (026, 031-034)/LWF, LfL (D)	N/A
D-BY-NKI	2012	2	imago	RT, dried	LfL, LTZ (D)	N/A

Table 20. ALB sampling information about collection year, developmental stage, storing condition, source and hosts per Pop-ID. – (2/2)

Pop-ID	Year(s)	n	Stage(s)	Storing condition	Source(s)	Host (s)
D-BY-SB	2015, 2018	11	larva, imago (110)	-20°C/RT, dried (110)	LWF, LfL (D)	*Acer sp., Betula sp.*, N / A (pheromone trap, 110)
D-ST-MB	2015, 2018, 2019	35	larva, imago (200, 201)	-20°C, 80% ethanol/RT, 80% ethanol (200, 201)	Llg, JKI, LTZ (D) (035-039)/Llg, JKI (D)/WSL (CH) (200, 201)	*Salix sp., Populus sp., Acer sp., Fraxinus sp.*
I-ANM-O	2017	2	larva, imago	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-BGL-TB	2017, 2018	8	larva, imago	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-CNP-CN	2018	3	larva	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-FMM-FM	2015	2	larva, imago	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-MIL-CB	2016, 2017, 2019	8	larva	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-MIL-VI	2014, 2016	7	larva	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A
I-TOP-V	2018	3	larva	-20°C, 80% ethanol	Lombardia Plant Health office (I)	N/A

n is the number of specimens for the respective Pop-ID. Underlined numbers in brackets are the respective sample-no for the given information, which is also represented in sample-IDs. The mentioned storing conditions refer to the time from collection until the arrival at UHO and was provided from the sample sources. If there are more than one sample source, the specimens were passed over all the mentioned institutions. All samples from Switzerland and China, as well as D-ST-MB-19-164 and D-ST-MB-19-165 were obtained as DNA extracts. All other samples were received as tissue or whole specimens.

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (1/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
CH-AG-BK-15-221	1	0	1	0	0	0	0	0	0	0
CH-AG-BK-15-222	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-209	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-210	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-211	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-212	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-213	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-214	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-215	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-216	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-217	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-218	1	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-219	**0**	0	1	0	0	0	0	0	0	0
CH-FR-MA-14-220	1	0	1	0	0	0	0	0	0	0
CH-FRS-BR-13-204	1	0	1	0	0	0	0	0	0	0
CH-FRS-BR-13-205	1	0	1	0	0	0	0	0	0	0
CH-FRS-BR-13-206	1	0	1	0	0	0	0	0	0	0
CH-FRS-BR-13-207	1	0	1	0	0	0	0	0	0	0
CH-FRS-BR-13-208	1	0	1	0	0	0	0	0	0	0
CH-ZH-WT-12-202	1	0	1	0	0	0	0	0	0	0
CH-ZH-WT-12-203	1	0	1	0	0	0	0	0	0	0
CN-HA15	1	0	1	0	0	0	0	0	0	0
CN-JI11	1	0	1	0	0	0	0	0	0	0
CN-TO7	1	0	1	0	0	0	0	0	0	0
D-BW-HLD-16-003	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-004	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-005	1	1	1	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (2/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-BW-HLD-16-006	0	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-007	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-008	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-009	0	1	0	0	0	0	0	0	0	0
D-BW-HLD-16-011	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-012	0	1	0	0	0	0	0	0	0	0
D-BW-HLD-16-013	1	1	1	1	1	1	1	1	1	1
D-BW-HLD-16-014	0	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-015	1	1	1	0	0	0	0	0	0	0
D-BW-HLD-16-016	0	1	0	0	0	0	0	0	0	0
D-BW-HLD-16-017	0	1	0	0	0	0	0	0	0	0
D-BW-HLD-16-018	0	1	0	0	0	0	0	0	0	0
D-BW-KA-16-020	0	1	0	0	0	0	0	0	0	0
D-BW-WAR-11-187	0	0	1	0	0	0	0	0	0	0
D-BW-WAR-11-188	0	0	1	0	0	0	0	0	0	0
D-BW-WAR-11-189	0	0	1	1	1	0	1	0	1	0
D-BW-WAR-11-190	0	0	1	0	0	0	0	0	0	0
D-BW-WAR-11-191	0	0	0	0	0	0	0	0	0	0
D-BW-WAR-11-192	0	0	0	0	0	0	0	0	0	0
D-BW-WAR-15-023	1	1	1	0	0	0	0	0	0	0
D-BW-WAR-15-024	1	1	1	1	1	0	1	1	1	1
D-BW-WAR-15-025	1	1	1	0	0	0	0	0	0	0
D-BY-FK-13-027	0	1	0	0	0	0	0	0	0	0
D-BY-FK-13-030	0	1	0	0	0	0	0	0	0	0
D-BY-FK-13-065	1	1	1	0	0	0	0	0	0	0
D-BY-FK-13-066	1	1	1	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (3/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-BY-FK-13-067	1	1	1	0	0	0	0	0	0	0
D-BY-FK-13-069	1	0	1	0	0	0	0	0	0	0
D-BY-FK-13-070	1	0	1	0	0	0	0	0	0	0
D-BY-FK-13-071	1	0	1	0	0	0	0	0	0	0
D-BY-FK-13-072	1	0	1	0	0	0	0	0	0	0
D-BY-FK-13-073	1	0	1	0	0	0	0	0	0	0
D-BY-FK-13-074	1	0	1	**0**	1	1	1	1	1	1
D-BY-FK-14-028	0	1	0	0	0	0	0	0	0	0
D-BY-FK-14-029	0	1	0	0	0	0	0	0	0	0
D-BY-FK-14-055	**0**	1	1	0	0	0	0	0	0	0
D-BY-FK-14-056	1	1	1	0	0	0	0	0	0	0
D-BY-FK-14-057	**0**	1	1	0	0	0	0	0	0	0
D-BY-FK-14-058	1	1	1	1	1	1	1	1	1	1
D-BY-FK-14-059	1	1	1	0	0	0	0	0	0	0
D-BY-FK-14-060	0	1	0	0	0	0	0	0	0	0
D-BY-FK-14-061	1	1	1	0	0	0	0	0	0	0
D-BY-FK-14-062	1	1	1	0	0	0	0	0	0	0
D-BY-FK-14-063	1	1	1	0	0	0	0	0	0	0
D-BY-FK-14-064	1	1	1	0	0	0	0	0	0	0
D-BY-FKR-16-075	1	0	1	0	0	0	0	0	0	0
D-BY-FKR-16-076	1	0	1	1	1	1	1	1	1	1
D-BY-FKR-16-077	1	0	1	0	0	0	0	0	0	0
D-BY-FKR-16-078	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-100	**0**	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-101	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-102	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-103	1	0	1	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (4/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-BY-KEH-16-104	1	0	1	1	**0**	1	1	1	1	1
D-BY-KEH-16-105	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-106	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-107	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-108	1	0	1	0	0	0	0	0	0	0
D-BY-KEH-16-109	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-193	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-194	1	0	1	**0**	1	**0**	1	1	1	1
D-BY-MI-19-195	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-196	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-197	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-198	1	0	1	0	0	0	0	0	0	0
D-BY-MI-19-199	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-090	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-091	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-092	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-093	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-094	**0**	0	**0**	0	0	0	0	0	0	0
D-BY-MU-17-095	1	0	1	1	1	1	1	1	1	1
D-BY-MU-17-096	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-097	1	0	1	0	0	0	0	0	0	0
D-BY-MU-17-098	**0**	0	1	0	0	0	0	0	0	0
D-BY-MU-17-099	**0**	0	1	0	0	0	0	0	0	0
D-BY-NBB-13-026	0	1	0	0	0	0	0	0	0	0
D-BY-NBB-14-031	0	1	0	0	0	0	0	0	0	0
D-BY-NBB-14-032	0	1	0	0	0	0	0	0	0	0
D-BY-NBB-14-033	0	1	0	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (5/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-BY-NBB-14-034	0	1	0	0	0	0	0	0	0	0
D-BY-NBB-14-079	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-080	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-081	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-082	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-083	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-084	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-085	1	0	1	**0**	1	1	1	1	1	1
D-BY-NBB-14-086	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-087	1	0	1	0	0	0	0	0	0	0
D-BY-NBB-14-088	1	0	1	0	0	0	0	0	0	0
D-BY-NKI-12-021	**0**	0	1	0	0	0	0	0	0	0
D-BY-NKI-12-022	**0**	0	1	0	0	0	0	0	0	0
D-BY-SB-15-045	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-046	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-047	1	1	**1**	1	1	1	1	1	1	1
D-BY-SB-15-048	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-049	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-050	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-051	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-052	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-053	1	1	1	0	0	0	0	0	0	0
D-BY-SB-15-054	1	0	1	0	0	0	0	0	0	0
D-BY-SB-18-110	1	0	1	0	0	0	0	0	0	0
D-ST-MB-15-035	1	1	1	0	0	0	0	0	0	0
D-ST-MB-15-036	**0**	1	**0**	0	0	0	0	0	0	0
D-ST-MB-15-037	1	1	1	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (6/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-ST-MB-15-038	1	1	1	**0**	**0**	1	1	1	1	1
D-ST-MB-15-039	**0**	1	1	0	0	0	0	0	0	0
D-ST-MB-18-128	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-129	1	0	1	**0**	1	**0**	1	1	1	1
D-ST-MB-18-130	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-132	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-133	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-134	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-135	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-136	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-137	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-138	1	0	1	0	0	0	0	0	0	0
D-ST-MB-18-139	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-126	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-127	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-140	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-142	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-143	**0**	0	1	0	0	0	0	0	0	0
D-ST-MB-19-144	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-145	**0**	0	1	0	0	0	0	0	0	0
D-ST-MB-19-146	**0**	0	1	0	0	0	0	0	0	0
D-ST-MB-19-147	**0**	0	1	0	0	0	0	0	0	0
D-ST-MB-19-148	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-149	**0**	0	1	1	**0**	1	1	1	1	**0**
D-ST-MB-19-150	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-151	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-152	1	0	1	0	0	0	0	0	0	0

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (7/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
D-ST-MB-19-154	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-164	1	0	1	1	1	1	1	1	1	1
D-ST-MB-19-165	1	0	1	0	0	0	0	0	0	0
D-ST-MB-19-200	**0**	0	**0**	0	0	0	0	0	0	0
D-ST-MB-19-201	**0**	0	**0**	0	0	0	0	0	0	0
I-ANM-O-17-121	1	0	1	0	0	0	0	0	0	0
I-ANM-O-17-170	1	0	1	1	1	1	1	1	1	1
I-BGL-TB-17-111	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-17-172	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-17-174	1	0	1	1	1	1	1	1	1	1
I-BGL-TB-17-175	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-17-177	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-1718-112	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-1718-113	1	0	1	0	0	0	0	0	0	0
I-BGL-TB-18-186	1	0	1	0	0	0	0	0	0	0
I-CNP-CN-18-124	1	0	1	**0**	1	1	1	1	1	1
I-CNP-CN-18-125	1	0	1	0	0	0	0	0	0	0
I-CNP-CN-18-168	1	0	1	0	0	0	0	0	0	0
I-FMM-FM-15-120	1	0	1	**0**	1	1	1	1	1	1
I-FMM-FM-15-169	1	0	1	0	0	0	0	0	0	0
I-MIL-CB-16-119	1	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-117	1	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-118	**0**	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-173	1	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-178	1	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-179	**0**	0	1	0	0	0	0	0	0	0
I-MIL-CB-17-180	**0**	0	1	1	**0**	**0**	1	1	**0**	1

Table 21. Sample-IDs and the corresponding sequenced PCR products and DNA markers per specimen. – (8/8)

Sample-ID	mtCOI	ALB-COI-II	GBS-SNPs	ALBsc158-Numt1(f)	ALBsc158-Numt1(int)	ALBsc158-Numt1(b)	ALBsc1709-Numt2	ALBsc1709-Numt3(f)	ALBsc1709-Numt3(int)	ALBsc1709-Numt3(b)
I-MIL-CB-19-171	**0**	0	1	0	0	0	0	0	0	0
I-MIL-VI-14-115	1	0	1	0	0	0	0	0	0	0
I-MIL-VI-14-181	**0**	0	1	**1**	**0**	**0**	1	1	1	**0**
I-MIL-VI-14-182	1	0	1	0	0	0	0	0	0	0
I-MIL-VI-14-183	**0**	0	1	0	0	0	0	0	0	0
I-MIL-VI-14-184	1	0	1	0	0	0	0	0	0	0
I-MIL-VI-14-185	**0**	0	**0**	0	0	0	0	0	0	0
I-MIL-VI-16-116	1	0	1	0	0	0	0	0	0	0
I-TOP-V-18-122	1	0	1	1	1	1	1	1	1	1
I-TOP-V-18-123	1	0	1	0	0	0	0	0	0	0
I-TOP-V-18-167	1	0	1	0	0	0	0	0	0	0
Specimen count	150	55	176	15	17	16	22	21	21	19

The binary code (0, 1) was used here to show for each specimen (sample-ID) which one of the PCR products and DNA-markers were sequenced (1) and which one not (0). When a "**0**" is red, this means the corresponding PCR product/DNA marker was sequenced, but the results were not sufficient, either due to no or poor PCR product or poor-quality sequences. The sample-IDs marked in **red** show which of the specimen could not result in any sufficient sequence information. The red "**1**" of I-MIL-VI-14-181 represents a single clone sequence, which was excluded from the dataset. The sample-IDs marked in **blue** represent the specimens that were used for NUMT PCR product sequencing as representatives for their Pop-ID. All PCR products/DNA markers that were cloned before sequencing were marked with black frames. The green marked "**1**" for the GBS-SNP marker of D-BY-SB-15-047 symbolizes the role as internal control of this specimen in the three GBS runs.

II. Chapter A

Table 22. megaBLAST of the haplotypes COI_Type1-Type8 and the reference sequence. – (1/2)

limited to	Nucleotide Collection				representative genomes of *Anoplophora glabripennis*			
Query	Subject	Query coverage	E value	Percent identity	Subject	Query coverage	E value	Percent identity
COI_Type1	COI partial cds (KY357648.1)	100%	0.0	100%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.85%
	COI partial cds (MW243612.1)	100%	0.0	100%	scaffold158 (NW_019416265.1)	99%	0.0	95.28%
	mitochondrion, complete genome (DQ768215.1)	100%	0.0	99.85%	scaffold1709 (NW_019417922.1)	99%	0.0	92.41%
COI_Type2	COI partial cds (KY357648.1)	100%	0.0	99.85%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.70%
	COI partial cds (MW243612.1)	100%	0.0	99.85%	scaffold158 (NW_019416265.1)	99%	0.0	95.13%
	mitochondrion, complete genome (DQ768215.1)	100%	0.0	99.70%	scaffold1709 (NW_019417922.1)	99%	0.0	92.26%
COI_Type3	COI partial cds (KY357651.1)	100%	0.0	100%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.39%
	COI partial cds (KY357652.1)	100%	0.0	99.85%	scaffold158 (NW_019416265.1)	99%	0.0	94.82%
	COI complete cds (KF737825.1)	100%	0.0	99.85%	scaffold1709 (NW_019417922.1)	99%	0.0	92.26%
COI_Type4	COI partial cds (KY357648.1)	100%	0.0	99.85%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.70%
	COI partial cds (MW243612.1)	100%	0.0	99.85%	scaffold158 (NW_019416265.1)	99%	0.0	95.13%
	mitochondrion, complete genome (DQ768215.1)	100%	0.0	99.70%	scaffold1709 (NW_019417922.1)	99%	0.0	92.26%

Table 22. megaBLAST of the haplotypes COI_Type1-Type8 and the reference sequence. – (2/2)

Query	Subject	E value	Percent identity	Percent identity	Subject	Query coverage	E value	Percent identity
COI_Type5	COI partial cds (KY357652.1)	100%	0.0	99.85%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.70%
	COI partial cds (KY357648.1)	100%	0.0	99.85%	scaffold158 (NW_019416265.1)	99%	0.0	95.13%
	COI complete cds (KF737825.1)	100%	0.0	99.85%	scaffold1709 (NW_019417922.1)	99%	0.0	92.26%
COI_Type6	COI complete cds (GU003927.1)	100%	0.0	100%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	98.63%
	COI partial cds (AB439153.1)	99%	0.0	100%	scaffold158 (NW_019416265.1)	99%	0.0	95.59%
	COI partial cds (KY357647.1)	100%	0.0	99.54%	scaffold1709 (NW_019417922.1)	99%	0.0	93.17%
COI_Type7	COI partial cds (KY357651.1)	100%	0.0	99.85%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.24%
	COI partial cds (KY357652.1)	100%	0.0	99.70%	scaffold158 (NW_019416265.1)	99%	0.0	94.67%
	COI complete cds (KF737825.1)	100%	0.0	99.70%	scaffold1709 (NW_019417922.1)	99%	0.0	92.11%
COI_Type8	COI partial cds (KY357651.1)	100%	0.0	99.54%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	98.94%
	COI partial cds (KY357652.1)	100%	0.0	99.39%	scaffold158 (NW_019416265.1)	99%	0.0	94.67%
	COI complete cds (KF737825.1)	100%	0.0	99.39%	scaffold1709 (NW_019417922.1)	99%	0.0	92.11%
COI_Ref	mitochondrion, complete genome (NC_008221.1)	100%	0.0	100.00%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	100.00%
	COI partial cds (KY357648.1)	100%	0.0	99.85%	scaffold158 (NW_019416265.1)	99%	0.0	95.13%
	COI partial cds (MW243612.1)	100%	0.0	99.85%	scaffold1709 (NW_019417922.1)	99%	0.0	92.26%

Table 23. F_{ST}-matrix – Population differentiation on nine haplotypes (mtCOI-658 bp-a) within mtCOI-658bp conducted in Arlequin v. 3.5.

	Pop-ID	1	2	3	4	5	6	7	8	9	10	11	12	13
1	CH_FR_MA	0.000	–	+	+	+	+	+	+	+	+	+	+	+
2	CH_FRS_BR	0.000	0.000	+	+	+	+	+	+	+	+	+	+	+
3	D_BW_HLD	1.000	1.000	0.000	+	–	–	–	–	+	+	–	+	–
4	D_BY_FK	0.936	0.920	0.852	0.000	+	+	+	+	+	+	+	+	+
5	D_BY_KEH	0.473	0.338	0.230	0.655	0.000	–	–	–	+	+	–	+	–
6	D_BY_MI	1.000	1.000	0.000	0.847	0.208	0.000	–	–	+	+	–	+	–
7	D_BY_MU	1.000	1.000	0.000	0.847	0.208	0.000	0.000	–	+	+	–	+	–
8	D_BY_NBB	1.000	1.000	0.000	0.862	0.268	0.000	0.000	0.000	+	+	–	+	–
9	D_BY_SB	1.000	1.000	1.000	0.968	0.655	1.000	1.000	1.000	0.000	+	+	+	+
10	D_ST_MB	1.000	1.000	1.000	0.957	0.782	1.000	1.000	1.000	1.000	0.000	+	+	+
11	I_BGL_TB	0.544	0.404	0.143	0.663	–0.11504	0.120	0.120	0.181	0.755	0.805	0.000	+	–
12	I_MIL	1.000	1.000	1.000	0.985	0.911	1.000	1.000	1.000	1.000	1.000	0.927	0.000	+
13	out	0.281	0.196	0.029	0.421	–0.05321	0.016	0.016	0.049	0.513	0.524	–0.07507	0.804	0.000

In the lower matrix, the exact F_{ST}-values are shown, while the upper matrix shows if they were significant (+) or not significant (– and red); 110 permutations; significance level = 0.0500. Population labels were used as set in **Table 4**.

.

Table 24. p-values to F_{ST}-values – Population differentiation on nine haplotypes (mtCOI-658 bp-a) within mtCOI-658bp conducted in Arlequin v. 3.5.

Pop-ID	1	2	3	4	5	6	7	8	9	10	11	12	13
CH_FR_MA		–	+	+	+	+	+	+	+	+	+	+	+
CH_FRS_BR	-1.00000+-1.0000		+	+	+	+	+	+	+	+	+	+	+
D_BW_HLD	0.00008+-0.0000	0.00078+-0.0002	+	–	–	–	–	+	+	–	+	–	–
D_BY_FK	0.00000+-0.0000	0.00003+-0.0000	0.00000+-0.0000		+	+	+	+	+	+	+	+	+
D_BY_KEH	0.00000+-0.0000	0.00047+-0.0001	0.20794+-0.0017	0.00000+-0.0000		–	–	–	+	+	–	+	–
D_BY_MI	0.00010+-0.0000	0.00105+-0.0002	-1.00000+-1.0000	0.00003+-0.0000	0.21305+-0.0019		–	–	+	+	–	+	–
D_BY_MU	0.00001+-0.0000	0.00124+-0.0002	-1.00000+-1.0000	0.00003+-0.0000	0.21111+-0.0025	-1.00000+-1.0000		–	+	+	–	+	–
D_BY_NBB	0.00000+-0.0000	0.00025+-0.0001	-1.00000+-1.0000	0.00001+-0.0000	0.08515+-0.0021	-1.00000+-1.0000	-1.00000+-1.0000		+	+	–	+	–
D_BY_SB	0.00000+-0.0000	0.00020+-0.0001	0.00008+-0.0000	0.00000+-0.0000	0.00222+-0.0003	0.00005+-0.0000	0.00000+-0.0000	0.00000+-0.0000		+	+	+	+
D_ST_MB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000		+	+	+
I_BGL_TB	0.00005+-0.0001	0.00087+-0.0003	0.46452+-0.0007	0.00000+-0.0000	1.00000+-0.0000	0.46917+-0.0018	0.46541+-0.0008	0.17943+-0.0015	0.00078+-0.0002	0.00000+-0.0000		+	–
I_MIL	0.00000+-0.0000	0.00099+-0.0003	0.00012+-0.0001	0.00000+-0.0000	0.00001+-0.0000	0.00018+-0.0000	0.00015+-0.0001	0.00001+-0.0000	0.00001+-0.0000	0.00000+-0.0000	0.00043+-0.0003		+
out	0.00000+-0.0000	0.00021+-0.0002	0.87874+-0.0031	0.00000+-0.0000	0.91768+-0.0028	0.86077+-0.0037	0.86072+-0.0033	0.65529+-0.0044	0.00000+-0.0000	0.00000+-0.0000	1.00000+-0.0000	0.00000+-0.0000	

110 permutations, 1-13 represent the Pop-IDs on the left in the same order. Population labels were used as set in **Table 4**. Values above the significance level of 0.05 are marked in red.

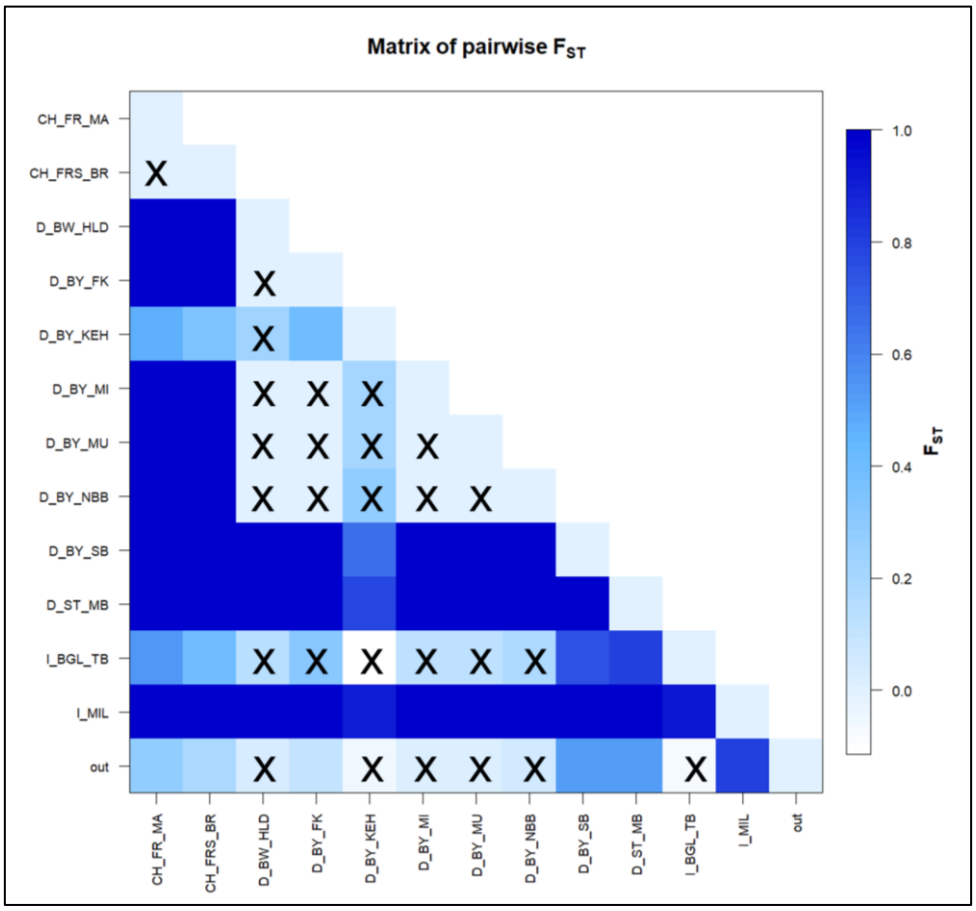


Figure 91. Population differentiation on eight haplotypes (mtCOI-658 bp-b) within mtCOI-658bp conducted in Arlequin v. 3.5.
Pairwise F_{ST}-values among populations mtCOI-658 bp-b; x = not significant (significance level=0.05). Population labels were used as set in **Table 4**.

Table 25. F_{ST}-matrix – Population differentiation on eight haplotypes (mtCOI-658 bp-b) within mtCOI-658bp conducted in Arlequin v. 3.5.

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	CH_FR_MA	0.000	–	+	+	+	+	+	+	+	+	+	+	+
2	CH_FRS_BR	0.000	0.000	+	+	+	+	+	+	+	+	+	+	+
3	D_BW_HLD	1.000	1.000	0.000	–	–	–	–	–	+	+	–	+	–
4	D_BY_FK	1.000	1.000	0.000	0.000	+	–	–	–	+	+	–	+	+
5	D_BY_KEH	0.473	0.338	0.230	0.401	0.000	–	–	–	+	+	–	+	–
6	D_BY_MI	1.000	1.000	0.000	0.000	0.208	0.000	–	–	+	+	–	+	–
7	D_BY_MU	1.000	1.000	0.000	0.000	0.208	0.000	0.000	–	+	+	–	+	–
8	D_BY_NBB	1.000	1.000	0.000	0.000	0.268	0.000	0.000	0.000	+	+	–	+	–
9	D_BY_SB	1.000	1.000	1.000	1.000	0.655	1.000	1.000	1.000	0.000	+	+	+	+
10	D_ST_MB	1.000	1.000	1.000	1.000	0.782	1.000	1.000	1.000	1.000	0.000	+	+	+
11	I_BGL_TB	0.544	0.404	0.143	0.312	–0.115	0.120	0.120	0.181	0.755	0.805	0.000	+	–
12	I_MIL	1.000	1.000	1.000	1.000	0.911	1.000	1.000	1.000	1.000	1.000	0.927	0.000	+
13	out	0.281	0.196	0.029	0.106	–0.053	0.016	0.016	0.049	0.513	0.524	–0.075	0.804	0.000

In the lower matrix, the exact F_{ST}-values are shown, while the upper matrix shows if they were significant (+) or not significant (– and red); 110 permutations; significance level = 0.0500. Population labels were used as set in **Table 4**.

Table 26. p-values to F_{ST}-values – Population differentiation on eight haplotypes (mtCOI-658 bp-b) within mtCOI-658bp conducted in Arlequin v. 3.5.

Pop-ID	1	2	3	4	5	6	7	8	9	10	11	12	13
CH_FR_MA	*	-	+	+	+	+	+	+	+	+	+	+	+
CH_FRS_BR	0.99099+-0.0030	*	+	+	+	+	+	+	+	+	+	+	+
D_BW_HLD	0.00000+-0.0000	0.00000+-0.0000	*	-	-	-	-	-	+	+	-	+	-
D_BY_FK	0.00000+-0.0000	0.00000+-0.0000	0.99099+-0.0030	*	+	-	-	-	+	+	-	+	+
D_BY_KEH	0.00000+-0.0000	0.03604+-0.0148	0.12613+-0.0309	0.00000+-0.0000	*	-	-	-	+	+	-	+	-
D_BY_MI	0.00000+-0.0000	0.00000+-0.0000	0.99099+-0.0030	0.99099+-0.0030	0.23423+-0.0473	*	-	-	+	+	-	+	-
D_BY_MU	0.00000+-0.0000	0.00000+-0.0000	0.99099+-0.0030	0.99099+-0.0030	0.24324+-0.0526	0.99099+-0.0030	*	-	+	+	-	+	-
D_BY_NBB	0.00000+-0.0000	0.00000+-0.0000	0.99099+-0.0030	0.99099+-0.0030	0.09009+-0.0192	0.99099+-0.0030	0.99099+-0.0030	*	+	+	-	+	-
D_BY_SB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.01802+-0.0121	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*	+	+	+	+
D_ST_MB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*	+	+	+
I_BGL_TB	0.00000+-0.0000	0.00000+-0.0000	0.45946+-0.0394	0.05405+-0.0309	0.99099+-0.0030	0.44144+-0.0628	0.43243+-0.0633	0.22523+-0.0493	0.00000+-0.0000	0.00000+-0.0000	*	+	-
I_MIL	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*	+
out	0.00000+-0.0000	0.00901+-0.0091	0.36937+-0.0533	0.01802+-0.0121	0.79279+-0.0298	0.30631+-0.0309	0.40541+-0.0365	0.15315+-0.0385	0.00000+-0.0000	0.00000+-0.0000	0.99099+-0.0030	0.00000+-0.0000	*

110 permutations,1-13 represent the Pop-IDs on the left in the same order. Population labels were used as set in **Table 4**. Values above the significance level of 0.05 are marked in red.

Table 27. F_{ST}-matrix – Population differentiation on 33 haplotypes (mtCOI-360 bp) based on 959 sequences conducted in Arlequin v. 3.5.

		1	2	3	4	5	6	7	8	9	10
1	Austria	0.000	+	+	-	-	-	+	+	+	+
2	Canada	0.992	0.000	+	+	+	+	+	+	+	+
3	China	0.126	0.362	0.000	+	+	-	+	+	+	+
4	France	0.080	0.825	0.147	0.000	+	+	+	+	+	+
5	Germany	0.045	0.751	0.155	0.098	0.000	+	+	+	+	+
6	Italy	0.104	0.702	0.009	0.159	0.171	0.000	+	+	-	+
7	SouthKorea	0.478	0.811	0.255	0.556	0.561	0.291	0.000	+	+	+
8	Switzerland	0.360	0.474	0.112	0.278	0.280	0.206	0.585	0.000	+	+
9	unknown	0.219	0.548	0.020	0.270	0.310	0.030	0.209	0.227	0.000	+
10	USA	0.242	0.279	0.029	0.224	0.240	0.070	0.349	0.087	0.042	0.000

In the lower matrix, the exact F_{ST}-values are shown, while the upper matrix shows if they were significant (+) or not significant (– and red); 110 permutations; significance level = 0.0500. Population labels were used as set in **Table 4**.

Table 28. p-values to F_{ST}-values – Population differentiation on 33 haplotypes (mtCOI-360 bp) based on 959 sequences conducted in Arlequin v. 3.5.

		1	2	3	4	5	6	7	8	9	10
1	Austria	*									
2	Canada	0.00000+-0.0000	*								
3	China	0.00000+-0.0000	0.00000+-0.0000	*							
4	France	0.24324+-0.0430	0.00000+-0.0000	0.00000+-0.0000	*						
5	Germany	0.10811+-0.0297	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*					
6	Italy	0.09910+-0.0252	0.00000+-0.0000	0.20721+-0.0470	0.00000+-0.0000	0.00000+-0.0000	*				
7	SouthKorea	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*			
8	Switzerland	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*		
9	unknown	0.00901+-0.0091	0.00000+-0.0000	0.02703+-0.0194	0.00000+-0.0000	0.00000+-0.0000	0.08108+-0.0252	0.00000+-0.0000	0.00000+-0.0000	*	
10	USA	0.00901+-0.0091	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.02703+-0.0139	*

110 permutations,1-13 represent the Pop-IDs on the left in the same order. Population labels were used as set in **Table 4**.

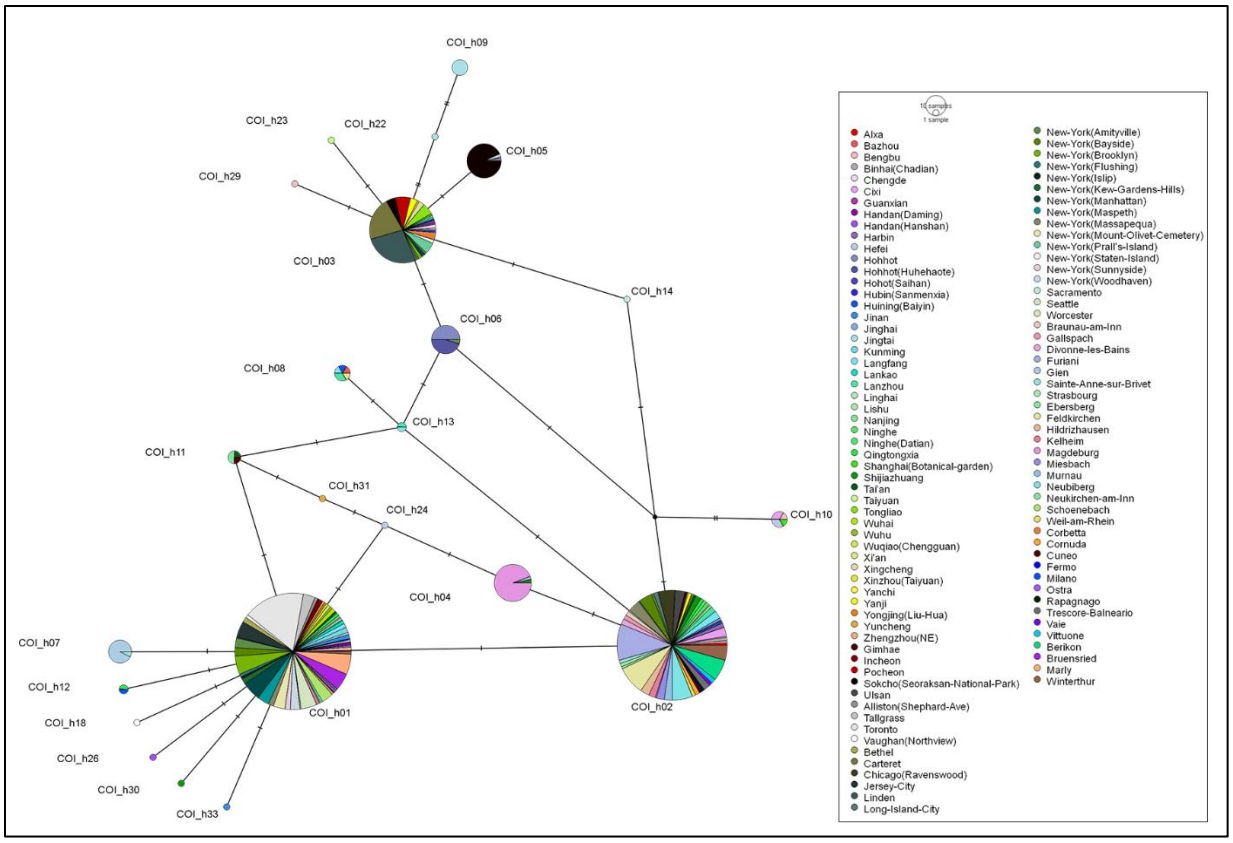


Figure 92. TCS haplotype network of 33 haplotypes of mtCOI-360 bp from 875 specimens from Asia, Europa and North-America (PopART v. 1.7).
Trait colours are drawn per city (109). Mutations are represented by hatch marks; the size of the circles is proportional to the number of specimens contributing to the haplotype.

III. Chapter B

Table 29. megaBLAST of six example cloned ALB-COI-II sequences from Cluster A and B respectively. – (1/3).

	limited to	nucleotide collection				refseq/representative genomes			
Cluster	Query	Subject	Query coverage	E value	Percent identity	Subject	Query coverage	E value	Percent identity
A	4HLD.8_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.50%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.50%
		COI-tRNA-Leu-COII partial cds (EU914812.1)*	100%	0.0	92.20%	scaffold158 (NW_019416265.1)	100%	0.0	93.09%
		COI complete cds (KF737826.1) **	100%	0.0	91.93%	scaffold1356 (NW_019417445.1)	100%	0.0	91.75%
B	4HLD.3_(COI_II)	COI complete cds (KF737825.1)	100%	0.0	99.82%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	95.07%
		COI-tRNA-Leu-COII partial cds (EU914794.1)	100%	0.0	99.82%	scaffold158 (NW_019416265.1)	100%	0.0	94.44%
		COI-tRNA-Leu-COII partial cds (EU914771.1)	100%	0.0	99.82%	scaffold1356 (NW_019417445.1)	100%	0.0	92.02%
A	24WAR.4_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.32%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.32%
		COI-tRNA-Leu-COII partial cds (EU914812.1) *	100%	0.0	92.02%	scaffold158 (NW_019416265.1)	100%	0.0	92.83%
		COI-tRNA-Leu-COII partial cds (EU914688.1)	100%	0.0	91.76%	scaffold1356 (NW_019417445.1)	100%	0.0	91.66%
B	24WAR.1_(COI_II)	COI complete cds (KF737825.1)	100%	0.0	99.91%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	95.16%
		COI-tRNA-Leu-COII partial cds (EU914794.1)	100%	0.0	99.91%	scaffold158 (NW_019416265.1)	100%	0.0	94.53%
		COI-tRNA-Leu-COII partial cds (EU914771.1)	100%	0.0	99.91%	scaffold1356 (NW_019417445.1)	100%	0.0	92.11%

Table 29. megaBLAST of six example cloned ALB-COI-II sequences from Cluster A and B respectively. – (2/3).

	limited to	nucleotide collection				refseq/representative genomes			
Cluster	Query	Subject	Query coverage	E value	Percent identity	Subject	Query coverage	E value	Percent identity
A	33NBB.2_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.50%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.50%
		COI-tRNA-Leu-COII partial cds (EU914812.1) *	100%	0.0	92.38%	scaffold158 (NW_019416265.1)	100%	0.0	93.27%
		COI complete cds (KF737826.1) **	100%	0.0	91.93%	scaffold1356 (NW_019417445.1)	100%	0.0	91.93%
B	33NBB.1_(COI_II)	COI complete cds (KF737825.1)	100%	0.0	99.73%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	94.98%
		COI-tRNA-Leu-COII partial cds (EU914794.1)	100%	0.0	99.73%	scaffold158 (NW_019416265.1)	100%	0.0	94.35%
		COI-tRNA-Leu-COII partial cds (EU914771.1)	100%	0.0	99.73%	scaffold1356 (NW_019417445.1)	100%	0.0	92.11%
A	39MBH.6_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.68%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.68%
		COI-tRNA-Leu-COII partial cds (EU914812.1) *	100%	0.0	92.38%	scaffold158 (NW_019416265.1)	100%	0.0	93.27%
		COI-tRNA-Leu-COII partial cds (EU914688.1)	100%	0.0	92.11%	scaffold1356 (NW_019417445.1)	100%	0.0	91.93%
B	39MBH.1_(COI_II)	COI complete cds (KF737825.1)	100%	0.0	99.91%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	95.16%
		COI-tRNA-Leu-COII partial cds (EU914794.1)	100%	0.0	99.91%	scaffold158 (NW_019416265.1)	100%	0.0	94.53%
		COI-tRNA-Leu-COII partial cds (EU914771.1)	100%	0.0	99.91%	scaffold1356 (NW_019417445.1)	100%	0.0	92.11%

Table 29. megaBLAST of six example cloned ALB-COI-II sequences from Cluster A and B respectively. – (3/3).

	limited to	nucleotide collection				refseq/representative genomes			
Cluster	Query	Subject	Query coverage	E value	Percent identity	Subject	Query coverage	E value	Percent identity
A	50SB.2_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.32%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.32%
		COI-tRNA-Leu-COII partial cds (EU914812.1) *	100%	0.0	92.02%	scaffold158 (NW_019416265.1)	100%	0.0	92.83%
		COI-tRNA-Leu-COII partial cds (EU914688.1)	100%	0.0	91.76%	scaffold1356 (NW_019417445.1)	100%	0.0	91.66%
B	50SB.1_(COI_II)	COI-tRNA-Leu-COII partial cds (EU914759.1)	100%	0.0	99.73%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	94.98%
		COI-tRNA-Leu-COII partial cds (EU914698.1)	100%	0.0	99.73%	scaffold158 (NW_019416265.1)	100%	0.0	94.62%
		COI-tRNA-Leu-COII partial cds (EU914665.1)	100%	0.0	99.73%	scaffold1356 (NW_019417445.1)	100%	0.0	92.20%
A	66FK.2_(COI_II)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.06%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	96.06%
		COI-tRNA-Leu-COII partial cds (EU914812.1) *	100%	0.0	91.85%	scaffold158 (NW_019416265.1)	100%	0.0	92.66%
		COI-tRNA-Leu-COII partial cds (EU914713.1)	100%	0.0	91.50%	scaffold1356 (NW_019417445.1)	100%	0.0	91.41%
B	66FK.1_(COI_II)	COI complete cds (KF737825.1)	100%	0.0	99.82%	mitochondrion, complete genome (NC_008221.1)	100%	0.0	95.07%
		COI-tRNA-Leu-COII partial cds (EU914794.1)	100%	0.0	99.82%	scaffold158 (NW_019416265.1)	100%	0.0	94.44%
		COI-tRNA-Leu-COII partial cds (EU914771.1)	100%	0.0	99.82%	scaffold1356 (NW_019417445.1)	100%	0.0	92.02%

* *Anoplophora chinensis*, ** *Anoplophora lurida*

Table 30. megaBLAST of the complete mitochondrial genome of ALB (NC_008221) and the three mitochondrial counterparts of the sequenced NUMTs.

Query	Subject	Query coverage	E value	Percent identity
374-ND2-COX1-2273(-)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	100.00%
	scaffold1709 (NW_019417922.1)	100%	0.0	95.06%
	scaffold1356 (NW_019417445.1)	100%	0.0	91.48%
906-ND2-COX1-2925(+)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	100.00%
	scaffold1356 (NW_019417445.1)	100%	0.0	92.14%
	scaffold158 (NW_019416265.1)	98%	0.0	94.41%
3555-COX2-ATP8-3856(-)	mitochondrion, complete genome (NC_008221.1)	100%	1,00E-157	100%
	scaffold1356 (NW_019417445.1)	100%	4,00E-122	93.38%
	scaffold1709 (NW_019417922.1)	100%	2,00E-96	87.78%
mitochondrion, complete genome (NC_008221.1)	mitochondrion, complete genome (NC_008221.1)	100%	0.0	100.00%
	scaffold1356 (NW_019417445.1)	67%	0.0	91.96%
	scaffold27 (NW_019416292.1)	18%	0.0	93.42%

limited to representative genomes; not in the first three hits from NC_008221.1 BLAST: scaffold1709 (NW_019417922.1), 22% Query, 0.0 E value, 95.65% identity; scaffold158 (NW_019416265.1) 12% Query, 0.0 E value, 94.41% identity.

Table 31. megaBLAST of the haplotypes of the PCR products for ALBsc158-Numt1. – (1/2)

Query	Subject	Query coverage	E value	Percent identity
Numt1_f_Ref	scaffold158 (NW_019416265.1)	100%	0.0	100%
	scaffold1709 (NW_019417922.1)	95%	0.0	95.21%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.80%
Numt1_f_Type1	scaffold158 (NW_019416265.1)	100%	0.0	97.77%
	scaffold1709 (NW_019417922.1)	95%	0.0	92.50%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	90.91%
Numt1_f_Type2	scaffold158 (NW_019416265.1)	100%	0.0	99.30%
	scaffold1709 (NW_019417922.1)	95%	0.0	94.59%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.20%
Numt1_f_Type3	scaffold1709 (NW_019417922.1)	95%	0.0	93.91%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.49%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.25%
Numt1_f_Type4	scaffold1709 (NW_019417922.1)	95%	0.0	93.76%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.47%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.09%

Table 31. megaBLAST of the haplotypes of the PCR products for ALBsc158-Numt1. – (2/2)

Query	Subject	Query coverage	E value	Percent identity
Numt1_int_Ref	scaffold158 (NW_019416265.1)	100%	0.0	100%
	mitochondrion, complete genome (NC_008221.1)	100%	3.00E-167	94.30%
	scaffold1709 (NW_019417922.1)	99%	1.00E-155	92.71%
Numt1_int_Type 1	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.72%
	scaffold1709 (NW_019417922.1)	100%	0.0	94.33%
	scaffold1356 (NW_019417445.1)	99%	0.0	92.32%
Numt1_int_Type 2	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.29%
	scaffold1709 (NW_019417922.1)	100%	0.0	94.48%
	scaffold1356 (NW_019417445.1)	99%	0.0	92.46%
Numt1_int_Type 3	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.57%
	scaffold1709 (NW_019417922.1)	100%	0.0	94.19%
	scaffold1356 (NW_019417445.1)	99%	0.0	92.18%
Numt1_int_Type 4	mitochondrion, complete genome (NC_008221.1)	100%	0.0	99.57%
	scaffold1709 (NW_019417922.1)	100%	0.0	94.19%
	scaffold1356 (NW_019417445.1)	99%	0.0	92.18%
Numt1_int_Type 5	scaffold158 (NW_019416265.1)	100%	0.0	99.40%
	mitochondrion, complete genome (NC_008221.1)	100%	8.00E-163	93.52%
	scaffold1709 (NW_019417922.1)	99%	8.00E-158	92.97%
Numt1_b_Ref	scaffold158 (NW_019416265.1)	100%	0.0	100%
	mitochondrion, complete genome (NC_008221.1)	94%	0.0	94.64%
	scaffold1356 (NW_019417445.1)	94%	0.0	94.61%
Numt1_b_Type1	scaffold158 (NW_019416265.1)	100%	0.0	99.76%
	scaffold1356 (NW_019417445.1)	95%	0.0	94.50%
	mitochondrion, complete genome (NC_008221.1)	94%	0.0	94.38%
Numt1_b_Type2	scaffold158 (NW_019416265.1)	100%	0.0	99.68%
	scaffold1356 (NW_019417445.1)	95%	0.0	94.41%
	mitochondrion, complete genome (NC_008221.1)	94%	0.0	94.30%
Numt1_b_Type3	scaffold158 (NW_019416265.1)	100%	0.0	98.73%
	scaffold1356 (NW_019417445.1)	95%	0.0	94.33%
	mitochondrion, complete genome (NC_008221.1)	94%	0.0	94.13%
Numt1_b_Type4	scaffold158 (NW_019416265.1)	99%	0.0	99.76%
	scaffold1356 (NW_019417445.1)	95%	0.0	94.50%
	mitochondrion, complete genome (NC_008221.1)	94%	0.0	94.38%

limited to refseq/representative genomes

Table 32. megaBLAST of the haplotypes of the PCR products for ALBsc1709-Numt2.

Query	Subject	Query coverage	E value	Percent identity
Numt2_Ref_AA	scaffold1709 (NW_019417922.1)	100%	0.0	100.00%
	scaffold164 (NW_019416480.1)	100%	0.0	99.77%
	mitochondrion, complete genome (NC_008221.1)	72%	2.00E-96	87.78%
Numt2_TypeA	scaffold164 (NW_019416480.1)	100%	0.0	100.00%
	scaffold1709 (NW_019417922.1)	100%	0.0	99.77%
	mitochondrion, complete genome (NC_008221.1)	72%	1.00E-94	87.46%
Numt2_TypeB	scaffold164 (NW_019416480.1)	100%	0.0	99.30%
	scaffold1709 (NW_019417922.1)	100%	0.0	99.07%
	mitochondrion, complete genome (NC_008221.1)	72%	2.00E-96	87.78%
Numt2_TypeC	scaffold164 (NW_019416480.1)	100%	0.0	99.53%
	scaffold1709 (NW_019417922.1)	100%	0.0	99.30%
	mitochondrion, complete genome (NC_008221.1)	72%	5.00E-98	88.10%

limited to refseq/representative genomes

Table 33. megaBLAST of the haplotypes of the PCR products for ALBsc1709-Numt3. – (1/3)

Query	Subject	Query coverage	E value	Percent identity
Numt3_f_Ref	scaffold1709 (NW_019417922.1)	100%	0.0	100.00%
	scaffold158 (NW_019416265.1)	84%	0.0	94.32%
	mitochondrion, complete genome (NC_008221.1)	84%	0.0	92.76%
Numt3_f_Type1	scaffold1709 (NW_019417922.1)	100%	0.0	98.80%
	scaffold158 (NW_019416265.1)	84%	0.0	94.06%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.76%
Numt3_f_Type2	scaffold1709 (NW_019417922.1)	100%	0.0	98.69%
	scaffold158 (NW_019416265.1)	84%	0.0	93.93%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.64%
Numt3_f_Type3	scaffold1709 (NW_019417922.1)	100%	0.0	98.58%
	scaffold158 (NW_019416265.1)	84%	0.0	94.06%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.76%
Numt3_f_Type4	scaffold1709 (NW_019417922.1)	100%	0.0	98.47%
	scaffold158 (NW_019416265.1)	84%	0.0	93.93%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.64%
Numt3_f_Type5	scaffold1709 (NW_019417922.1)	100%	0.0	98.58%
	scaffold158 (NW_019416265.1)	84%	0.0	93.80%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.51%
Numt3_f_Type6	scaffold1709 (NW_019417922.1)	100%	0.0	98.58%
	scaffold158 (NW_019416265.1)	84%	0.0	94.06%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.76%
Numt3_f_Type7	scaffold1709 (NW_019417922.1)	100%	0.0	98.69%
	scaffold158 (NW_019416265.1)	84%	0.0	93.93%
	scaffold1356 (NW_019417445.1)	84%	0.0	92.64%

Table 33. megaBLAST of the haplotypes of the PCR products for ALBsc1709-Numt3. – (2/3)

Query	Subject	Query coverage	E value	Percent identity
Numt3_int_Ref	scaffold1709 (NW_019417922.1)	100%	0.0	100.00%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	94.00%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.96%
Numt3_int_Type1	scaffold1709 (NW_019417922.1)	100%	0.0	99.36%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.78%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.74%
Numt3_int_Type2	scaffold1709 (NW_019417922.1)	100%	0.0	99.15%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.56%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.51%
Numt3_int_Type3	scaffold1709 (NW_019417922.1)	100%	0.0	99.15%
	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.56%
	scaffold1356 (NW_019417445.1)	95%	0.0	93.51%
Numt3_b_Ref	scaffold1709 (NW_019417922.1)	100%	0.0	100.00%
	scaffold1356 (NW_019417445.1)	100%	0.0	93.27%
	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.45%
Numt3_b_Type1	scaffold1709 (NW_019417922.1)	100%	0.0	99.08%
	scaffold1356 (NW_019417445.1)	100%	0.0	93.37%
	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.24%
Numt3_b_Type2	scaffold1709 (NW_019417922.1)	100%	0.0	98.98%
	scaffold1356 (NW_019417445.1)	100%	0.0	93.27%
	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.14%
Numt3_b_Type3	scaffold1709 (NW_019417922.1)	100%	0.0	98.98%
	scaffold1356 (NW_019417445.1)	100%	0.0	93.27%
	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.14%
Numt3_b_Type4	scaffold1709 (NW_019417922.1)	100%	0.0	98.98%
	scaffold1356 (NW_019417445.1)	100%	0.0	93.27%
	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.14%
Numt3_Ref	scaffold1709 (NW_019417922.1)	100%	0.0	100.00%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.86%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.88%
Numt3_Type1	scaffold1709 (NW_019417922.1)	100%	0.0	98.97%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.86%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.66%
Numt3_Type2	scaffold1709 (NW_019417922.1)	100%	0.0	98.82%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.80%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.60%

Table 33. megaBLAST of the haplotypes of the PCR products for ALBsc1709-Numt3. – (3/3)

Query	Subject	Query coverage	E value	Percent identity
Numt3_Type3	scaffold1709 (NW_019417922.1)	100%	0.0	98.82%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.80%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.60%
Numt3_Type4	scaffold1709 (NW_019417922.1)	100%	0.0	98.87%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.75%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.55%
Numt3_Type5	scaffold1709 (NW_019417922.1)	100%	0.0	98.82%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.80%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.60%
Numt3_Type6	scaffold1709 (NW_019417922.1)	100%	0.0	98.92%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.80%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.60%
Numt3_Type7	scaffold1709 (NW_019417922.1)	100%	0.0	98.87%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.75%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.55%
Numt3_Type8	scaffold1709 (NW_019417922.1)	100%	0.0	98.87%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.86%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.66%
Numt3_Type9	scaffold1709 (NW_019417922.1)	100%	0.0	98.92%
	scaffold1356 (NW_019417445.1)	92%	0.0	92.80%
	mitochondrion, complete genome (NC_008221.1)	92%	0.0	91.60%

limited to refseq/representative genomes

Table 34. megaBLAST of the NUMT references of ALBsc158-Numt1, ALBsc1709-Numt2 and ALBsc1709-Numt3.

Query	Subject	Query coverage	E value	Percent identity
Numt1_f_Ref	mitochondrion, complete genome (NC_008221.1)	95%	0.0	93.80%
	COX1 complete CDS (KF737825.1)	28%	5E-80	92.72%
	COX1 complete CDS (GU003926.1)	28%	5E-80	92.72%
Numt1_int_Ref	COX1 complete CDS (GU003925.1)	57%	0.0	94.56%
	COX1 complete CDS (KF737825.1)	57%	0.0	94.30%
	COX1 complete CDS (GU003927.1)	57%	0.0	94.30%
Numt1_b_Ref	COX1 partial CDS (EU914747.1)	95%	0.0	95.33%
	COX1 partial CDS (EU914688.1)	95%	0.0	95.33%
	COX1 partial CDS (EU914811.1)	95%	0.0	95.25%
Numt2_Ref	mitochondrion, complete genome (NC_008221.1)	72%	2E-97	87.78%
Numt3_f_Ref	COX1 complete CDS (GU003927.1)	84%	0.0	93.64%
	COX1 complete CDS (GU003925.1)	84%	0.0	93.64%
	COX1 partial CDS (AB439153.1)	84%	0.0	93.64%
Numt3_int_Ref	mitochondrion, complete genome (NC_008221.1)	95%	0.0	94.00%
	COX1 complete CDS (KF737825.1)	53%	4E-99	92.80%
	COX1 complete CDS (GU003927.1)	53%	4E-99	92.80%
Numt3_b_Ref	mitochondrion, complete genome (NC_008221.1)	99%	0.0	91.45%

limited to nucleotide collection of *A. glabripennis*

Table 35. Estimates of evolutionary divergence between 16 sequences of ALBsc158_Numt1(f) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) in %.

		1	17
1	NC_008221_-_906-ND2-COX1-2925_(+)_f2		6.405%
2	Numt1(f2)_D-BW-HLD-16-013_A	6.241%	0.152%
3	Numt1(f2)_D-BW-HLD-16-013_B	6.260%	0.762%
4	Numt1(f2)_D-BW-WAR-11-189	6.405%	0.000%
5	Numt1(f2)_D-BW-WAR-15-024	6.405%	0.000%
6	Numt1(f2)_D-BY-FK-14-058	6.405%	0.000%
7	Numt1(f2)_D-BY-FKR-16-076	6.405%	0.000%
8	Numt1(f2)_D-BY-KEH-16-104	6.241%	0.152%
9	Numt1(f2)_D-BY-MU-17-095_A	6.405%	0.000%
10	Numt1(f2)_D-BY-SB-15-047	6.241%	0.152%
11	Numt1(f2)_D-ST-MB-19-149	6.241%	0.152%
12	Numt1(f2)_D-ST-MB-19-164	6.405%	0.000%
13	Numt1(f2)_I-ANM-O-17-170	6.086%	0.609%
14	Numt1(f2)_I-BGL-TB-17-174_B	6.086%	0.609%
15	Numt1(f2)_I-MIL-CB-17-180	6.086%	0.609%
16	Numt1(f2)_I-TOP-V-18-122	6.405%	0.000%
17	Numt1(f2)_Ref	6.405%	

This analysis involved 17 nucleotide sequences and a total of 660 positions in the final dataset.

Table 36. Estimates of evolutionary divergence between 19 sequences of ALBsc158_Numt1-(int) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) in %.

		1	20
1	NC_008221_-_906-ND2-COX1-2925_(+)_int2		5.167%
2	Numt1(int2)_D-BW-HLD-16-013	0.301%	4.837%
3	Numt1(int2)_D-BW-WAR-11-189_B	5.518%	0.302%
4	Numt1(int2)_D-BW-WAR-15-024	0.301%	4.837%
5	Numt1(int2)_D-BY-FK-13-074_A	0.301%	4.837%
6	Numt1(int2)_D-BY-FK-13-074_B	0.453%	5.002%
7	Numt1(int2)_D-BY-FK-14-058	0.301%	4.837%
8	Numt1(int2)_D-BY-FKR-16-076	0.301%	4.837%
9	Numt1(int2)_D-BY-MI-19-194	0.301%	4.837%
10	Numt1(int2)_D-BY-MU-17-095	0.301%	4.837%
11	Numt1(int2)_D-BY-NBB-14-085	0.301%	4.837%
12	Numt1(int2)_D-BY-SB-15-047	0.756%	5.327%
13	Numt1(int2)_D-ST-MB-18-129	0.453%	5.002%
14	Numt1(int2)_D-ST-MB-19-164	0.453%	5.002%
15	Numt1(int2)_I-ANM-O-17-170	0.756%	5.327%
16	Numt1(int2)_I-BGL-TB-17-174	0.756%	5.327%
17	Numt1(int2)_I-CNP-CN-18-124	0.301%	4.837%
18	Numt1(int2)_I-FMM-FM-15-120	0.756%	5.327%
19	Numt1(int2)_I-TOP-V-18-122	0.301%	4.837%
20	Numt1(int2)_Ref	5.167%	

This analysis involved 20 nucleotide sequences and a total of 666 positions in the final dataset.

Table 37. Estimates of evolutionary divergence between 15 sequences of ALBsc158_Numt1-(b) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) in %.

		1	16
1	NC_008221_-_906-ND2-COX1-2925_(+)_b2		5.598%
2	Numt1(b2)_D-BY-FKR-16-076	5.966%	0.336%
3	Numt1(b2)_D-BY-KEH-16-104	5.873%	0.252%
4	Numt1(b2)_D-BY-MU-17-095	5.873%	0.252%
5	Numt1(b2)_D-BY-NBB-14-085_A	5.873%	0.252%
6	Numt1(b2)_D-BY-NBB-14-085_B	5.966%	0.336%
7	Numt1(b2)_D-BY-SB-15-047	5.873%	0.252%
8	Numt1(b2)_D-ST-MB-15-038	5.873%	0.252%
9	Numt1(b2)_D-ST-MB-19-149	5.873%	0.252%
10	Numt1(b2)_D-ST-MB-19-164	5.873%	0.252%
11	Numt1(b2)_I-ANM-O-17-170	6.157%	1.269%
12	Numt1(b2)_I-BGL-TB-17-174	5.873%	0.252%
13	Numt1(b2)_I-CNP-CN-18-124	5.966%	0.336%
14	Numt1(b2)_I-FMM-FM-15-120	6.157%	1.269%
15	Numt1(b2)_I-TOP-V-18-122	5.873%	0.252%
16	Numt1(b2)_Ref	5.598%	

This analysis involved 16 nucleotide sequences and a total of 1193 positions in the final dataset.

Table 38. Estimates of evolutionary divergence between 30 sequences of ALBsc1709-Numt2 and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) converted into percentage.

		1	31
1	NC_008221_-_3555-COX2-ATP8-3856_(-)		10.575%
2	Numt2_D-BW-HLD-16-013_BB	10.581%	1.303%
3	Numt2_D-BW-WAR-11-189_A	10.995%	0.324%
4	Numt2_D-BW-WAR-11-189_B	10.581%	1.303%
5	Numt2_D-BW-WAR-15-024_AA	10.995%	0.324%
6	Numt2_D-BY-FK-13-074_AA	10.995%	0.324%
7	Numt2_D-BY-FK-14-058_AA	10.995%	0.324%
8	Numt2_D-BY-FKR-16-076_AA	10.995%	0.324%
9	Numt2_D-BY-KEH-16-104_AA	10.995%	0.324%
10	Numt2_D-BY-MI-19-194_BB	10.581%	1.303%
11	Numt2_D-BY-MU-17-095_A	10.995%	0.324%
12	Numt2_D-BY-MU-17-095_C	10.204%	0.976%
13	Numt2_D-BY-NBB-14-085_A	10.995%	0.324%
14	Numt2_D-BY-NBB-14-085_B	10.581%	1.303%
15	Numt2_D-BY-SB-15-047_B	10.581%	1.303%
16	Numt2_D-BY-SB-15-047_C	10.204%	0.976%
17	Numt2_D-ST-MB-15-038_A	10.995%	0.324%
18	Numt2_D-ST-MB-15-038_C	10.204%	0.976%
19	Numt2_D-ST-MB-18-129_A	10.995%	0.324%
20	Numt2_D-ST-MB-18-129_B	10.581%	1.303%
21	Numt2_D-ST-MB-19-149_A	10.995%	0.324%
22	Numt2_D-ST-MB-19-149_C	10.204%	0.976%
23	Numt2_D-ST-MB-19-164_BB	10.581%	1.303%
24	Numt2_I-ANM-O-17-170_AA	10.995%	0.324%
25	Numt2_I-BGL-TB-17-174_AA	10.995%	0.324%
26	Numt2_I-CNP-CN-18-124_AA	10.995%	0.324%
27	Numt2_I-FMM-FM-15-120_BB	10.581%	1.303%
28	Numt2_I-MIL-CB-17-180_AA	10.995%	0.324%
29	Numt2_I-MIL-VI-14-181_AA	10.995%	0.324%
30	Numt2_I-TOP-V-18-122_AA	10.995%	0.324%
31	Numt2_Ref2_AA	10.575%	

The number of base substitutions per site from between sequences are shown. Analyses were conducted using the Tamura 3-parameter model (Tamura 1992). This analysis involved 31 nucleotide sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 310 positions in the final dataset.

Table 39. Estimates of evolutionary divergence between 19 sequences of ALBsc1709-Numt3(f2) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) converted into percentage.

		1	30
1	NC_008221_-_374-ND2-COX1-2273_(-)_f2		7.575%
2	Numt3(f2)_D-BW-HLD-16-013	7.890%	0.913%
3	Numt3(f2)_D-BW-WAR-15-024_A	7.890%	0.913%
4	Numt3(f2)_D-BW-WAR-15-024_B	8.183%	1.175%
5	Numt3(f2)_D-BY-FK-13-074	7.891%	1.177%
6	Numt3(f2)_D-BY-FK-14-058_A	7.890%	0.913%
7	Numt3(f2)_D-BY-FK-14-058_B	8.031%	1.043%
8	Numt3(f2)_D-BY-FKR-16-076	8.031%	1.043%
9	Numt3(f2)_D-BY-KEH-16-104_A	8.031%	1.043%
10	Numt3(f2)_D-BY-KEH-16-104_B	7.891%	1.177%
11	Numt3(f2)_D-BY-MI-19-194	8.031%	1.043%
12	Numt3(f2)_D-BY-MU-17-095	8.031%	1.043%
13	Numt3(f2)_D-BY-NBB-14-085	7.890%	0.913%
14	Numt3(f2)_D-BY-SB-15-047	8.031%	1.043%
15	Numt3(f2)_D-ST-MB-15-038	7.890%	0.913%
16	Numt3(f2)_D-ST-MB-18-129	7.890%	0.913%
17	Numt3(f2)_D-ST-MB-19-149	7.890%	0.913%
18	Numt3(f2)_D-ST-MB-19-164	7.890%	0.913%
19	Numt3(f2)_I-ANM-O-17-170	7.891%	1.177%
20	Numt3(f2)_I-BGL-TB-17-174_A	7.890%	0.913%
21	Numt3(f2)_I-BGL-TB-17-174_B	8.031%	1.043%
22	Numt3(f2)_I-CNP-CN-18-124_A	8.031%	1.043%
23	Numt3(f2)_I-CNP-CN-18-124_B	7.750%	1.046%
24	Numt3(f2)_I-FMM-FM-15-120_A	8.041%	1.044%
25	Numt3(f2)_I-FMM-FM-15-120_B	7.890%	0.913%
26	Numt3(f2)_I-MIL-CB-17-180	8.043%	1.309%
27	Numt3(f2)_I-MIL-VI-14-181_A	8.043%	1.309%
28	Numt3(f2)_I-MIL-VI-14-181_B	7.891%	1.177%
29	Numt3(f2)_I-TOP-V-18-122	8.031%	1.043%
30	Numt3(f2)_Ref	7.575%	

This analysis involved 30 nucleotide sequences and a total of 773 positions in the final dataset.

Table 40. Estimates of evolutionary divergence between 23 sequences of ALBsc1709-Numt3(int2) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) converted into percentage.

		1	24
1	NC_008221_-_374-ND2-COX1-2273_(-)_int2		5.447%
2	Numt3(int2)_D-BW-HLD-16-013	5.689%	0.645%
3	Numt3(int2)_D-BW-WAR-11-189	5.689%	0.645%
4	Numt3(int2)_D-BW-WAR-15-024_A	5.689%	0.645%
5	Numt3(int2)_D-BY-FK-13-074	5.934%	0.862%
6	Numt3(int2)_D-BY-FK-14-058_A	5.689%	0.645%
7	Numt3(int2)_D-BY-FKR-16-076	5.689%	0.645%
8	Numt3(int2)_D-BY-KEH-16-104_A	5.689%	0.645%
9	Numt3(int2)_D-BY-KEH-16-104_B	5.933%	0.862%
10	Numt3(int2)_D-BY-MI-19-194	5.689%	0.645%
11	Numt3(int2)_D-BY-MU-17-095	5.689%	0.645%
12	Numt3(int2)_D-BY-NBB-14-085	5.689%	0.645%
13	Numt3(int2)_D-BY-SB-15-047	5.689%	0.645%
14	Numt3(int2)_D-ST-MB-15-038	5.689%	0.645%
15	Numt3(int2)_D-ST-MB-18-129	5.689%	0.645%
16	Numt3(int2)_D-ST-MB-19-149	5.689%	0.645%
17	Numt3(int2)_D-ST-MB-19-164	5.689%	0.645%
18	Numt3(int2)_I-ANM-O-17-170	5.934%	0.862%
19	Numt3(int2)_I-BGL-TB-17-174_A	5.689%	0.645%
20	Numt3(int2)_I-CNP-CN-18-124_A	5.689%	0.645%
21	Numt3(int2)_I-FMM-FM-15-120_A	5.689%	0.645%
22	Numt3(int2)_I-MIL-VI-14-181	5.716%	0.648%
23	Numt3(int2)_I-TOP-V-18-122	5.689%	0.645%
24	Numt3(int2)_Ref	5.447%	

This analysis involved 24 nucleotide sequences and a total of 469 positions in the final dataset.

Table 41. Estimates of evolutionary divergence between 26 sequences of ALBsc1709-Numt3(b2) and the mitochondrial counterpart conducted in MEGA X (Kumar et al. 2018) converted into percentage.

		1	27
1	NC_008221_-_374-ND2-COX1-2273_(-)_b2		8.126%
2	Numt3(b2)_D-BW-HLD-16-013	8.372%	0.928%
3	Numt3(b2)_D-BW-WAR-15-024_A	8.372%	0.928%
4	Numt3(b2)_D-BW-WAR-15-024_B	8.372%	0.928%
5	Numt3(b2)_D-BY-FK-13-074	8.499%	1.033%
6	Numt3(b2)_D-BY-FK-14-058_A	8.372%	0.928%
7	Numt3(b2)_D-BY-FK-14-058_B	8.372%	0.928%
8	Numt3(b2)_D-BY-FKR-16-076	8.372%	0.928%
9	Numt3(b2)_D-BY-KEH-16-104_A	8.372%	0.928%
10	Numt3(b2)_D-BY-KEH-16-104_B	8.497%	1.032%
11	Numt3(b2)_D-BY-MI-19-194	8.372%	0.928%
12	Numt3(b2)_D-BY-MU-17-095	8.372%	0.928%
13	Numt3(b2)_D-BY-NBB-14-085	8.372%	0.928%
14	Numt3(b2)_D-BY-SB-15-047	8.372%	0.928%
15	Numt3(b2)_D-ST-MB-15-038	8.372%	0.928%
16	Numt3(b2)_D-ST-MB-18-129	8.372%	0.928%
17	Numt3(b2)_D-ST-MB-19-164	8.372%	0.928%
18	Numt3(b2)_I-ANM-O-17-170	8.499%	1.033%
19	Numt3(b2)_I-BGL-TB-17-174_A	8.486%	1.031%
20	Numt3(b2)_I-BGL-TB-17-174_B	8.486%	1.031%
21	Numt3(b2)_I-CNP-CN-18-124_A	8.372%	0.928%
22	Numt3(b2)_I-CNP-CN-18-124_B	8.372%	0.928%
23	Numt3(b2)_I-FMM-FM-15-120_A	8.372%	0.928%
24	Numt3(b2)_I-FMM-FM-15-120_B	8.372%	0.928%
25	Numt3(b2)_I-MIL-CB-17-180	8.497%	1.032%
26	Numt3(b2)_I-TOP-V-18-122	8.372%	0.928%
27	Numt3(b2)_Ref	8.126%	

This analysis involved 27 nucleotide sequences and a total of 980 positions in the final dataset.

IV. Chapter C

Table 42. F_{ST}-matrix – Population differentiation on genomic ALB SNP subset 170Indvf8g50m50 (32432 SNPs) conducted in Arlequin v. 3.5.

	Pop-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	CH-FR-MA	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	CH-FRS-BR	**0.051**	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+
3	D-BW-HLD	**0.500**	**0.472**	0.000	+	+	+	+	+	+	+	+	+	+	+	+
4	D-BW-WAR	**0.368**	**0.348**	**0.308**	0.000	+	+	+	+	+	+	+	+	+	+	–
5	D-BY-FK	**0.460**	**0.437**	**0.420**	**0.221**	0.000	+	+	+	+	+	+	+	+	+	+
6	D-BY-KEH	**0.307**	0.223	0.223	0.086	**0.288**	0.000	+	+	+	+	+	+	+	+	+
7	D-BY-MI	**0.538**	**0.521**	**0.421**	**0.328**	**0.422**	0.247	0.000	+	+	+	+	+	+	+	+
8	D-BY-MU	**0.475**	**0.438**	**0.359**	**0.294**	**0.306**	0.202	**0.357**	0.000	+	+	+	+	+	+	+
9	D-BY-NBB	**0.535**	**0.520**	**0.403**	**0.281**	**0.404**	**0.250**	**0.432**	**0.420**	0.000	+	+	+	+	+	+
10	D-BY-SB	**0.468**	**0.448**	**0.342**	**0.345**	**0.457**	0.233	**0.407**	**0.387**	**0.429**	0.000	+	+	+	+	+
11	D-ST-MB	**0.529**	**0.452**	**0.409**	**0.287**	**0.425**	**0.280**	**0.414**	**0.411**	**0.331**	**0.475**	0.000	+	+	+	+
12	I-BGL-TB	0.198	0.168	**0.327**	0.189	**0.316**	0.137	**0.369**	**0.315**	**0.387**	**0.352**	**0.407**	0.000	+	+	+
13	I-MIL-CB	**0.449**	**0.505**	**0.561**	**0.502**	**0.541**	**0.359**	**0.617**	**0.527**	**0.629**	**0.580**	**0.579**	**0.260**	0.000	+	+
14	I-MIL-VI	**0.486**	**0.549**	**0.498**	**0.497**	**0.485**	**0.342**	**0.584**	**0.477**	**0.637**	**0.542**	**0.575**	**0.255**	**0.260**	0.000	+
15	out	0.215	0.132	0.214	0.073	**0.260**	**0.080**	**0.266**	0.220	0.244	0.191	**0.261**	0.103	**0.357**	**0.328**	0.000

Distance method: pairwise differences. In the lower matrix, the exact F_{ST}-values are shown, while the upper matrix shows if they were significant (+) or not significant (– and red) (110 permutations); significance level = 0.0500. Population labels were used as set in **Table 4**. Boxes mark F_{ST}-values >0.5 and bold black letters F_{ST}-values >0.25, while bold green letters mark <0.1.

Table 43. p-values to F_{ST}-values – Population differentiation on genomic SNP subset 170Indvf8g50m50 conducted in Arlequin v. 3.5.

Pop-ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CH-FR-MA	*														
CH-FRS-BR	0.00901+-0.0091	*													
D-BW-HLD	0.00000+-0.0000	0.00000+-0.0000	*												
D-BW-WAR	0.00000+-0.0000	0.00000+-0.0000	0.00901+-0.0091	*											
D-BY-FK	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*										
D-BY-KEH	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*									
D-BY-MI	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*								
D-BY-MU	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*							
D-BY-NBB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*						
D-BY-SB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*					
D-ST-MB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*				
I-BGL-TB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*			
I-MIL-CB	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*		
I-MIL-VI	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00901+-0.0091	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*	
out	0.00000+-0.0000	0.00901+-0.0091	0.00000+-0.0000	0.05405+-0.0242	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*

110 permutations, 1-15 represent the Pop-IDs on the left in the same order. Population labels were used as set in **Table 4**. Values above the significance level of 0.05 are marked in red.

Table 44. Eigenvalues from PCA analysis performed in PLINK v. 1.9 on LD-pruned 170Indvf8g50m50 subset of genomic ALB SNP-Data.

PC	Eigenvalues	% variance explained
1	25.785	17.513
2	16.198	11.002
3	13.670	9.284
4	10.155	6.897
5	9.597	6.518
6	8.883	6.033
7	7.812	5.306
8	7.601	5.163
9	6.958	4.726
10	5.711	3.879
11	4.750	3.226
12	4.109	2.791
13	3.878	2.634
14	3.363	2.284
15	3.135	2.130
16	3.042	2.066
17	2.315	1.572
18	2.277	1.547
19	2.166	1.471
20	2.033	1.380
21	1.944	1.321
22	1.856	1.261

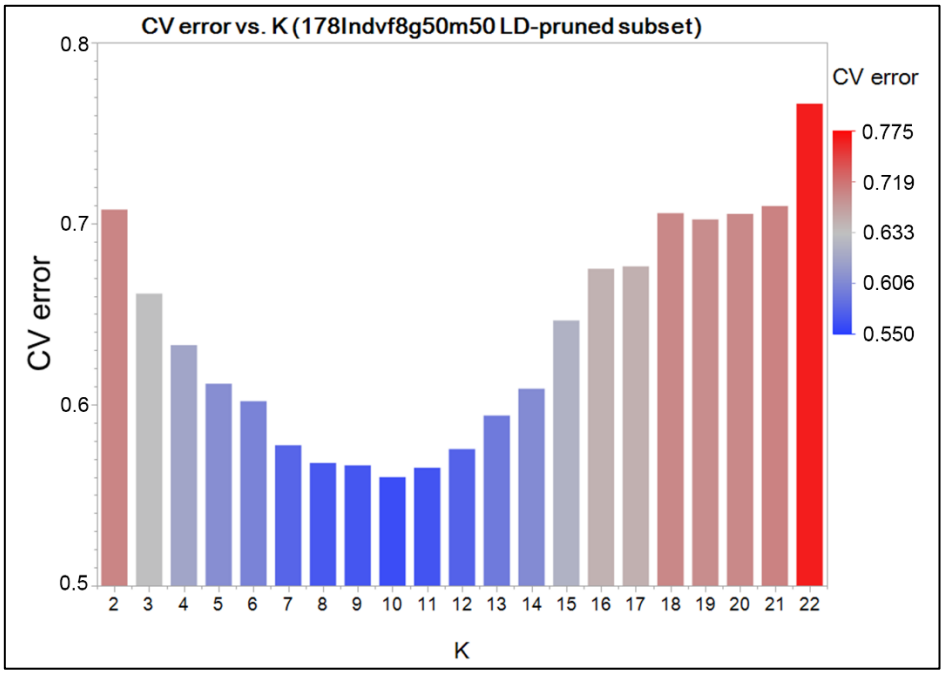


Figure 93. Cross-validation errors to choose the right K-value on the ancestry in the LD-pruned 178Indvf8g50m50 ALB SNP data subset conducted in ADMIXTURE v. 1.3.0.

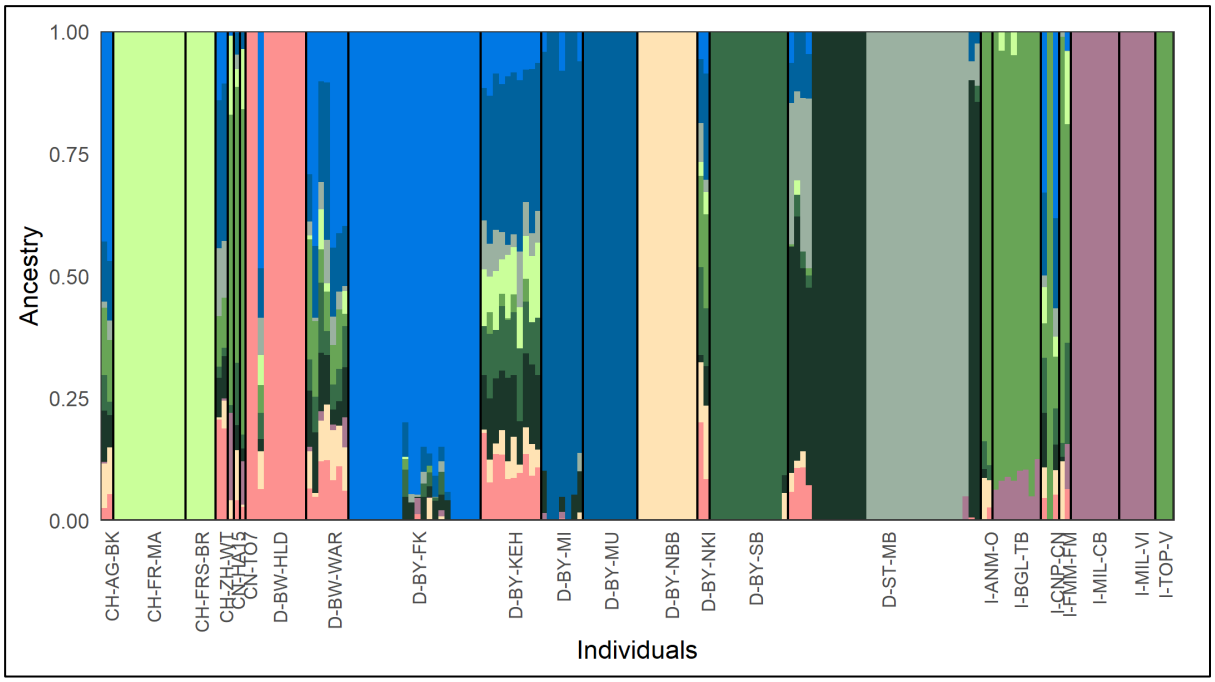


Figure 94. Analysis of Ancestry of 178 ALB samples from Switzerland, China, Germany and Italy conducted in ADMIXTURE v. 1.3.0; K = 10; 7673 SNPs.
The LD-pruned genomic ALB SNP subset 178Indvf8g50m50 was analysed here.

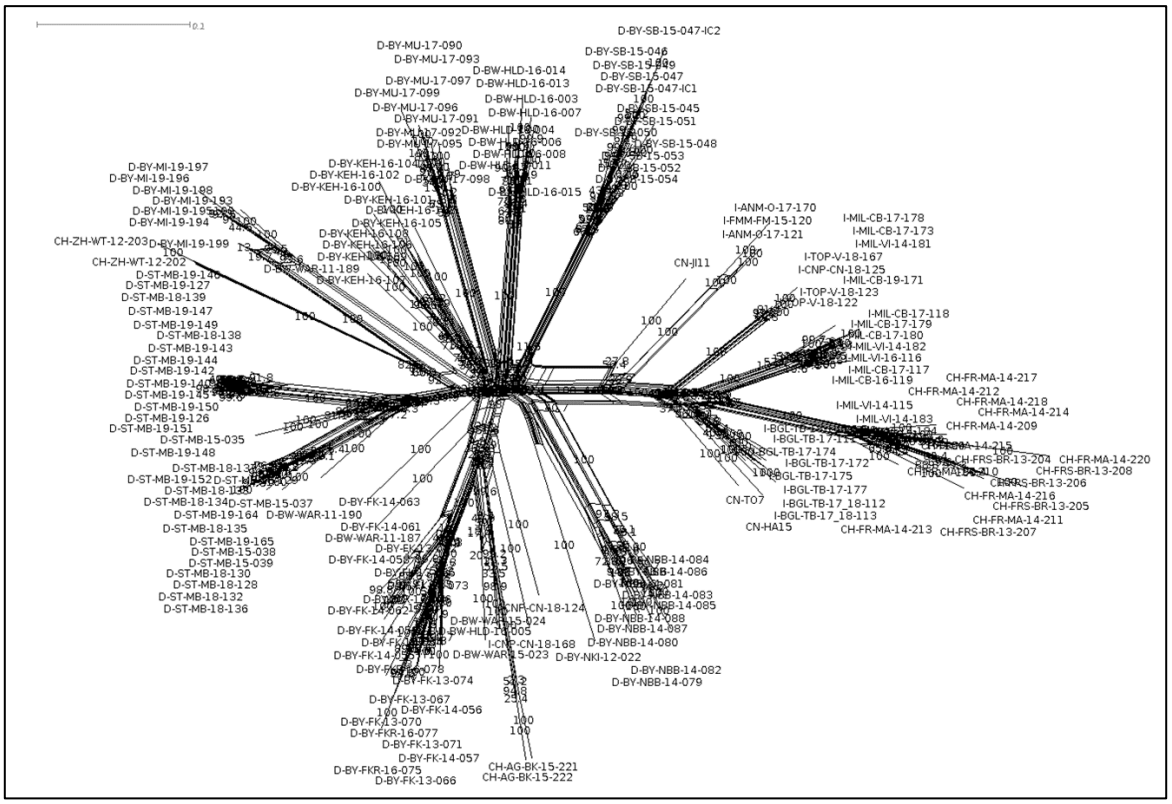


Figure 95. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on ALB SNP data subset 170f8g50m50 from Switzerland, China, Germany and Italy.
The ɣ-distributed (+G 3.73) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 588 splits; Fit = 96.515; 1000 runs Bootstrap; 170 taxa and 32432 SNPs are shown.

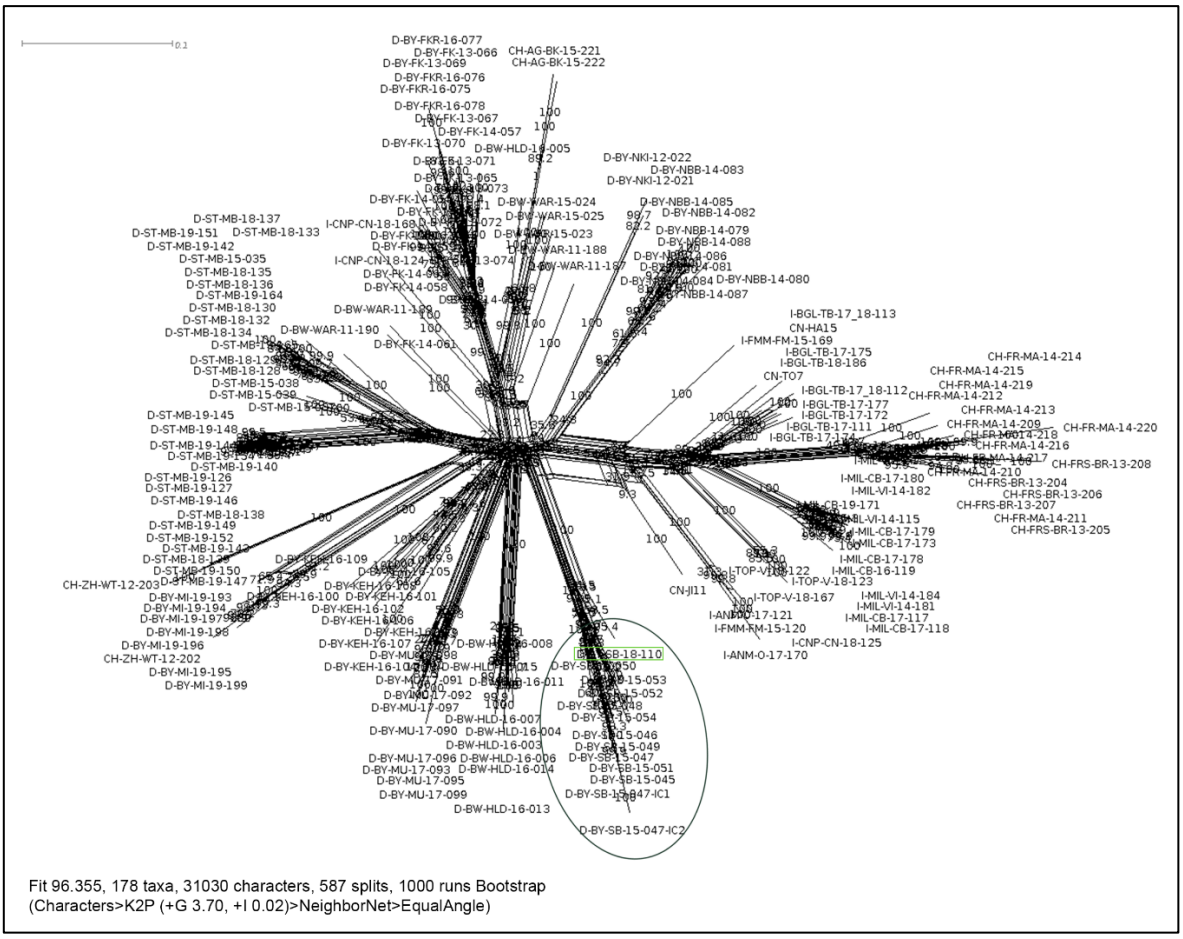


Figure 96. NeighborNet drawn with Splitstree v. 4.17.1 (Huson and Bryant 2006) on ALB SNP data subset 178f8g50m50 from Switzerland, China, Germany and Italy.
The ɣ-distributed (+G 3.70) K2P distances with invariable sites (+I 0.02) and Equal Angle were used to draw the network with 587 splits; Fit = 96.355; 1000 runs Bootstrap; 178 taxa and 31030 SNPs are shown. The dark green ellipse marks the population D-BY-SB, the light green rectangle the single individual D-BY-SB-110.